Adapting Ordinal-Risk Alignment for Clinically Safe Diabetic Retinopathy Grading

Yuhan Zhang[1], Xi Wang[2], Mingchao Li[3], Xiao Ma[4,*], Wufeng Xue[1], Dong Ni[5,6]

[1] School of Biomedical Engineering, Shenzhen University, Shenzhen, China.
[2] Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, China.
[3] Department of Computer Science and Technology, Ocean University of China, Qingdao, China.
[4] Department of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, China.
[5] School of Artificial Intelligence, Shenzhen University, Shenzhen, China.
[6] School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China.
maxiao@njust.edu.cn

Abstract. Automatic diabetic retinopathy (DR) grading is essential for scalable screening and timely referral to prevent vision loss. However, most existing methods treat DR grading as a multi-class classification problem, overlooking the ordinal nature of disease severity and the asymmetric clinical risk, where underestimation is far more detrimental than overestimation. In this paper, we propose Risk-Averse Ordinal Reinforcement Learning (RAO-RL), a novel framework that reformulates DR grading as a risk-aware decision-making task, enabling the direct optimization of non-differentiable clinical risks. First, we introduce ordinal-inductive feature learning with a strict monotonicity constraint to embed the progressive manifold of DR severity directly into the latent space. Then, we employ the cumulative probability regularization that enforces probabilistic consistency across cumulative severity during pretraining. Finally, we propose a reinforcement learning strategy, called group relative risk-averse policy optimization (GR^2PO), to improve accuracy while imposing asymmetric penalties on high-risk misclassifications by clinic-guided rewards. Extensive experiments on four public benchmarks demonstrate that RAO-RL achieves state-of-the-art accuracy while significantly reducing clinically severe errors, offering a safer solution for automatic DR grading. The code is available at github.com/ZhangYH0502/RAO-RL.

Keywords: DR Grading · Fundus Images · Reinforcement Learning.

1 Introduction

Diabetic retinopathy (DR) remains a leading cause of irreversible blindness among the working-age population globally [5]. As shown in Fig. 1, clinical protocols categorize DR severity into a progressive spectrum of five levels: No

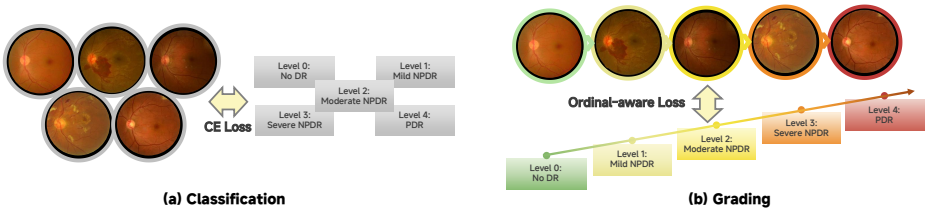

Fig. 1: Differences between the classification task and grading task.

DR (0)→Mild Non-Proliferative DR (NPDR) (1)→Moderate NPDR (2)→Severe NPDR (3)→Proliferative DR (4). DR grading is an inherently predictive task that reflects the progressive accumulation of pathological changes [18]. Consequently, an effective automatic DR grading model should capture the progressive continuity of DR severity and adhere to clinical safety constraints.

Deep learning has achieved remarkable success on DR grading [2], from convolutional neural networks [4, 14, 27] to Transformer-based models [9, 26, 28, 29]. Most existing methods formulated DR grading as a standard multi-class classification problem [3, 25], which leaves two fundamental methodological gaps. First, standard training objectives, such as cross-entropy (CE) loss, impose a uniform misclassification cost and ignore ordinal distances between grades [16]. As a result, large severity jumps (e.g., Level 4 as Level 0) are not penalized more than minor deviations (e.g., Level 4 as Level 3) [7, 21]. Second, more critically, these methods are typically risk-agnostic. Clinically, underestimating severe DR (e.g., Level 4 as Level 1) may delay referral and treatment, whereas the reverse error mainly leads to unnecessary follow-up. However, standard differentiable loss functions are symmetric and struggle to explicitly optimize such asymmetric clinical constraints or maximize complex, non-differentiable evaluation metrics.

To bridge these gaps, we cast DR grading as both an ordinal classification problem and a risk-aware decision-making task. Reinforcement learning (RL), which enables direct optimization of non-differentiable reward functions [17, 30], offers a potent theoretical framework to integrate asymmetric clinical costs into model training [12, 22, 30]. In DR grading, this formulation allows us to design a fine-grained reward mechanism [19, 20, 24] that reflects ordinal distances between grades. Simultaneously, it provides the flexibility to impose asymmetric penalties, explicitly discouraging the model from underestimating severe cases, which is the most critical failure mode in clinical screening. Such risk-aware optimization is crucial because the true clinical value of a DR grading system relies on metrics that account for both grade adjacency and error severity, properties that are often not captured by standard differentiable losses. By optimizing these clinically aligned objectives through RL, the model learns that different misclassifications entail fundamentally different costs.

This paper aims to develop a clinically safe DR grading method by explicitly modeling the progressive nature of DR severity. Our main contributions include: 1) we propose Risk-Averse Ordinal Reinforcement Learning (RAO-RL), a unified

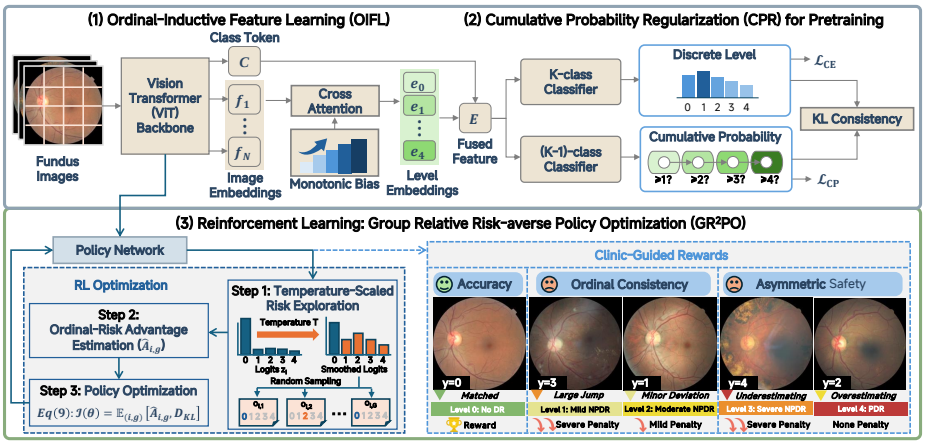

Fig. 2: Overview of RAO-RL, which is first pretrained using OIFL and CPR, and then enhanced via the reinforcement learning strategy GR^{2}PO.

framework that integrates ordinal representation learning and risk-sensitive RL for DR grading; 2) we propose ordinal-inductive feature learning (OIFL) and cumulative probability regularization (CPR) to ensure the latent space respects the progressive manifold of DR severity; 3) we propose a novel RL strategy, called group relative risk-averse policy optimization (GR^{2}PO) with clinic-guided rewards, enabling the direct minimization of high-risk clinical penalty; 4) extensive experiments on four public benchmarks (APTOS, DDR, Messidor-2 and EyePACS) demonstrate that RAO-RL achieves state-of-the-art accuracy while significantly reducing clinically dangerous misclassifications.

2 Method

Let $\boldsymbol{x}$ denote a fundus image and $\mathcal{Y}=\{0,\cdots,K-1\}$ the discrete action set corresponding to the $K=5$ progressive DR levels. RAO-RL (Fig. 2) aims to learn an accurate and reliable mapping $\boldsymbol{x} \mapsto y \in \mathcal{Y}$. To address the class imbalance problem, we use a stratified resampling strategy to draw samples uniformly across DR levels for forming mini-batches. We then sent the mini-batch into the ViT backbone [8] for feature encoding, upon which Ordinal-Inductive Feature Learning (OIFL) injects level-specific progressive cues into the resulting representations. RAO-RL is first pretrained by supervised cross-entropy (CE) loss with Cumulative Probability Regularization (CPR), and then is further enhanced by Group Relative Risk-averse Policy Optimization (GR^{2}PO) with clinic-guided rewards.

2.1 Ordinal-Inductive Feature Learning (OIFL)

Given the class token $\boldsymbol{C}$ and the set of patch embeddings $\boldsymbol{F} = \{\boldsymbol{f}_n\}_{n=1}^{N}$ from the ViT output, we aim to obtain an ordinal-aware representation by injecting level-

specific progression priors. Instead of using independent learnable biases for each level, we introduce a strict monotonicity constraint on the level-specific biases so that they are ordered with increasing DR severity, thereby encoding disease progression in the model. We define the level-specific bias $\boldsymbol{b}_k$ as a cumulative sum of non-negative learnable intervals Δ:

$$\boldsymbol{b}_k = \boldsymbol{b}_0 + \sum_{j=1}^{k} (\Delta_j)^2, \quad \text{where } \boldsymbol{b}_0 = 0,\ k \in [1, K-1]. \tag{1}$$

This formulation ensures that $\boldsymbol{b}_0 < \boldsymbol{b}_1 < \cdots < \boldsymbol{b}_{K-1}$, mathematically guaranteeing that the attention mechanism captures the continuous accumulation of pathological features. The level-specific attention weights are then computed as:

$$\alpha_{n,k} = \exp\left(-\frac{(\sum(\tanh(\text{Linear}(\boldsymbol{f}_n))) + \boldsymbol{b}_k)^2}{2\tau^2}\right), \quad k \in [0, K-1]. \tag{2}$$

By fusing the embeddings $\boldsymbol{f}_n$ with these monotonically constrained weights, the resulting level embeddings $\boldsymbol{e}_k$ inherently respect the ordinal manifold of DR severity. Finally, average and max operations are used over K level embeddings, and we concatenated them with the class token $\boldsymbol{C}$ for the level prediction:

$$\boldsymbol{E} = \text{Linear}(\text{Concat}(\frac{1}{K}\sum_k \boldsymbol{e}_k,\ \max_k \boldsymbol{e}_k,\ \boldsymbol{C})), \quad \boldsymbol{e}_k = \sum_n \frac{\alpha_{n,k}}{\sum_n \alpha_{n,k}} \cdot \boldsymbol{f}_n. \tag{3}$$

2.2 Cumulative Probability Regularization (CPR) for Pretraining

Standard CE loss treats labels as unordered and mutually exclusive categories, failing to capture the ordinal relationships of levels. We introduce cumulative probability regularization [15] to model the inherent order of DR levels in CE. Specifically, we use two independent classifiers to project $\boldsymbol{E}$ to DR level space as $\hat{\boldsymbol{Z}} \in \mathbb{R}^K$ and cumulative probability space as $\hat{\boldsymbol{Z}}_{\text{cp}} \in \mathbb{R}^{K-1}$. $\hat{\boldsymbol{Z}}$ is supervised with a CE loss ($\mathcal{L}_{\text{ce}}$) and $\hat{\boldsymbol{Z}}_{\text{cp}}$ is supervised by a cumulative probability loss ($\mathcal{L}_{\text{cp}}$).

$\mathcal{L}_{\text{cp}}$ models the grading task using $K-1$ binary classifiers for K levels to capture the ordinal relationships. The j-th classifier learns to distinguish whether the predicted level $\hat{y} \geq j$. For a true level y, the cumulative probability label for the $K-1$ classifiers are defined as:

$$\boldsymbol{z}_{\text{cp}}^j = \begin{cases} 1 & \text{if } y \geq j \\ 0 & \text{otherwise} \end{cases}, \quad j \in [1, K-1]. \tag{4}$$

Given the output probabilities for $K-1$ binary classifiers $\hat{\boldsymbol{Z}}_{\text{cp}} = \{\hat{\boldsymbol{z}}_{\text{cp}}^j\}_{j=1}^{K-1}$ that has been normalized via the sigmoid function, $\mathcal{L}_{\text{cp}}$ is the average of binary CE losses over all $K-1$ classifiers. This formulation explicitly enforces the ordinal structure by learning cumulative relationships between DR levels.

The total supervised loss for pretraining is the combination of $\mathcal{L}_{\text{ce}}$, $\mathcal{L}_{\text{cp}}$ and their KL consistency regularization:

$$\mathcal{L}_{\text{pre}} = \mathcal{L}_{\text{ce}}(\hat{\boldsymbol{Z}}, \boldsymbol{Z}) + \mathcal{L}_{\text{cp}}(\hat{\boldsymbol{Z}}_{\text{cp}}, \boldsymbol{Z}_{\text{cp}}) + \text{KL}(\hat{\boldsymbol{Z}} || \mathcal{T}(\hat{\boldsymbol{Z}}_{\text{cp}})), \tag{5}$$

where $\mathcal{T}(\cdot)$ converts the $K-1$ binary cumulative probabilities $\hat{\boldsymbol{Z}}_{\text{cp}}$ into a discrete probability distributions over K levels.

2.3 Group Relative Risk-averse Policy Optimization (GR^{2}PO)

While pretraining establishes a probabilistic representation, standard optimizes for average accuracy, often failing to capture the ordinal-risk asymmetry inherent in clinical diagnosis. To address this, we propose Group Relative Risk-averse Policy Optimization (GR^{2}PO), which explicitly embeds clinical safety constraints into the advantage estimation process, achieving ordinal-risk alignment directly within the optimization loop without requiring a separate value network.

Clinic-Guided Reward Construction: The foundation of GR^{2}PO is a structured reward function designed to translate clinical protocols into scalar feedback. For a sampled prediction $\hat{y}$ given ground truth y, the reward $r(\hat{y}, y)$ is composed of accuracy, ordinal consistency, and asymmetric safety:

$$r(\hat{y}, y) = \underbrace{\mathbb{I}(\hat{y} = y)}_{\text{Accuracy}} - \underbrace{\lambda_{\text{dist}} \cdot \left| \frac{\hat{y} - y}{K - 1} \right|^{\gamma_1}}_{\text{Ordinal Consistency}} - \underbrace{\lambda_{\text{safe}} \cdot \mathbb{I}(\hat{y} < y) \cdot |\hat{y} - y|^{\gamma_2}}_{\text{Asymmetric Safety}}. \tag{6}$$

The first term rewards exact matches. The second term, weighted by λ_{dist}, imposes a distance-based penalty to enforce the ordinal nature of DR severity. The third term is the asymmetric safety penalty, which imposes a severe penalty when the model fails to refer a patient requiring urgent treatment.

Optimization Objective: Both the active policy π_θ and the reference model π_{ref} are built upon the CPR pre-trained ViT model. GR^{2}PO utilizes group-wise statistics to optimize π_θ towards the constructed reward for in three steps.

Step 1: Temperature-Scaled Risk Exploration. Standard GRPO relies on sampling a diverse group to estimate relative advantages. However, pretrained classification models typically yield overconfident predictions, which collapses the group variance to zero. Tailored for DR grading task, we design a *temperature-scaled multinomial sampling* strategy to generate the required group. For an input image $\boldsymbol{x}_i$ with logits $\boldsymbol{z}_i$, we draw a group of G predictions $\{o_{i,1}, \ldots, o_{i,G}\}$:

$$o_{i,g} \sim \text{Multinomial}\left(\text{Softmax}\left(\frac{\boldsymbol{z}_i}{T}\right)\right), \quad g \in [1, G], \tag{7}$$

where the Multinomial$(\cdot)$ operation performs G independent random samplings with replacement. By applying a temperature $T>1$, this operation explicitly softens the distribution, forcing the policy to explore adjacent levels and exposing unsafe decision boundaries to the reward function.

Step 2: Ordinal-Risk Advantage Estimation. To isolate the relative quality of each sampled prediction from the intrinsic difficulty of the input image,

we compute the standardized advantage $\hat{A}_{i,g}$ using the group-wise mean μ_i and standard deviation σ_i of the rewards:

$$\hat{A}_{i,g} = \frac{r(o_{i,g}, y_i) - \mu_i}{\sigma_i + \epsilon}, \quad \text{where } \mu_i = \text{Mean}(\{r_{i,j}\}_{j=1}^{G}),\ \sigma_i = \text{Std}(\{r_{i,j}\}_{j=1}^{G}). \quad (8)$$

Step 3: Policy Optimization. We maximize the expected advantage using a clipped surrogate objective to ensure monotonic improvement without destructive parameter updates. Let $\rho_{i,g}(\theta) = \frac{\pi_\theta(o_{i,g}|\boldsymbol{x}_i)}{\pi_{\theta_{old}}(o_{i,g}|\boldsymbol{x}_i)}$ denotes the probability ratio between the current and old policies, the final objective incorporates both the PPO clipping and a KL divergence penalty to prevent catastrophic forgetting:

$$\mathcal{J}(\theta) = \mathbb{E}_{(i,g)}\left[\min(\rho_{i,g}\hat{A}_{i,g}, \text{clip}(\rho_{i,g}, 1-\epsilon, 1+\epsilon)\hat{A}_{i,g}) - \beta D_{\text{KL}}(\pi_\theta||\pi_{\text{ref}})\right], \quad (9)$$

where ϵ is the clipping hyperparameter and β controls the regularization strength relative to the frozen pretrained model π_{ref}.

3 Experiments

3.1 Materials and Details

We conducted comprehensive experiments on four publicly available DR grading datasets, including APTOS [11], DDR [13], Messidor-2 [1] and EyePACS [6]. In the four datasets, all fundus images were captured under a variety of imaging conditions at multiple primary care sites. Each fundus image was annotated on a scale of 0 to 4 according to the DR severity by ophthalmologists. For each dataset, we followed its original split of the training set and test set, and randomly selected 10% samples from training set as the validation set.

We automatically removed the black background using Sobel edge detection followed by cropping. The cropped fundus images were then resized to 224×224 and normalized to $[0, 1]$ as consistent model inputs. We chose the Adam optimizer with an initial learning rate of 10^{-4} and a weight decay of 0.1. The batch size was set to 32. Basic data augmentation strategies were used in the training process. Experiments were built on two NVIDIA 4090 GPUs.

3.2 Overall Evaluation

We compared RAO-RL against five recent state-of-the-art DR grading methods, including Triple-DRNet (2023) [10], CLIP-DR (2024) [25], AOR-DR (2025) [26], CSFHANet (2025) [14] and LECNet (2025) [7], on four public datasets. As shown in Table 1, RAO-RL achieves the best results in terms of quadratic weighted kappa (QWK), F1-score and accuracy (ACC) as evaluation metrics over all competing methods. The improvements in QWK are particularly notable, with absolute gains over the second-best methods being approximately +3.1% (APTOS), +5.7% (DDR), +5.0% (Messidor-2) and +5.6% (EyePACS). Improvements in F1 (ranging from +2.0% to +8.5%) and ACC (ranging from +3.4% to +6.1%)

Table 1: Overall evaluations on four publicly available DR grading datasets between RAO-RL and other five state-of-the-art DR grading methods. Three quantitative metrics (QWK, F1-score and ACC) are used.

Model	APTOS [11]			DDR [13]			Messidor-2 [1]			EyePACS [6]		
	QWK↑	F1↑	ACC↑	QWK↑	F1↑	ACC↑	QWK↑	F1↑	ACC↑	QWK↑	F1↑	ACC↑
Triple-DRNet [10]	0.909	0.677	0.857	0.757	0.550	0.622	0.758	0.523	0.710	0.838	0.754	0.799
CLIP-DR [25]	0.844	0.606	0.791	0.719	0.513	0.594	0.726	0.421	0.676	0.781	0.671	0.738
AOR-DR [26]	0.861	0.604	0.818	0.731	0.541	0.617	0.739	0.501	0.704	0.819	0.718	0.770
CSFHANet [14]	0.873	0.641	0.852	0.805	0.663	0.741	0.795	0.562	0.733	0.851	0.802	0.847
LECNet [7]	0.915	0.707	0.865	0.821	0.636	0.738	0.777	0.519	0.740	0.841	0.816	0.861
RAO-RL (Ours)	**0.946**	**0.792**	**0.905**	**0.878**	**0.683**	**0.802**	**0.845**	**0.610**	**0.774**	**0.907**	**0.865**	**0.903**

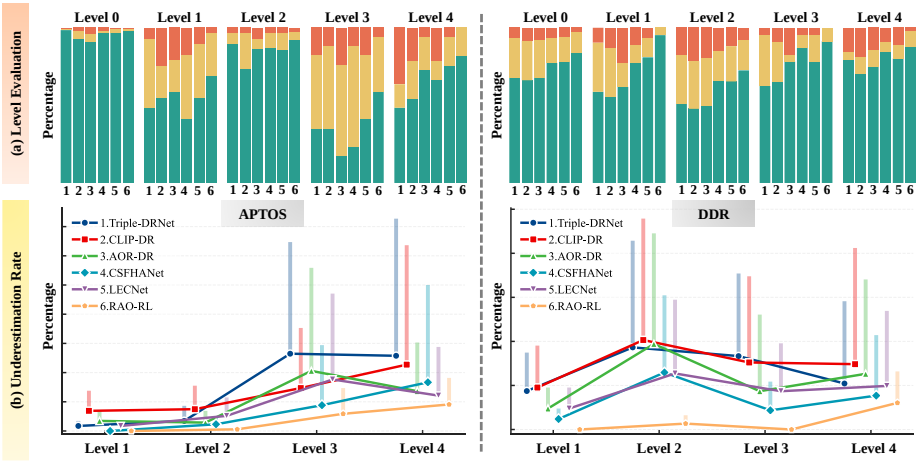

Fig. 3: Evaluation of six methods (numbers 1-5 are the competing methods and number 6 is our proposed RAO-RL) for each level on APTOS and DDR. (a): The percentage of test sets that are Correct Classification (green), Adjacent Misclassification (yellow) and Farther Misclassification (red). (b): Underestimation analysis across DR levels, where the lines denote the absolute underestimation rate (AUR), the bars denote the weighted underestimation rate (WUR).

are of similar magnitude and consistent with the QWK gains. Overall, these consistent gains across metrics and benchmarks indicate that RAO-RL delivers reliable improvements over recent methods, supporting its effectiveness for DR grading under practical clinical scenarios.

3.3 Clinically Safe Evaluation

To evaluate the clinical safety of RAO-RL, we presented the distribution of grading errors across five DR levels on APTOS and DDR. As shown in Fig. 3(a), the bars of each level are decomposed into three components: Correct Classification

(green; the level is predicted correctly), Adjacent Misclassification (yellow; the level is misclassified into an adjacent grade), and Farther Misclassification (red; the level is misclassified into a non-adjacent grade). On APTOS, RAO-RL consistently yielded the highest correct proportion, demonstrating superior discriminative power, particularly at Levels 1-4. Besides, RAO-RL drastically minimized the farther misclassification, the most clinically hazardous error, ensuring that the majority of misclassified samples remain in the adjacent category, which is more acceptable in a screening context. The results on DDR further validated the robustness of RAO-RL.

To evaluate the clinical safety, we performed a detailed underestimation analysis by using the absolute underestimation rate (AUR) and weighted underestimation rate (WUR) [23] to evaluate the DR levels 1-4:

$$\mathrm{AUR}_k = \frac{\#\{i : y_i^k > \hat{y}_i^k\}}{\#\{i : y_i^k = k\}}, \quad \mathrm{WUR}_k = \frac{\sum_i \max(0, y_i^k - \hat{y}_i^k)}{\#\{i : y_i^k = k\}}, \quad k \in [1, 4]. \tag{10}$$

In Fig. 3(b), while most baselines exhibited a high AUR for Levels 2-4, RAO-RL maintained a remarkably low error profile. Even in the most critical Level 4 cases, RAO-RL significantly restricted the WUR, indicating that even when misclassifications occur, they were predominantly assigned to the adjacent Level 3 rather than being dangerously dismissed as Normal or Mild.

3.4 Ablation Study

Component-level ablation. We quantify the contribution of each module in RAO-RL by progressively adding OIFL, CPR, and GR^{2}PO under the same backbone and evaluation protocol. In Table 2(a), introducing OIFL consistently improves the baseline on both datasets, indicating that enforcing ordinal structure in the latent space benefits DR grading. Adding CPR further strengthens ordinal calibration and safety, yielding higher QWK and lower under-grading risks (AUR/WUR). Finally, incorporating GR^{2}PO (Table 2(c)) leads to the most substantial safety gains, markedly reducing clinically severe underestimation while simultaneously improving overall performance. These results demonstrate the advantage of directly optimizing risk-aware objectives via RL.

Reward ablation in GR^{2}PO. We further dissect the clinic-guided reward by ablating its components while keeping the group-relative sampling architecture of GR^{2}PO fixed (Table 2(b)). An "Only Accuracy" reward optimizes purely for group-relative correctness. However, it treats near-miss errors identically to clinically severe under-grading during advantage estimation. Introducing the *asymmetric safety* penalty explicitly penalizes critical false negatives within the group, noticeably reducing safety metrics. Conversely, the *ordinal consistency* penalty penalizes distance errors, serving as a global smoother that substantially improves ordinal metrics like QWK. Combining all terms (Table 2(c)) exhibits a profound synergy, achieving the highest QWK and lowest WUR. This confirms that the ordinal term encourages smoothly ordered predictions, while the asymmetric term enforces strict clinical boundaries.

Table 2: Ablation Study on APTOS and DDR datasets based on five metrics.

	Model	APTOS [11]					DDR [13]				
		QWK↑	F1↑	ACC↑	AUR↓	WUR↓	QWK↑	F1↑	ACC↑	AUR↓	WUR↓
(a)	Baseline: ViT [8]	0.835	0.612	0.776	0.259	0.382	0.714	0.505	0.588	0.295	0.415
	ViT+OIFL	0.864	0.646	0.821	0.228	0.325	0.749	0.564	0.633	0.259	0.334
	ViT+OIFL+CPR	0.918	0.720	0.858	0.184	0.218	0.817	0.659	0.746	0.222	0.296
(b)	Only Accuracy	0.920	0.728	0.865	0.179	0.204	0.828	0.663	0.750	0.176	0.240
	Accuracy+Asym	0.929	0.743	0.869	0.144	0.180	0.835	0.668	0.761	0.141	0.169
	Accuracy+Ordinal	0.933	0.771	0.882	0.122	0.171	0.853	0.675	0.776	0.124	0.153
(c)	RAO-RL (Full)	**0.946**	**0.792**	**0.905**	**0.078**	**0.092**	**0.878**	**0.683**	**0.802**	**0.037**	**0.045**

4 Conclusion

In this paper, we propose RAO-RL, a unified framework for DR grading that integrates ordinal feature learning with risk-aware RL. Our experiments demonstrate two practical advantages. First, RAO-RL significantly reduces the dangerous underestimation of severe cases, ensuring higher sensitivity for patients requiring urgent treatment. Second, it restricts most grading errors to adjacent levels, avoiding extreme misclassifications that violate the natural progression of the disease. By effectively balancing high accuracy with minimized clinical risk, RAO-RL offers a safer and more trustworthy solution for automatic DR grading.

Acknowledgments. This study was funded by the National Natural Science Foundation of China under grants 62572324, 62402275 and 62402458.

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

References

1. Abràmoff, M.D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J.C., Niemeijer, M.: Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Investigative ophthalmology & visual science **57**(13), 5200–5206 (2016)
2. Almattar, W., Luqman, H., Khan, F.A.: Diabetic retinopathy grading review: Current techniques and future directions. Image and Vision Computing **139**, 104821 (2023)
3. Bi, Q., Yi, J., Zheng, H., Zhan, H., Huang, Y., Ji, W., Li, Y., Zheng, Y.: Gad: Domain generalized diabetic retinopathy grading by grade-aware de-stylization. Pattern Recognition **164**, 111484 (2025)
4. Che, H., Cheng, Y., Jin, H., Chen, H.: Towards generalizable diabetic retinopathy grading in unseen domains. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 430–440. Springer (2023)
5. Cho, N.H., Shaw, J.E., Karuranga, S., Huang, Y., da Rocha Fernandes, J.D., Ohlrogge, A., Malanda, B.: Idf diabetes atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes research and clinical practice **138**, 271–281 (2018)

6. Dugas, E., Jared, Jorge, Cukierski, W.: Diabetic retinopathy detection. https://kaggle.com/competitions/diabetic-retinopathy-detection (2015), kaggle
7. Guo, Y., Du, H., Zhang, Y., Ma, F., Meng, J.: Lecnet: Lesion semantic feature enhancement and complementary networks for diabetic retinopathy grading. Biomedical Signal Processing and Control **107**, 107813 (2025)
8. Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., et al.: A survey on vision transformer. IEEE transactions on pattern analysis and machine intelligence **45**(1), 87–110 (2022)
9. Huang, Y., Lyu, J., Cheng, P., Tam, R., Tang, X.: Ssit: Saliency-guided self-supervised image transformer for diabetic retinopathy grading. IEEE Journal of Biomedical and Health Informatics **28**(5), 2806–2817 (2024)
10. Jian, M., Chen, H., Tao, C., Li, X., Wang, G.: Triple-drnet: A triple-cascade convolution neural network for diabetic retinopathy grading using fundus images. Computers in Biology and Medicine **155**, 106631 (2023)
11. Karthik, Maggie, Dane, S.: Aptos 2019 blindness detection. https://kaggle.com/competitions/aptos2019-blindness-detection (2019), kaggle
12. Le, N., Rathour, V.S., Yamazaki, K., Luu, K., Savvides, M.: Deep reinforcement learning in computer vision: a comprehensive survey. Artificial Intelligence Review **55**(4), 2733–2819 (2022)
13. Li, T., Gao, Y., Wang, K., Guo, S., Liu, H., Kang, H.: Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. Information Sciences **501**, 511–522 (2019)
14. Lin, Z., He, Z., Wang, X., Su, W., Tan, J., Deng, Y., Xie, S.: Cross-scale fuzzy holistic attention network for diabetic retinopathy grading from fundus images. IEEE Transactions on Emerging Topics in Computational Intelligence (2025)
15. Liu, X., Ge, Y., Ye, P., You, J., Lu, J.: Ordinal unsupervised domain adaptation with recursively conditional gaussian imposed variational disentanglement. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022)
16. Luo, X., Xu, Q., Wang, Z., Huang, C., Liu, C., Jin, X., Zhang, J.: A lesion-fusion neural network for multi-view diabetic retinopathy grading. IEEE Journal of Biomedical and Health Informatics (2024)
17. Minakshee, M.: Enhanced diabetic retinopathy diagnosis: A comparative analysis of models with an ensemble classifier and deep q learning. J. Electrical Systems **20**(6s), 2613–2624 (2024)
18. Nage, P., Shitole, S., Kokare, M.: An intelligent approach for detection and grading of diabetic retinopathy and diabetic macular edema using retinal images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization **11**(5), 1625–1640 (2023)
19. Rayavel, P., Priyadharshini, S.A., Midhunadharshini, G., Divyedharshini, G.: Detection and classification of diabetic retinopathy in retinal fundus images using deep spiking q network optimized with partial reinforcement optimizer. Computers in Biology and Medicine **196**, 110863 (2025)
20. Rehman, A., Naijie, G., Ojo, S., Nathaniel, T.I., Samee, N.A., Umer, M., Jamjoom, M.M.: Fism: harnessing deep learning and reinforcement learning for precision detection of microaneurysms and retinal exudates for early diabetic retinopathy diagnosis. BioData Mining **18**(1), 75 (2025)
21. Romero-Oraá, R., Herrero-Tudela, M., López, M.I., Hornero, R., García, M.: Attention-based deep learning framework for automatic fundus image processing to aid in diabetic retinopathy grading. Computer Methods and Programs in Biomedicine **249**, 108160 (2024)

22. Shakya, A.K., Pillai, G., Chakrabarty, S.: Reinforcement learning algorithms: A brief survey. Expert Systems with Applications **231**, 120495 (2023)
23. Telo, G.H., Friedrich Fontoura, L., Avila, G.O., Gheno, V., Bertuzzo Brum, M.A., Teixeira, J.B., Erthal, I.N., Alessi, J., Telo, G.H.: Obesity bias: How can this underestimated problem affect medical decisions in healthcare? a systematic review. Obesity Reviews **25**(4), e13696 (2024)
24. Vadduri, M., Kuppusamy, P.: Eho-q-lgan: an eho-based q-learning gan for the timely diagnosis of diabetic retinopathy. IEEE Access (2025)
25. Yu, Q., Xie, J., Nguyen, A., Zhao, H., Zhang, J., Fu, H., Zhao, Y., Zheng, Y., Meng, Y.: Clip-dr: Textual knowledge-guided diabetic retinopathy grading with ranking-aware prompting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 667–677. Springer (2024)
26. Yu, Q., Zhou, W., Liu, H., Xu, Y., Wang, M., Zhao, Y., Fu, H., Ye, X., Zheng, Y., Meng, Y.: Parameterized diffusion optimization enabled autoregressive ordinal regression for diabetic retinopathy grading. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 450–460. Springer (2025)
27. Yue, G., Li, Y., Zhou, T., Zhou, X., Liu, Y., Wang, T.: Attention-driven cascaded network for diabetic retinopathy grading from fundus images. Biomedical Signal Processing and Control **80**, 104370 (2023)
28. Zang, F., Ma, H.: Cra-net: Transformer guided category-relation attention network for diabetic retinopathy grading. Computers in Biology and Medicine **170**, 107993 (2024)
29. Zhang, Y., Ma, X., Huang, K., Li, M., Heng, P.A.: Semantic-oriented visual prompt learning for diabetic retinopathy grading on fundus images. IEEE Transactions on Medical Imaging **43**(8), 2960–2969 (2024)
30. Zhou, S.K., Le, H.N., Luu, K., Nguyen, H.V., Ayache, N.: Deep reinforcement learning in medical imaging: A literature review. Medical image analysis **73**, 102193 (2021)