

Recurrent Contrastive Learning for Imbalanced Medical Image Classification

Zhiyuan Zhu^{1,2*}, Xinling Meng^{3*}, Junxuan Yu^{1,2*}, Jiongquan Chen^{1,2*},
Qiongying Ni^{1,2}, Tuhang Shao^{1,2}, Yuhao Huang⁴, Luping Zhou^{1,2}, Ruiyang
Huang^{1,2}, Yuxue Wang⁵, Rongliang Zhang⁵, Xue Wang⁵, Tianhong Tang⁶,
Likun Wang⁷, Junbo Chen³, Yong Jiang⁶, Yongping Lu⁵(✉), and Xin
Yang^{1,2}(✉)

¹ School of Biomedical Engineering, Medical School, Shenzhen University, China

² Ultra-X AI Lab, Shenzhen University, Guangdong, China

³ School of Biomedical Engineering, South-Central Minzu University, Wuhan, China

⁴ Department of Radiology, Boston Children's Hospital and Harvard Medical School,
Boston, MA, USA

⁵ The Affiliated Hospital of Yunnan University, Kunming, China

⁶ Fuwai Shenzhen Hospital, Chinese Academy of Medical Sciences, Shenzhen, China

⁷ First Affiliated Hospital of Hebei North University, Zhangjiakou, China

Abstract. Medical image classification often suffers from class imbalance due to the inherent disparities in disease incidence. Existing approaches, such as class resampling and loss reweighting, mainly improve learning within the observed feature distribution, but do not explicitly enlarge the latent support region of tail classes. As a result, tail-class representations remain overly compact and are easily encroached upon by head classes, leading to biased decision boundaries. In this work, we propose **Recurrent Contrastive Learning (RCL)** for imbalanced medical image classification. RCL progressively expands the support region of tail classes by recurrently reusing historical feature states across training phases. Specifically, we adopt DINOv3 with LoRA adapters as the backbone to provide robust feature embeddings. We then devise a **Temporal Memory Queue (TMQ)** to preserve corpus-level features across training phases and provide diversified global references for contrastive learning. Based on TMQ, we construct **Temporal Anchors (TARs)** to form an anchor field around tail classes. This field enlarges the support region of tail classes, suppresses head-class encroachment, and improves inter-class separation. Extensive experiments on three imbalanced medical datasets demonstrate that RCL achieves consistent improvements over strong baselines. The code is available at <https://github.com/dndins/RCL>.

Keywords: Recurrent contrastive learning · Imbalanced learning · Medical image classification

* Zhiyuan Zhu, Xinling Meng, Junxuan Yu and Jiongquan Chen contribute equally to this work. Corresponding emails: luyongp@163.com and xinyang@szu.edu.cn

1 Introduction

Deep learning often benefits from large and balanced datasets to effectively learn and model feature distributions across different categories [11]. However, in real-world medical scenarios, as shown in Fig. 1 (a), the prevalence of different diseases varies substantially, resulting in a limited number of samples for rare conditions [2]. This imbalance biases model training toward head classes and weakens representation learning for tail classes [3]. Notably, tail classes often correspond to severe or rare disease types, where accurate identification is more critical for real-world clinical decision-making.

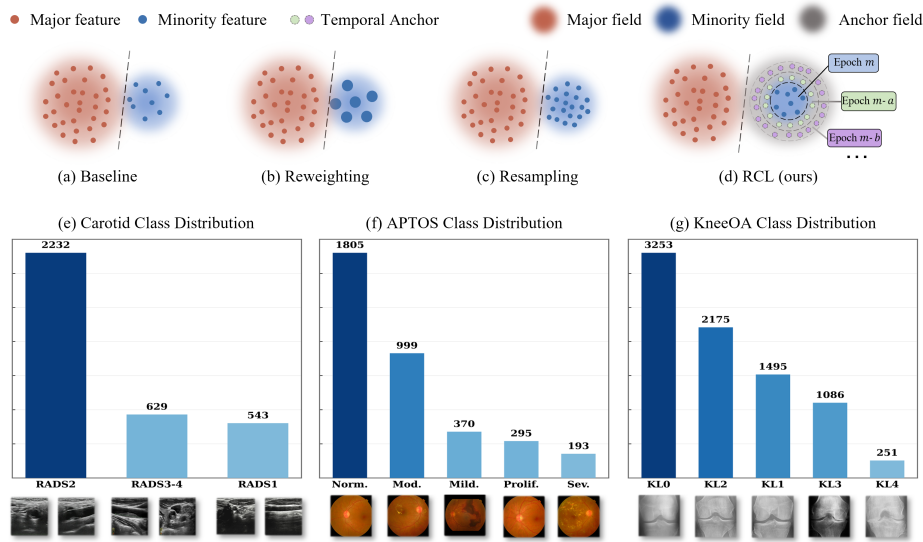


Fig. 1. Comparison of different imbalanced classification strategies and data distributions. (a): Data imbalance situation; (b-c): Data reweighting and resampling; (d): our anchor-based RCL; (e-g): The class distribution of three datasets

To address this class-imbalanced challenge, various strategies have been proposed in recent years. Menon et al. [8] introduced a logit adjustment strategy that incorporates class-prior information into logits to correct prediction bias under class-imbalanced distributions, providing a theoretically grounded reweighting approach. To improve representation fairness in contrastive learning, Zhu et al. [19] reformulated supervised contrastive learning by balancing gradient contributions across classes to restore a more regular geometric structure in feature space. Beyond loss-level adjustments, Zhao et al. [18] enhanced minority learning by incorporating neighboring tail categories to increase granularity. Wang et al. [14] proposed a mixture-of-experts framework to decouple category-specific knowledge into dedicated experts. In addition, Shao et al. [16] designed

a dual-axis style-based recalibration network to enhance minority feature discrimination through statistical loss constraints, while Zhou et al. [5] adopted hierarchical CNN–Transformer fusion to strengthen global–local feature modeling. For ordinal medical grading tasks, Tang et al. [13] modeled asymmetric Gaussian label distributions to better capture intra-class severity variations. Although these methods improve minority recognition from different perspectives, they primarily operate within the existing feature distribution through reweighting, resampling, or architectural modification (Fig. 1(b, c)). They improve how features are used, but do not explicitly enlarge the support region of tail classes. Consequently, tail-class representations remain spatially compact and vulnerable to head-class encroachment, leading to reduced class separability.

In this study, we propose a novel learning framework named **Recurrent Contrastive Learning (RCL)**, which interprets class-imbalanced representation learning from a field-expansion perspective. As shown in Fig. 1(d), we introduce **Temporal Anchors (TARs)**, which are recurrently derived from historical epoch features and used to occupy the surrounding region of tail-class features in latent space. These anchors expand the support region of tail classes and create a buffer margin against nearby head classes. To realize this mechanism, we maintain a **Temporal Memory Queue (TMQ)** to preserve corpus-level features from previous epochs and reuse them for anchor construction. By integrating TARs into contrastive learning, RCL progressively reshapes the feature geometry toward more balanced inter-class separation. Our contributions are threefold. *First*, we propose RCL, an imbalanced medical image classification framework that enlarges tail-class support regions by injecting recurrently constructed temporal anchors into contrastive learning, mitigating head-class encroachment. *Second*, we introduce TMQ to preserve corpus-level historical embeddings across training phases and use them to construct TARs for tail-class field expansion. *Third*, as shown in Fig. 1(e-g), we validate RCL on one multi-center private carotid dataset and two public datasets, KneeOA [1] and APTOS 2019 [6], where it achieves consistent improvements over strong baselines. Experimental results demonstrate the effectiveness and generalization of the proposed method.

2 Methodology

Fig. 2 provides an overview of RCL. For each input image x , the DINOv3 backbone with LoRA adapters (rank=4) extracts a representation feature, which is used for both classification and projection into the latent space. RCL maintains a TMQ that stores features from previous training phases. We retrieve TARs from TMQ for tail classes and organize them into an **anchor field**. This field expands the latent support region of tail classes and serves as a buffer to enhance inter-class separability within RCL. More details are shown in the text below.

Formally, let $\mathcal{X} = \{(x_i, y_i) | 1 \leq i \leq N\}$ be a class-imbalanced training set with C categories. The network consists of a backbone $\mathbf{F}(\cdot)$, a classifier head $\mathbf{G}(\cdot)$,

and a projection head $\mathbf{P}(\cdot)$, and maintains stage-wise memory queues $\mathcal{Q}^{(m)} = \{(q_i^{(m)}, y_i) | 1 \leq i \leq N\}$, $m = (0, 1, 2, \dots, M)$ as the anchor source.

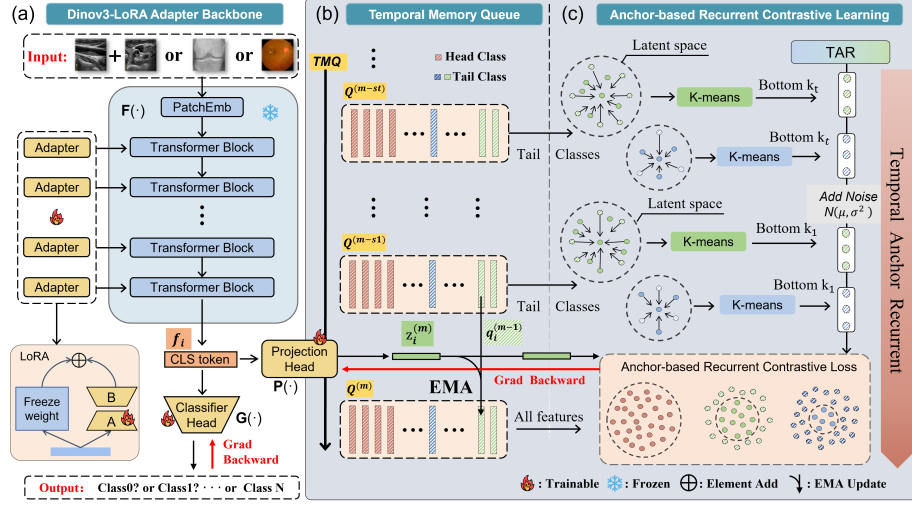


Fig. 2. Overview of the proposed RCL framework.

2.1 DINOv3 with LoRA Adapter Fine-tuning

The foundation model acquires strong representation capability through self-supervised learning on large-scale data, which enables effective adaptation to downstream domains with limited labeled data. As shown in Fig. 2(a), we feed an input image x_i into backbone $\mathbf{F}(\cdot)$ and extract the final-layer 1280-dimensional class token f_i . We employ a classifier head $\mathbf{G}(\cdot)$ that outputs predictions p_i for cross-entropy supervision, and a projection head $\mathbf{P}(\cdot)$ that maps f_i to a 256-dimensional normalized embedding z_i for memory-bank update and contrastive learning. Both $\mathbf{G}(\cdot)$ and $\mathbf{P}(\cdot)$ adopt a Linear–ReLU–Linear design. During training, we freeze the DINOv3 backbone and optimize only the LoRA adapters with $\mathbf{G}(\cdot)$ and $\mathbf{P}(\cdot)$, enabling efficient adaptation with few trainable parameters.

2.2 Temporal Memory Queue

Contrastive learning has shown clear benefits in imbalanced medical image classification tasks [9, 15]. However, many existing methods perform contrastive matching only at the batch-level or within a FIFO queue, which limits access to global and temporally diverse feature references. Inspired by [12, 20], we introduce a TMQ that retains multiple memory banks across training stages and provides diversified global references for contrastive learning. Specifically, in epoch

0, we extract the projection embedding as $z_i^{(0)}$ for each training sample x_i , and use $\{(z_i^{(0)}, y_i) \mid 1 \leq i \leq N\}$ to initialize the memory bank $\mathcal{Q}^{(0)}$, which is defined as:

$$\mathcal{Q}^{(0)} = \{(q_i^{(0)}, y_i) \mid 1 \leq i \leq N\} = \{(z_i^{(0)}, y_i) \mid 1 \leq i \leq N\}. \quad (1)$$

In the subsequent epoch m , as shown in Fig. 2(b), the projection embedding $z_i^{(m)}$ is used to update $q_i^{(m)}$ in $\mathcal{Q}^{(m)}$ via exponential moving average:

$$q_i^{(m)} = (1 - \lambda)q_i^{(m-1)} + \lambda z_i^{(m)}. \quad (2)$$

where $\lambda \in [0, 1)$ is the EMA momentum controlling the contribution of the current embedding. Importantly, we retain the updated bank $\mathcal{Q}^{(m)}$ after each epoch and enqueue it over epochs, forming the $TMQ = \{\mathcal{Q}^{(m)}\}_{m=0}^M$, which enables recurrent access to historical global embeddings.

This design provides two key advantages. First, each epoch-wise bank $\mathcal{Q}^{(m)}$ approximates the global feature distribution at a specific optimization stage, and EMA updates keep stored features aligned with the evolving representation space, enabling richer and more reliable positives/negatives for contrastive learning. Second, by preserving multiple historical banks, TMQ serves as a carrier of temporal knowledge across training phases, which will be leveraged for temporal anchor-based field construction in Sec. 2.3.

2.3 Anchor-based Recurrent Contrastive Learning

The temporal memory queue $TMQ = \{\mathcal{Q}^{(m)}\}_{m=0}^M$ is introduced to provide a global view of the feature distribution, enabling the model to learn dataset-level representations. However, the features stored in $\mathcal{Q}^{(m)}$ still follow a class-imbalanced distribution. To alleviate this inter-class imbalance, we retrieve TARs from historical memory banks to enrich tail-class support and promote balanced learning.

As illustrated in Fig. 2(b, c), for a query $z_i^{(m)}$ with label c at epoch m , we compute the contrastive loss L_{RCL} using features in memory bank queue $\mathcal{Q}$.

In current epoch m , the positive set $\mathcal{Q}_+^{(m)}$ consists of features sharing same class label as z_i , while the negative set $\mathcal{Q}_-^{(m)}$ contains features from other classes:

$$\mathcal{Q}_+^{(m)} = \{q_{i+}^{(m)} \mid y_i = c\}, \quad \mathcal{Q}_-^{(m)} = \{q_{i-}^{(m)} \mid y_i \neq c\}. \quad (3)$$

In earlier epoch $m - st$, we collect candidate tail-class features as $TAR^{(m-st)}$ from $\mathcal{Q}^{(m-st)}$. Specifically, for each tail class in $\mathcal{Q}^{(m-st)}$, we apply K-means selection to intra-class features, compute each feature's similarity to the corresponding cluster center, and select the bottom- k_t features. These selected features are then perturbed with Gaussian noise as in Eq. 4:

$$TAR^{(m-st)} = Bottom-K(k - means(\{(q_i^{(m-st)}, y_i) \mid y_i \in \mathcal{T}\})) + \varepsilon \sim \mathcal{N}(0, 0.1^2). \quad (4)$$

where $\mathcal{T}$ denotes the set of tail classes. Aggregating anchors from multiple previous epochs yields the final anchor set A :

$$A = \bigcup_{t=1}^n TAR^{(m-st)} = \{(a_j, y_j) \mid y_j \in \mathcal{T}\}. \quad (5)$$

where s is the sample step, n is the sample number. Accordingly, the anchor-positive and anchor-negative sets are defined as:

$$A_+ = \{(a_{j+}, y_j) \mid y_j = c\}, \quad A_- = \{(a_{j-}, y_j) \mid y_j \neq c\}.$$

The recurrent contrastive loss is defined as:

$$\mathcal{L}_{RCL} = -\log \frac{\sum_{q \in Q_+^{(m)}} \exp\left(\frac{\text{sim}(z_i^{(m)}, q)}{\beta}\right) + \sum_{a \in A_+} \exp\left(\frac{\text{sim}(z_i^{(m)}, a)}{\beta}\right)}{\sum_{q \in Q^{(m)}} \exp\left(\frac{\text{sim}(z_i^{(m)}, q)}{\beta}\right) + \sum_{a \in A} \exp\left(\frac{\text{sim}(z_i^{(m)}, a)}{\beta}\right)}. \quad (6)$$

The final loss L is defined as $L = L_{ce} + \alpha L_{RCL}$, where α denotes the loss balancing coefficient and β is the contrastive temperature.

3 Experimental Results

Datasets and Implementations. To evaluate RCL, we collected a private multi-center carotid ultrasound dataset with 3,404 paired images (543 RADS1, 2,232 RADS2, and 629 RADS3–4) from 12 centers, graded according to carotid plaque RADS guidelines [10]. The study was approved by the local Institutional Review Board (Anonymized ID: SP2023135 (01)). To verify generalizability, we further evaluated RCL on two public imbalanced datasets: APTOS 2019 [6] for diabetic retinopathy grading (3,662 images) and KneeOA [1] for Kellgren–Lawrence (KL) grading (8,260 images). All images were padded to a square shape and then resized to 256×256 . The Carotid and APTOS 2019 datasets were split at an 8:2 ratio into training and test sets, with four-fold cross-validation performed on the training set and mean test performance reported. The split was conducted at the patient level to prevent data leakage. KneeOA followed the official split protocol. RCL was implemented in PyTorch and trained for 100 epochs on an NVIDIA GeForce RTX 3090 GPU using AdamW (initial learning rate 1×10^{-4}). The EMA momentum λ and contrastive temperature β were set to 0.01 and 0.1, respectively. The loss weight α was linearly decayed from 1 to 0.1. The sample step s was set as 3. Among the n sampled historical banks $\{Q^{(m-s)}, Q^{(m-2s)}, \dots, Q^{(m-ns)}\}$, the bottom selection ratio k_t was linearly decreased from 50% for the most recent bank to 10% for the earliest, so that temporally closer banks contribute more anchor candidates. Five standard augmentation operations from the *timm* were randomly applied during training.

Table 1. Performance comparison and ablation on three datasets. BAcc(%), F1(%), Acc(%), and QWK (three decimals). Best results (**bold**), second-best (underlined).

Method	Carotid				APTOS 2019				KneeOA			
	Bacc↑	F1↑	Acc↑	QWK↑	Bacc↑	F1↑	Acc↑	QWK↑	Bacc↑	F1↑	Acc↑	QWK↑
ResNet50 [4]	74.97	75.73	80.65	0.681	50.66	50.91	72.50	0.700	59.13	58.95	58.03	0.729
Focal* [7]	76.01	77.62	82.55	0.731	52.08	52.38	73.46	0.711	60.16	62.47	61.35	0.744
BCL* [19]	85.02	79.22	81.09	0.763	58.17	50.74	68.13	0.757	<u>69.58</u>	67.98	66.55	0.802
ECL [17]	82.86	83.41	86.95	0.790	64.85	66.71	83.31	0.890	68.24	<u>68.81</u>	67.19	0.787
HiFuse [5]	83.58	83.71	86.51	0.801	66.98	67.29	81.94	0.840	61.59	57.70	63.10	0.725
ADSR [16]	84.60	82.65	85.52	0.786	62.56	64.35	81.53	0.855	65.21	64.51	64.31	0.778
DGN [13]	81.65	83.26	87.10	0.779	67.54	68.51	83.58	0.898	62.56	62.26	<u>68.84</u>	0.813
DINOv3	79.75	82.33	86.14	0.793	<u>68.48</u>	69.20	83.75	<u>0.920</u>	61.10	61.82	63.35	0.611
RCL*	<u>86.92</u>	88.10	90.32	0.853	66.23	68.21	83.72	0.894	66.65	66.65	67.87	<u>0.825</u>
DINOv3+RCL	87.10	<u>87.87</u>	<u>90.06</u>	<u>0.852</u>	69.01	70.79	85.39	0.922	70.19	70.36	70.77	0.843

* using ResNet50 as the backbone.

Quantitative analysis. Balanced Accuracy (Bacc), F1-score (F1), Accuracy (Acc), and Quadratic Weighted Kappa (QWK) were adopted to comprehensively evaluate performance under class-imbalanced settings. Table 1 presents both quantitative comparisons with state-of-the-art class-imbalanced learning approaches [4, 5, 7, 13, 16, 17, 19] and the ablation study. Our framework achieves the best Bacc on all three datasets, reaching 87.10%, 69.01%, and 70.19% on Carotid, APTOS 2019, and KneeOA, respectively, outperforming all competing methods. Compared with vanilla DINOv3 fine-tuning, DINOv3+RCL improves Bacc by 7.35% on Carotid and 9.09% on KneeOA, demonstrating consistent gains across different imbalance scenarios. On the APTOS 2019, the DINOv3-based RCL achieves a QWK of 0.92, surpassing the top-ranked CNN-Ensemble method [6] (QWK = 0.91) in the original challenge, while the ResNet50-based RCL also exceeds the second-ranked Inception-v4 model [6] (QWK = 0.88). When we replace DINOv3 with ResNet50 (the second-to-last row), RCL also yields substantial improvements. On the Carotid and KneeOA datasets, RCL based on ResNet50 even surpasses the performance of DINOv3 fine-tuning, indicating that RCL enhances feature discrimination beyond the backbone alone.

Table 2. Effect of different sample numbers and tail-class settings on the Carotid dataset. Best results (**bold**), second-best (underlined).

n^*	Tail Classes	Bacc↑	F1↑	Acc↑	QWK↑	Tail Classes	Bacc↑	F1↑	Acc↑	QWK↑
3	1	86.86	<u>88.78</u>	91.06	0.862	2	<u>88.57</u>	88.68	<u>90.47</u>	0.860
5		87.72	88.56	<u>90.47</u>	0.858		87.42	87.70	89.59	0.847
7		86.85	87.34	89.59	0.839		88.68	89.18	91.06	<u>0.861</u>

* n : Sample Number.

Table 2 reports the effect of different sample numbers and tail-class settings on the Carotid dataset. Owing to space constraints, we present the hyperparam-

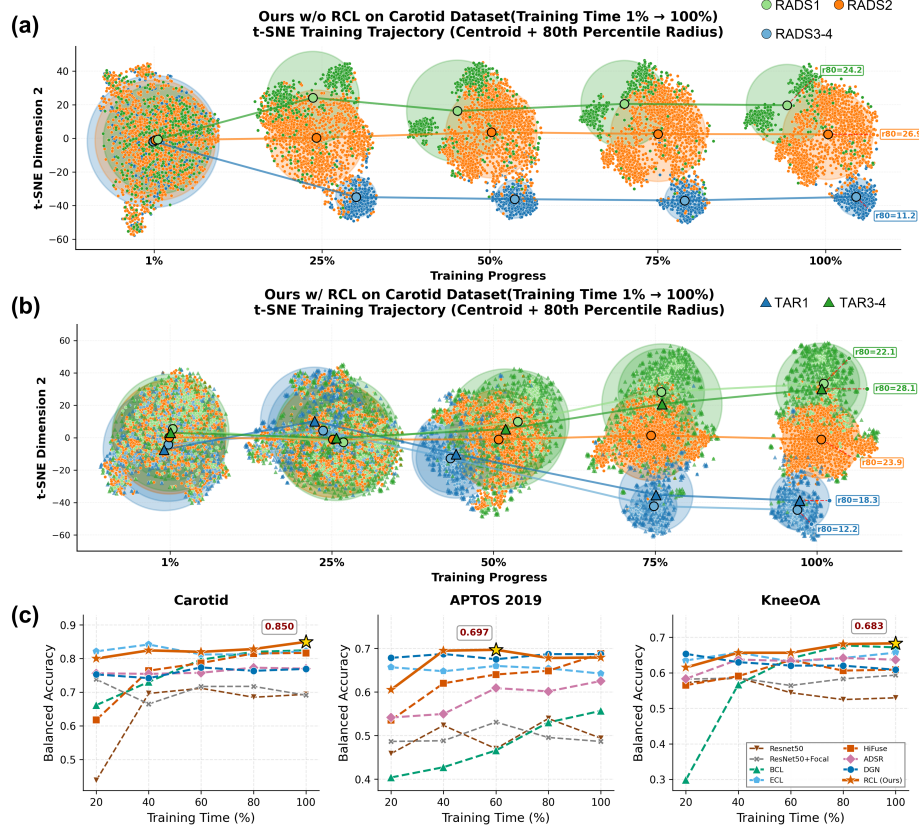


Fig. 3. t-SNE visualization and Bacc evolution at different training stages.

eter configurations that show the most significant influence on the experimental results. Here, Sample Number n denotes the number of historical memory banks sampled to construct anchors. Tail Classes denotes the number of tail classes involved as TARs. The setting with Tail Classes = 2 and $n = 7$ achieves the best overall performance, yielding the highest Bacc (88.68%) and F1 (89.18%), together with the best overall Acc (91.06%). By contrast, the highest QWK is obtained with Tail Classes = 1 and $n = 3$ (0.862). Overall, using two tail classes leads to stronger and more stable performance across different sample epochs, suggesting that richer tail-class information and progressively refined historical features are beneficial for robust representation learning under class imbalance.

Qualitative analysis. We visualize t-SNE embeddings at multiple training phases in Fig. 3(a, b), summarizing each class by its centroid trajectory and 80th-percentile radius. Fig. 3(a) shows the DINOv3 training without RCL, whereas Fig. 3(b) shows with RCL enabled. Compared with the DINOv3-baseline, three phenomena emerge under RCL: (i) Compared with Fig. 3(a), the relative field of head and tail classes becomes more balanced in Fig. 3(b), indicating a fairer distribution of feature support in latent space; (ii) TARs display a larger ra-

dus than current epoch features, suggesting an expanded tail-class boundary induced by the anchor field; and (iii) TAR1 field occupies the intermediate region between RADS2 and RADS1 fields, forming an effective buffer area. This buffer mechanism pushes RADS1 centers away from head-class distributions, thereby enlarging inter-class margins. These observations are all consistent with the field-guided mechanism of RCL. Fig. 3(c) reports validation Balanced Accuracy for checkpoints from different epochs. RCL achieves the best late-epoch performance on Carotid, APTOS 2019, and KneeOA, consistent with the progressively stabilized separation observed in the embedding dynamics.

4 Conclusion

In this work, we propose a novel Recurrent Contrastive Learning for imbalanced medical image classification. By leveraging a Temporal Memory Queue and constructing Temporal Anchors from historical feature states, RCL forms an anchor field to expand the support region of tail classes and mitigate head-class dominance. However, the framework relies on manual hyperparameter tuning. Future work will explore adaptive anchor selection to address this limitation.

Acknowledgments. This study was funded by the Yunnan Major Science and Technology Special Project Program (No. 202402AA310052), the Yunnan Key Research and Development Program(202503AP140014), and the Guangxi Province Science Program (No.2024AB17023).

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

References

1. Pingjun Chen. Knee osteoarthritis severity grading dataset. *Mendeley Data*, 1(10.17632):30784984, 2018.
2. Yuanyuan Chen, Xiaoqing Guo, Yongsheng Pan, Yong Xia, and Yixuan Yuan. Dynamic feature splicing for few-shot rare disease diagnosis. *Medical Image Analysis*, 90:102959, 2023.
3. Damien Dablain, Bartosz Krawczyk, and Nitesh V Chawla. Deepsmote: Fusing deep learning and smote for imbalanced data. *IEEE Transactions on Neural Networks and Learning Systems*, 34(9):6390–6404, 2022.
4. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.
5. Xiangzuo Huo, Gang Sun, Shengwei Tian, Yan Wang, Long Yu, Jun Long, Wengdong Zhang, and Aolun Li. Hifuse: Hierarchical multi-scale feature fusion network for medical image classification. *Biomedical Signal Processing and Control*, 87:105534, 2024.
6. M Karthick and D Sohier. Aptos 2019 blindness detection dataset. <https://www.kaggle.com/competitions/aptos2019-blindness-detection>, 2019.
7. Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *CVPR*, pages 2980–2988, 2017.

8. Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. Long-tail learning via logit adjustment. *arXiv preprint arXiv:2007.07314*, 2020.
9. Sumin Roh, Harim Kim, Ho Yun Lee, and Il Yong Chun. Equilibrium contrastive learning for imbalanced image classification. *arXiv preprint arXiv:2602.09506*, 2026.
10. Luca Saba, Riccardo Cau, Alessandro Murgia, Andrew N Nicolaides, Max Wintermark, Mauricio Castillo, Daniel Staub, Stavros K Kakkos, Qi Yang, Kosmas I Paraskevas, et al. Carotid plaque-rads: a novel stroke risk classification system. *Cardiovascular Imaging*, 17(1):62–75, 2024.
11. Mabrouka Salmi, Dalia Atif, Diego Oliva, Ajith Abraham, and Sebastian Ventura. Handling imbalanced medical datasets: review of a decade of research. *Artificial Intelligence Review*, 57(10):273, 2024.
12. Hao Tang, Zhiqing Guo, Liejun Wang, and Chao Liu. Similarity memory prior is all you need for medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23009–23018, 2025.
13. Wenqiang Tang and Zhouwang Yang. Disease-grading networks with asymmetric gaussian distribution for medical imaging. *IEEE Transactions on Medical Imaging*, 2025.
14. Mengzhu Wang, Jianlong Yuan, and Zhibin Wang. Mixture-of-experts learner for single long-tailed domain generalization. In *ACM MM*, pages 290–299, 2023.
15. Zhixiong Yang, Junwen Pan, Yanzhan Yang, Xiaozhou Shi, Hong-Yu Zhou, Zhicheng Zhang, and Cheng Bian. Proco: Prototype-aware contrastive learning for long-tailed medical image classification. In *MICCAI*, pages 173–182. Springer, 2022.
16. Xiaoqing Zhang, Zunjie Xiao, Jingzhe Ma, Xiao Wu, Jilu Zhao, Shuai Zhang, Runzhi Li, Yi Pan, and Jiang Liu. Adaptive dual-axis style-based recalibration network with class-wise statistics loss for imbalanced medical image classification. *IEEE Transactions on Image Processing*, 2025.
17. Yilan Zhang, Jianqi Chen, Ke Wang, and Fengying Xie. Ecl: Class-enhancement contrastive learning for long-tailed skin lesion classification. In *MICCAI*, pages 244–254. Springer, 2023.
18. Shizhen Zhao, Xin Wen, Jiahui Liu, Chuofan Ma, Chunfeng Yuan, and Xiaojuan Qi. Learning from neighbors: Category extrapolation for long-tail learning. In *CVPR*, pages 30483–30492, 2025.
19. Jianggang Zhu, Zheng Wang, Jingjing Chen, Yi-Ping Phoebe Chen, and Yu-Gang Jiang. Balanced contrastive learning for long-tailed visual recognition. In *CVPR*, pages 6908–6917, 2022.
20. Zhiyuan Zhu, Jian Wang, Yong Jiang, Tong Han, Yuhao Huang, Ang Zhang, Kaiwen Yang, Mingyuan Luo, Zhe Liu, Yaofei Duan, et al. Hierarchical corpus-view-category refinement for carotid plaque risk grading in ultrasound. In *MICCAI*, pages 255–264. Springer, 2025.