Conservative-Radical Complementary Learning for Class-incremental Medical Image Analysis with Pre-trained Foundation Models

Xinyao Wu*[1], Zhe Xu*(✉)[1], Donghuan Lu[2], Jinghan Sun[2], Hong Liu[3], Sadia Shakil[1], Jiawei Ma[4,5], Yefeng Zheng[6], and Raymond Kai-yu Tong(✉)[1]

[1] Dept. of Biomedical Engineering, The Chinese University of Hong Kong, Hong Kong, China
[2] Tencent Jarvis Lab, Shenzhen, China
[3] National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, China
[4] Dept. of Computer Science, City University of Hong Kong, Hong Kong, China
[5] Institute of Digital Medicine, City University of Hong Kong, Hong Kong, China
[6] Westlake University, Hangzhou, China
jackxz@link.cuhk.edu.hk; kytong@cuhk.edu.hk

Abstract. Class-incremental learning (CIL) in medical image-guided diagnosis requires models to retain diagnostic expertise on historical disease classes while adapting to newly emerging categories—a critical challenge for scalable clinical deployment. While pretrained foundation models (PFMs) have revolutionized CIL in the general domain by enabling generalized feature transfer, their potential remains underexplored in medical imaging, where domain-specific adaptations are critical yet challenging due to anatomical complexity and data heterogeneity. To address this gap, we first benchmark recent PFM-based CIL methods in the medical domain and further propose Conservative-Radical Complementary Learning (CRCL), a novel framework inspired by the complementary learning systems in the human brain. CRCL integrates two specialized learners built upon PFMs: (i) a neocortex-like **conservative learner**, which safeguards accumulated diagnostic knowledge through stability-oriented parameter updates, and (ii) a hippocampus-like **radical learner**, which rapidly adapts to new classes via dynamic and task-specific plasticity-oriented optimization. Specifically, dual-learner feature and cross-classification alignment mechanisms harmonize their complementary strengths, reconciling inter-task decision boundaries to mitigate catastrophic forgetting. To ensure long-term knowledge retention while enabling adaptation, a consolidation process progressively transfers learned representations from the radical to the conservative learner. During task-agnostic inference, CRCL integrates outputs from both learners for robust final predictions. Comprehensive experiments on four medical imaging datasets show CRCL's superiority over state-of-the-art methods.

Keywords: Class-incremental · Foundation Model · Disease Diagnosis.

* Equal contribution

1 Introduction

In medical image-guided disease diagnosis, continually updating diagnostic models to adapt to evolving healthcare data is critical, particularly as novel disease categories emerge [25, 31]. However, traditional deep learning paradigms often suffer from catastrophic forgetting when learning new tasks sequentially, significantly hindering their scalability and long-term effectiveness. Ideally, diagnostic models should effectively recognize both historical and newly introduced disease categories without access to explicit task identity during inference. Such a clinical need is typically framed as class-incremental learning (CIL), which necessitates balancing the trade-off between stability (retaining existing diagnostic expertise) and plasticity (integrating novel disease patterns).

Related Work. Traditional CIL methods fall into three categories: replay, which retains or synthesizes historical data [5, 8, 17]; regularization, which penalizes adversarial parameter updates to preserve prior knowledge [13, 11, 22]; and adaptive architecture, which dynamically expands or conditionally activates model components [1, 29]. However, these methods generally assume models are trained from scratch and require extensive parameter tuning, empirically falling short of practical deployment requirements. Recent advances in pretrained foundation models (PFMs) have revitalized CIL research by leveraging their generalized feature representations [24, 33, 20, 7, 32]. Prompt-based methods [24, 23, 18] adapt PFMs via dynamic prompt pools to capture task-specific features, though they remain constrained to Transformer architectures and necessitate prompt pool expansion for new tasks. Beyond prompt-based methods, SLCA [30] introduces dual learning rates for backbone and classifier tuning, coupled with Gaussian-based classifier rectification. LAE [7] inherits SLCA's learning rate calibration but further proposes model merging to consolidate knowledge. ADAM [33] demonstrates that a prototypical classifier serves as a robust baseline and merges embeddings from a frozen PFM and a first-session adapted downstream model for subsequent classification. EASE [34] and MOS [19] enhance feature representation by merging outputs from multiple task-specific adapters, while SSIAT [20] addresses feature drift by estimating class prototype shifts across tasks and enforcing unimodal distribution assumptions for replay-based unified training. Despite progress in natural image domains, PFM-based CIL remains underexplored in medical diagnosis, where domain-specific PFMs are scarce. To bridge this gap, we harness powerful PFMs from the general domain to take the initiative to benchmark recent PFM-based CIL methods on clinical datasets and advance replay-free CIL strategies for evolving disease diagnosis.

To address the stability-plasticity dilemma in PFM-based medical CIL, we present Conservative-Radical Complementary Learning (CRCL), a novel framework inspired by the complementary learning systems in the human brain [12] (as shown in Fig. 1). CRCL consists of two specialized learners built upon PFMs: (i) a neocortex-like conservative learner, which safeguards accumulated diagnostic knowledge through stability-oriented parameter updates, and (ii) a hippocampus-like radical learner, which rapidly adapts to new disease categories via dynamic and plasticity-oriented optimization. To mitigate catastrophic for-

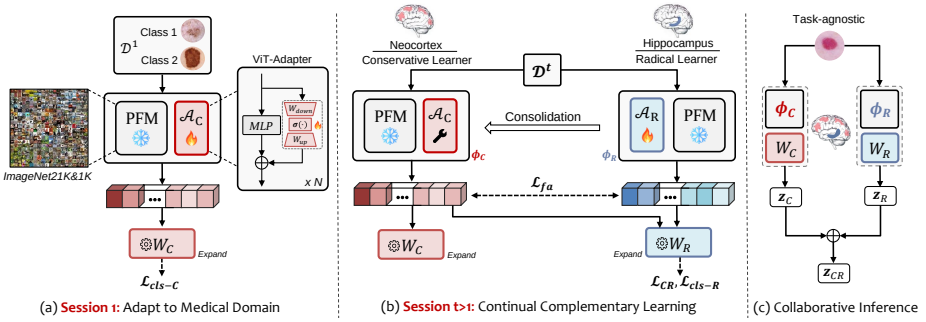

Fig. 1. Overview of our CRCL framework. At session $t = 1$, the PFM adapts to the medical domain by adapter tuning. For $t > 1$, the radical learner adapts to new classes while the conservative learner retains past knowledge and gradually consolidates new knowledge. During inference, their outputs are fused for robust predictions.

getting while ensuring flexible and efficient knowledge integration, we introduce dual-learner feature and cross-classification alignment mechanisms, harmonizing representations across tasks to reconcile inter-task decision boundaries. Besides, a consolidation process is incorporated, where knowledge acquired by the radical learner is progressively transferred to the conservative learner. This consolidation mechanism stabilizes long-term representations while allowing flexible adaptation to new tasks. During inference, CRCL leverages second-order statistics to enhance linear separability and integrates the complementary outputs of both learners for robust and task-agnostic predictions. Extensive experiments on four medical imaging benchmarks demonstrate the superiority of CRCL, significantly outperforming state-of-the-art (SOTA) PFM-based CIL approaches.

2 Method

2.1 Preliminary

Problem Definition. CIL is defined as training a model on a sequential data stream where new classes are incrementally introduced. Specifically, the training set for session t can be written as $\mathcal{D}^t = \{(x_{i,t}, y_{i,t})\}_{i=1}^{n_t}$, where $t \in \{1, 2, \ldots, T\}$ denotes the session index among T total sessions, and each session contains n_t instances. We assume that each session t introduces a unique set of classes Y_t, with no overlap between sessions: $Y_t \cap Y_{t'} = \varnothing$ for $t \neq t'$. The goal is for the model to perform well on a test set encompassing all classes introduced up to session t, i.e., $\mathcal{Y}_t = Y_1 \cup \cdots Y_t$.

PFM-based CIL. We aim to build a model $f(\mathbf{x}) : X \rightarrow \mathcal{Y}_t$ that can learn new classes incrementally without forgetting previously learned ones. Following prior replay-free PFM-based CIL studies [20, 30, 33, 34], we consider a pre-trained Vision Transformer (ViT) model is available for initializing $f(\mathbf{x})$. We then decompose it into a feature extraction backbone $\mathcal{F}_{\theta_{\text{bne}}}$ and a linear classification layer $f_{\theta_{\text{cls}}}$. The backbone $\mathcal{F}_{\theta_{\text{bne}}}$ extracts features from the input images,

serving as a feature embedding function $\phi(\cdot)$ mapping $\mathbb{R}^D \rightarrow \mathbb{R}^d$, while the classifier layer $f_{\theta_{\text{cls}}}$, represented by a weight matrix $W \in \mathbb{R}^{d\times|Y_t|}$, projects the feature embeddings to class predictions. The model can then be formularized as $f(\mathbf{x}) = W^\top\phi(\mathbf{x})$, where $W = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_j]$, with $\mathbf{w}_j$ denoting the classifier weights for class j. We treat the embedded `[CLS]` token as $\phi(\mathbf{x})$ for ViT.

Tuning with Adapter. Recent studies have explored various parameter-efficient tuning (PET) methods to adapt PFMs for downstream tasks, including SSF [14], VPT [9], and adapters [6]. Yet, our empirical analysis reveals that SSF and VPT often suffer from overfitting to the current distribution, yielding unstable results on complex medical data. Consequently, we advocate adapters as the PET strategy in our PFM-based CIL framework. Specifically, adapters are lightweight modules that consist of a down-projection $W_{\text{down}} \in \mathbb{R}^{k\times\hat{k}}$, a ReLU activation function σ, and an up-projection $W_{up} \in \mathbb{R}^{\hat{k}\times k}$, forming a bottleneck structure. Following [20, 34], we augment the ViT's multilayer perceptron (MLP) layers with adapters, as depicted in Fig. 1(a) with projected dimension $\hat{k}$ equal to 64 [6]. Let x_{in} be the input of the MLP layer, the output of the adapter-equipped MLP is: $x_{out} = MLP(x_{in}) + \sigma\left(x_{in} * W_{\text{down}}\right) * W_{up}$, where $*$ denotes matrix multiplication. During training, the PFM remains frozen, while only the adapters and classification heads are updated. Thus, the optimizable parameters can be denoted as: $\Theta = \theta_{W_{\text{down}}} \cup \theta_{W_{\text{up}}} \cup \theta_W$. This adapter-based ViT fine-tuning serves as the foundation of our replay-free PFM-based CIL model, offering a flexible and parameter-efficient approach for adapting Transformer models to new tasks.

2.2 Conservative-Radical Complementary Learning

Balancing plasticity and stability is a fundamental challenge in PFM-based CIL, further exacerbated when transferring generalized representations from general-domain PFMs to the medical domain. To address this, our CRCL (Fig. 1) features three key processes: (i) lightweight adaptation to initialize the conservative learner, enabling efficient PFM transfer while preserving generalization, (ii) radical learner integration, inspired by the hippocampus-neocortex system, to achieve continual complementary learning for balancing knowledge retention and adaptation, and (iii) collaborative inference for robust task-agnostic predictions.

Model Adaptation for Conservative Learner Initialization. As illustrated in Fig. 1(a), we aim to bridge the gap between the ImageNet21K&1K-pretrained PFM and downstream medical datasets in the first incremental session ($t = 1$). To achieve this, we efficiently adapt the PFM to the medical domain with the lightweight ViT-Adapters. Specifically, we freeze the PFM backbone to preserve its general knowledge and only update the adapters $\mathcal{A}_C$ along with the classifier W_C. Notably, while the classifier W_C expands to incorporate new classes and updates during training, it leverages class prototypes as the imprinted weights [16] rather than a conventional trainable layer. The *gear* icon in Fig. 1 visually represents this distinction, indicating that W_C undergoes structured updates rather than standard parameter learning. In the first session, we do not impose additional parameter constraints to ensure sufficient

adaptation. Our training objective is to minimize the cross-entropy (CE) loss: $\mathcal{L}_{\text{cls-C}} = \frac{1}{N_b}\sum_{i=1}^{N_b} \text{CE}\left(W_C^\top \phi_C\left(x_i\right), y_i\right)$, where N_b is the batch size, $\phi_C(x_i)$ represents the adapted feature embeddings for input x_i, and y_i is the corresponding ground-truth label.

Continual Complementary Learning. Fig. 1(b) illustrates the continual complementary learning paradigm (when $t > 1$) that balances stability and plasticity through the interaction of radical and conservative learners. The radical learner, akin to the hippocampus, rapidly encodes new information by optimizing all parameters of its dedicated adapters, $\mathcal{A}_R$, via backpropagation in each training session. Meanwhile, the conservative learner, reminiscent of the neocortex, preserves consolidated knowledge while gradually integrating new patterns. This is achieved by updating its adapters $\mathcal{A}_C$ using an exponential moving average (EMA) of the radical learner's updates (represented by a *wrench* icon), ensuring stable and incremental assimilation. Formally, in the the t-th ($t > 1$) session, this consolidation is defined as: $\theta_{\mathcal{A}_{C,t}} = \alpha\theta_{\mathcal{A}_{C,t-1}} + (1-\alpha)\theta_{\mathcal{A}_{R,t}}$, where α is the EMA decay rate and empirically set to 0.99 [26]. To learn new tasks, the radical learner is first optimized using a standard cross-entropy classification loss: $\mathcal{L}_{\text{cls-R}} = \frac{1}{N_b}\sum_{i=1}^{N_b} \text{CE}\left(W_R^\top \phi_R(x_i), y_i\right)$, where W_R represents the structurally updated classification weights (analogous to W_C), and $\phi_R(x_i)$ denotes the radical learner's feature embedding for input x_i. At task t, $W_R \in \mathbb{R}^{d\times|Y_{t-1}|}$ expands to $W_R \in \mathbb{R}^{d\times|Y_t|}$ with the class prototypes of newly introduced classes. To ensure compatibility between the learners' representations and mitigate semantic drift, we enforce feature alignment by minimizing:

$$\mathcal{L}_{fa} = \frac{1}{N_b}\sum_{i=1}^{N_b}\left(1 - \text{sim}\left(\phi_C(x_i), \phi_R(x_i)\right)\right), \tag{1}$$

where sim$(\cdot)$ utilizes cosine similarity. This regularizes the two learners to maintain stable feature spaces, mitigating catastrophic forgetting. Beyond feature alignment, the cross-classification regularization loss $\mathcal{L}_{CR}$ further encourages the conservative learner's features to remain compatible with the radical classifier, reinforcing consistency in representation learning of the radical learner and mitigating severe semantic shifts that could lead to biased decision boundaries:

$$\mathcal{L}_{CR} = \frac{1}{N_b}\sum_{i=1}^{N_b} \text{CE}\left(W_R^\top \phi_C(x_i)\right), y_i). \tag{2}$$

Overall, the final loss for the radical learner is: $\mathcal{L}_R = \mathcal{L}_{\text{cls-R}} + \lambda\mathcal{L}_{fa} + \mathcal{L}_{CR}$, where λ is set to 50 to ensure comparable scales for the three losses. This combined objective enables the radical learner to effectively acquire new knowledge while maintaining alignment with the conservative learner, ensuring a stable yet adaptable continual learning process.

Collaborative Inference. The nearest class mean (NCM) classifier has shown promise in few-shot classification. However, [15] reveals that raw class prototypes often exhibit high correlations between classes, leading to poorly calibrated cosine similarities. In this regard, we adapt the strategy that projects features

into higher-dimensional spaces and utilizes second-order statistics to decorrelate prototypes and enhance class separability (refer to [15] for further theoretical details). Briefly, given a test sample $\mathbf{x}$ and the embedded feature $\phi_l(\mathbf{x})$ from the learner l, a frozen random projection matrix $W_{\text{rand}} \in \mathbb{R}^{d\times M}$ is employed with each column sampled from $\mathcal{N}(0,1)$. This projection yields a length-M ($M > d$) feature vector $h_l = \sigma\left(\phi_l(\mathbf{x})^\top W_{\text{rand}}\right) \in \mathbb{R}^M$, where σ is the non-linear activation function. We then compute the Gram matrix G for the projected features, aggregating information across tasks: $G = \sum_t \sum_{n=1}^{N_t} h_{l,t,n} h_{l,t,n}^\top \in \mathbb{R}^{M\times M}$. This matrix is incrementally computed from feature vectors across sessions 1 to t, capturing second-order relationships and the variance structure within the feature space. Another matrix C_p consists of the accumulated sum of projected features corresponding to the same class labels, defined by $C_p = \sum_t \sum_{n=1}^{N_t} h_{l,t,n} y_{t,n}^\top \in \mathbb{R}^{M\times|Y_t|}$. The predicted logits for learner l are then computed as:

$$z_l = h_l W_l = h_l (G + \beta I)^{-1} C_p \in \mathbb{R}^{|Y_t|}, \tag{3}$$

with β the ridge regularization parameter selected by cross-validation-based optimization, and I the identity matrix. As such, W_l can be interpreted as decorrelated class prototypes [15]. This strategy facilitates that similarity calculations reflect both feature alignment and the underlying structure of the feature space. During collaborative inference, logits from the conservative learner and the radical learner are summed for final logits $z_{CR} = z_C + z_R$, effectively mimicking the brain's integrative decision-making process, where stable long-term knowledge is harmonized with rapidly acquired insights. The final prediction is determined by the index of the maximum logit: $y^* = \underset{y}{\text{argmax}}(z_{CR})$.

3 Experiments

Datasets. We benchmark various methods on four medical imaging datasets, as summarized in Table 1. Colon [10] contains H&E stained histopathology images of human colorectal cancer and healthy tissue, while Blood [2] includes normal peripheral blood cell images from blood smears. Skin8 [21], originating from the ISIC challenge for skin lesion classification from dermatoscopy images, exhibits significant class imbalance. MedMNISTv2 [28] is a standardized biomedical image classification benchmark comprising 12 2D datasets and 6 3D datasets, designed for tasks including multi-class, multi-label, and ordinal regression. Following [31], we focus on a subset of four 2D multi-class classification datasets, BloodMNIST, OrganAMNIST, PathMNIST, and TissueMNIST, dubbed MedMNIST-Sub. We use the same data splits as outlined in [4, 31]. During training, each image is resized to 224×224 pixels.

Implementation and Evaluation Protocol. The framework is implemented on PyTorch using NVIDIA A100 GPUs. Following [34, 33], we adopt the representative ViT-B/16-IN1K [3] as our PFM, which is pre-trained on ImageNet-21K and fine-tuned on ImageNet-1K. We conduct all experiments with a batch size of 48 for 20 epochs in the initial adaptation and 15 epochs in subsequent sessions,

Table 1. Overview of medical image classification datasets.

Dataset	Classes	Training set	Test set	Task Num.	Size
Colon [10]	9	70,000	30,000	4	224×224
Blood [2]	8	11,965	5,127	4	360×363
Skin8 [21]	8	3,555	705	4	$[600, 1024]$
MedMNIST-Sub [28]	36	302,002	75,659	4	28×28

Table 2. Performance on four medical datasets. "†" indicates the necessity of replaying prior data. The best and second-best results are **bolded** and underlined, respectively.

Method	Colon		Blood		Skin8		MedMNIST-Sub	
	Acc_{Avg}	Acc_{Last}	Acc_{Avg}	Acc_{Last}	Acc_{Avg}	Acc_{Last}	Acc_{Avg}	Acc_{Last}
Joint Training	-	99.98	-	99.61	-	67.73	-	73.61
Finetune	38.65	10.43	37.47	14.54	39.62	17.87	29.18	5.66
FOSTER† [22]	90.81	86.69	87.70	<u>90.24</u>	55.00	39.01	58.24	31.46
iCaRL† [17]	80.50	78.13	81.41	81.57	58.45	39.57	68.23	38.44
DER† [27]	85.92	85.63	86.48	87.10	48.57	23.55	69.97	42.11
ACL† [31]	85.57	85.16	76.09	82.33	58.30	<u>57.66</u>	<u>81.33</u>	<u>65.40</u>
L2P [24]	70.13	49.91	86.46	76.15	55.39	35.89	56.24	28.96
DualPrompt [23]	79.47	63.36	76.62	66.27	52.32	27.38	54.92	26.31
CodaPrompt [18]	81.96	66.70	72.64	61.57	50.17	32.48	61.12	28.65
LAE [7]	71.16	49.68	55.51	33.94	49.52	24.40	48.46	18.39
SimpleCIL [33]	90.10	85.41	83.85	79.79	56.61	38.30	68.07	50.63
ADAM-Adapter [33]	86.01	78.00	88.09	83.52	59.82	41.84	70.97	53.11
SLCA [30]	86.66	76.73	<u>90.23</u>	82.29	58.91	40.71	56.39	44.42
EASE [34]	89.62	82.48	68.85	67.60	60.40	40.43	65.11	39.26
SSIAT [20]	75.36	66.34	86.00	84.63	60.46	41.99	59.43	25.79
MOS [19]	<u>91.46</u>	<u>87.60</u>	89.27	87.39	<u>63.80</u>	45.82	74.59	51.80
CRCL (ours)	**98.16**	**97.58**	**97.13**	**96.04**	**73.76**	**61.32**	**84.70**	**66.46**

using SGD with momentum and a cosine-annealed learning rate starting at 0.01. We apply random flipping and rotating for weak data augmentation. Following [20, 30, 33, 34], we report the last session accuracy Acc_{Last} and the average accuracy across all incremental sessions t, formulated as: $Acc_{\text{Avg}} = \frac{1}{T}\sum_{t=1}^{T} Acc_t$. All other methods use the same seed and PFM for a fair comparison. The implementation is available at `https://github.com/CUHK-BMEAI/CRCL`.

Comparison with SOTA Methods. Table 3 presents the results of various methods across four medical datasets. Joint training serves as the upper bound, representing the ideal scenario with simultaneous access to all data, while continual PFM finetuning acts as the lower bound, highlighting severe forgetting. The compared methods can be categorized into traditional CIL (e.g., FOSTER [22], iCaRL [17] and DER [27]) and PFM-based CIL (e.g., ACL [31], CodaPrompt [18], ADAM-Adapter [33], EASE [34], SSIAT [20] and MOS [19]). While traditional top-performing CIL methods achieve competitive results on Colon and Blood, their reliance on computationally demanding tuning and data replay raises practical concerns. PFM-based CIL methods utilize PET to enable adaptability while reducing computational overhead. Yet, existing methods exhibit notable performance variability with suboptimal results. As shown, ACL [31] achieves competitive performance but depends on replaying prior data. ADAM-Adapter [33], leveraging prototypes and first-session adaptation, enhances efficiency but lacks continual adaptation. SSIAT [20], incorporating semantic shift estimation, improves feature alignment but struggles with Colon, Skin8 and

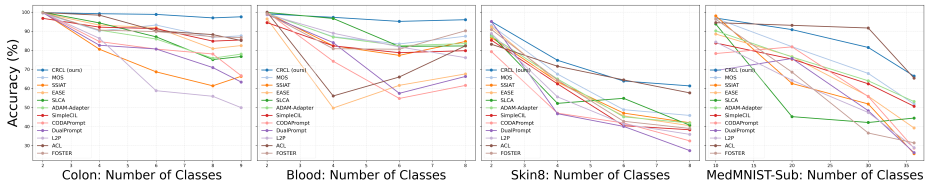

Fig. 2. The performance curves across learning sessions on four medical datasets.

Table 3. Ablation analysis on Skin8.

Setting	Exclusion	Acc_{Avg} (%)	Acc_{Last} (%)
CRCL	None	73.76	**61.32**
Abla-1	Model Adaptation	70.96	59.69
Abla-2	Feature Alignment $\mathcal{L}_{fa}$	60.35	40.69
Abla-3	Cross-classification Reg. $\mathcal{L}_{CR}$	71.06	60.50
Abla-4	$\mathcal{L}_{fa}$ and $\mathcal{L}_{CR}$	58.29	38.69
Abla-5	Collaborative Inference	**73.95**	60.51

MedMNIST. The recent MOS [19] utilizes adapter merging and a self-refined adapter retrieval mechanism, yet it still yields suboptimal performance. Meanwhile, prompt-based methods (L2P [24], DualPrompt [23] and CodaPrompt [18]) show overall weaker performance, likely due to the limited expressiveness of soft prompts in handling complex distributions of medical images. We also present the incremental performance trend of various methods in Fig. 2, observing that most methods struggle with maintaining stability across sessions, with significant performance drops in later stages. Encouragingly, CRCL consistently shows superior adaptability to evolving medical data, making it an appealing replay-free solution for class-incremental medical image analysis.

Ablation Analysis. To better understand CRCL, we exclude individual modules for an ablation study (Table 3) on Skin8. Excluding model adaptation (Abla-1) results in a notable performance drop, indicating that while general-domain PFMs provide generalizable features, domain adaptation remains essential for medical applications. Removing $\mathcal{L}_{fa}$ (Abla-2) leads to a sharp degradation, highlighting its key role in aligning representations across tasks and facilitating knowledge transfer. Excluding $\mathcal{L}_{CR}$ (Abla-3) causes a smaller decline, suggesting that it supports consistency and mitigates forgetting but is less critical than $\mathcal{L}_{fa}$. Removing both (Abla-4) results in the most severe degradation, underscoring their complementary roles. Omitting collaborative inference (Abla-5) slightly improves average accuracy but reduces final performance, indicating that its absence may cause overfitting to earlier tasks and compromise overall robustness. These results verify that each CRCL component contributes to improving performance, collectively ensuring an effective balance between stability and plasticity.

4 Conclusion

In this paper, we introduced the Conservative-Radical Complementary Learning (CRCL) framework, a replay-free approach for class-incremental learning in medical imaging. CRCL integrates a stability-focused conservative learner

with a plasticity-driven radical learner, effectively harmonized through feature and cross-classification alignment, while a progressive consolidation mechanism transfers knowledge from the radical to the conservative learner to ensure long-term retention. Extensive benchmarking on four medical datasets demonstrated that CRCL outperformed recent top-performing methods. By bridging general-domain pretrained foundation models with clinical demands, CRCL advances scalable and lifelong diagnostic systems that adapt to evolving disease diversity.

Acknowledgments. This research was partly supported by Research Impact Fund (R5039-23F) from Research Grants Council of Hong Kong.

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

References

1. Abati, D., Tomczak, J., Blankevoort, T., Calderara, S., Cucchiara, R., Bejnordi, B.E.: Conditional channel gated networks for task-aware continual learning. In: CVPR. pp. 3931–3940 (2020)
2. Acevedo, A., Alférez, S., Merino, A., Puigví, L., Rodellar, J.: Recognition of peripheral blood cell images using convolutional neural networks. Computer Methods and Programs in Biomedicine **180**, 105020 (2019)
3. Alexey, D.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv: 2010.11929 (2020)
4. Bayasi, N., Hamarneh, G., Garbi, R.: Continual-zoo: Leveraging zoo models for continual classification of medical images. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 4128–4138 (2024)
5. Castro, F.M., Marín-Jiménez, M.J., Guil, N., Schmid, C., Alahari, K.: End-to-end incremental learning. In: Proceedings of the European Conference on Computer Vision. pp. 233–248 (2018)
6. Chen, S., Ge, C., Tong, Z., Wang, J., Song, Y., Wang, Y., Luo, P.: Adaptformer: Adapting vision transformers for scalable visual recognition. In: Advances in Neural Information Processing Systems (2022)
7. Gao, Q., Zhao, C., Sun, Y., Xi, T., Zhang, G., Ghanem, B., Zhang, J.: A unified continual learning framework with general parameter-efficient tuning. In: IEEE/CVF International Conference on Computer Vision. pp. 11483–11493 (2023)
8. Hou, S., Pan, X., Loy, C.C., Wang, Z., Lin, D.: Learning a unified classifier incrementally via rebalancing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 831–839 (2019)
9. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: European Conference on Computer Vision (2022)
10. Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.A., Gaiser, T., Marx, A., Valous, N.A., Ferber, D., et al.: Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. PLoS Medicine **16**(1), e1002730 (2019)
11. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., et al.: Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences **114**(13), 3521–3526 (2017)

12. Kumaran, D., Hassabis, D., McClelland, J.L.: What learning systems do intelligent agents need? Complementary learning systems theory updated. Trends in Cognitive Sciences **20**(7), 512–534 (2016)
13. Li, Z., Hoiem, D.: Learning without forgetting. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(12), 2935–2947 (2017)
14. Lian, D., Zhou, D., Feng, J., Wang, X.: Scaling & shifting your features: A new baseline for efficient model tuning. In: Advances in Neural Information Processing Systems (2022)
15. McDonnell, M.D., Gong, D., Parvaneh, A., Abbasnejad, E., van den Hengel, A.: RANPAC: Random projections and pre-trained models for continual learning. Advances in Neural Information Processing Systems **36** (2024)
16. Qi, H., Brown, M., Lowe, D.G.: Low-shot learning with imprinted weights. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
17. Rebuffi, S.A., Kolesnikov, A., Sperl, G., Lampert, C.H.: iCARL: Incremental classifier and representation learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2001–2010 (2017)
18. Smith, J., Karlinsky, L., Gutta, V., Cascante-Bonilla, P., Kim, D., Arbelle, A., Panda, R., Feris, R., Kira, Z.: Coda-prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning. IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 11909–11919 (2023)
19. Sun, H.L., Zhou, D.W., Zhao, H., Gan, L., Zhan, D.C., Ye, H.J.: MOS: Model Surgery for Pre-Trained Model-Based Class-Incremental Learning. In: Proceedings of the AAAI Conference on Artificial Intelligence (2025)
20. Tan, Y., Zhou, Q., Xiang, X., Wang, K., Wu, Y., Li, Y.: Semantically-Shifted Incremental Adapter-Tuning is A Continual ViTransformer. IEEE/CVF Conference on Computer Vision and Pattern Recognition (2024)
21. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data **5**(1), 1–9 (2018)
22. Wang, F., Zhou, D., Ye, H., Foster, D.Z.: Feature boosting and compression for class-incremental learning. ECCV (2022)
23. Wang, Z., Zhang, Z., Ebrahimi, S., Ren, X., Su, G., Perot, V., Dy, J., Pfister, T.: DualPrompt: Complementary prompting for rehearsal-free continual learning. European Conference on Computer Vision pp. 631–648 (2022)
24. Wang, Z., Zhang, Z., Lee, C., Zhang, H., Sun, R., Ren, X., Su, G., Perot, V., Dy, J., Pfister, T.: Learning to prompt for continual learning. IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 139–149 (2022)
25. Wu, X., Xu, Z., Tong, R.K.y.: Continual learning in medical image analysis: A survey. Computers in Biology and Medicine **182**, 109206 (2024)
26. Xu, Z., Wang, Y., Lu, D., Luo, X., Yan, J., Zheng, Y., Tong, R.K.y.: Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation. Medical Image Analysis p. 102880 (2023)
27. Yan, S., Xie, J., He, X.: DER: Dynamically expandable representation for class incremental learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3014–3023 (2021)
28. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: MedMNIST v2-A large-scale lightweight benchmark for 2D and 3D biomedical image classification. Scientific Data **10**(1), 41 (2023)
29. Yoon, J., Yang, E., Lee, J., Hwang, S.J.: Lifelong learning with dynamically expandable networks. arXiv preprint arXiv:1708.01547 (2017)

30. Zhang, G., Wang, L., Kang, G., Chen, L., Wei, Y.: SLCA: Slow Learner with Classifier Alignment for Continual Learning on a Pre-trained Model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2023)
31. Zhang, W., Huang, Y., Zhang, T., Zou, Q., Zheng, W., Wang, R.: Adapter learning in pretrained feature extractor for continual learning of diseases. arXiv preprint arXiv:2304.09042 (2023)
32. Zhao, L., Zhang, X., Yan, K., Ding, S., Huang, W.: SAFE: slow and fast parameter-efficient tuning for continual learning with pre-trained models. arXiv preprint arXiv:2411.02175 (2024)
33. Zhou, D.W., Cai, Z.W., Ye, H.J., Zhan, D.C., Liu, Z.: Revisiting class-incremental learning with pre-trained models: Generalizability and adaptivity are all you need. International Journal of Computer Vision (2024)
34. Zhou, D.W., Sun, H.L., Ye, H.J., Zhan, D.C.: Expandable subspace ensemble for pre-trained model-based class-incremental learning. CVPR pp. 23554–23564 (2024)