# Eigen-Directed Random Projection for Medical Image Class Incremental Learning

Yongyi Wu[⋆1(✉)], Yichen Wu[⋆2], Quanzheng Li[2], Jianhua Ma[1], and Hong Wang[†1(✉)]

[1] Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China
[2] Massachusetts General Hospital, Harvard Medical School, Boston, USA
yongyiwu@stu.xjtu.edu.cn, hongwang01@xjtu.edu.cn

**Abstract.** Class-incremental learning (CIL) is essential for medical imaging systems that incorporate new diseases over time, as it avoids sacrificing performance on previously learned conditions. We first revisit a broad range of CIL methods on medical benchmarks and find, unexpectedly, that a previously proposed classifier-based analytic paradigm can outperform many widely used state-of-the-art alternatives. Motivated by this observation, we examine this paradigm more closely and observe that existing classifier-based methods fail to adequately separate features of similar examples and leave them too close in the representation space, resulting in insufficient margins and reducing separability as new classes arrive. To address these limitations, we specifically propose a training-free analytic margin-expansion framework that builds the classifier using an eigen-directed design, enlarging discriminative margins along principal data directions while preserving complementary detail. Across many commonly used medical benchmarks, our method consistently improves over a wide range of CIL baselines, and extensive ablations and theoretical analysis further support the proposed design. Code will be released at https://github.com/hmep313/EDRP_MICCAI.

**Keywords:** Continual learning · Replay-free · Disease diagnosis

## 1 Introduction

In medical image-guided diagnosis, deep learning models generally need to adapt over time as clinical practice evolves, including the appearance of new diseases and the adoption of new imaging modalities. This continual evolution is naturally cast as Class-Incremental Learning (CIL), where new classes arrive sequentially [18,3,27]. However, updating the model for new classes can disrupt previously learned representations, leading to catastrophic forgetting and reflecting the stability-plasticity dilemma [25,19]. In addition, privacy and governance

⋆ Equal contribution.
† Corresponding author.

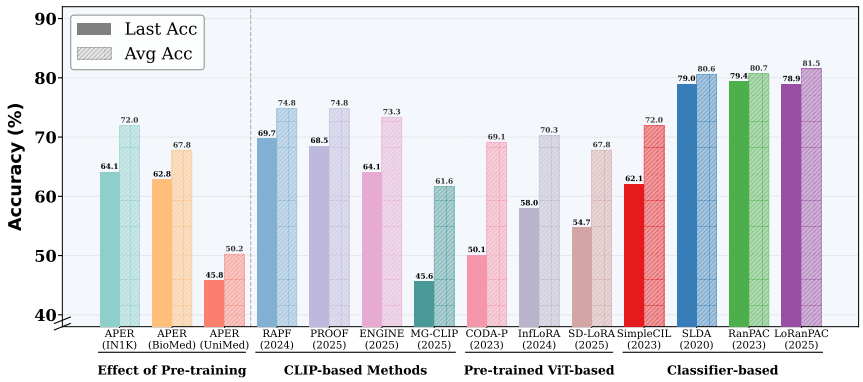



**Fig. 1.** Performance comparison of various CL methods on MedMNIST-Sub.

constraints often prohibit retaining historical patient data [23], which makes rehearsal-based solutions impractical in many clinical settings. Consequently, robust replay-free CIL is essential for scalable clinical deployment.

To alleviate forgetting without data replay, recent CIL methods increasingly start from Pre-trained Foundation Models (PFMs) and adapt them incrementally using parameter-efficient tuning [31,12,4]. We first systematically reevaluate a broad set of representative CIL methods in the medical setting, spanning CLIP-based [6,30,7,33], Pre-trained ViT-based [13,8,24], and classifier-based paradigms [29,5,9,11]. As shown in Fig. 1, a counterintuitive trend emerges: an earlier classifier-based analytic method consistently achieves the strongest performance, outperforming many more recent alternatives. This observation may reflect a characteristic challenge of medical imaging, where many classes differ only subtly and form tightly clustered features [21], making performance highly sensitive to representation shift during incremental updates [15,2].

This motivates renewed attention to classifier-based approaches, which leverage fixed pre-trained features and analytic classifier updates. For instance, RanPAC [9] builds on pre-trained features and updates the classifier analytically using random projection and closed-form solutions. While this design is efficient and stable, our empirical analysis shows that it can struggle in fine-grained medical diagnosis. A key reason is that random projection are agnostic to the structure of medical features, so they do not reliably expand the separation between highly similar classes, leaving them mixed in the projected space. To address this, we replace purely random mapping with a deterministic, data-driven projection that explicitly emphasizes the feature directions responsible for class confusion, and thus enlarges decision margins between entangled classes. We further provide theoretical justification showing that the achievable margins after projection are governed by the most vulnerable directions of the feature space, which motivates and supports our design. In summary, our contributions are:

- We revisit representative CIL baselines in the medical incremental setting and find that a classifier-based analytic paradigm unexpectedly delivers the strongest performance.
- We propose EDRP (Eigen-Directed Random Projection), a training-free dual-view analytic CIL method that enhances fine-grained separability through an eigen-directed projection while maintaining low memory overhead.
- We provide theoretical support and extensive experiments with ablations on multiple medical datasets.

## 2 Method

### 2.1 Preliminaries

**Problem Setting.** We consider the CIL scenario, where tasks $\{\mathcal{D}_t\}_{t=1}^{T}$ arrive sequentially. Each task $\mathcal{D}_t = \{(\mathbf{x}_i^t, y_i^t)\}_{i=1}^{n_t}$ contains samples whose labels belong to a task-specific class set $\mathcal{Y}_t$, and the class sets are mutually exclusive, i.e., $\mathcal{Y}_t \cap \mathcal{Y}_{t'} = \emptyset$ for $t \neq t'$. Let $\mathcal{Y}_{1:t} = \bigcup_{k=1}^{t} \mathcal{Y}_k$, the goal is to learn a single model that predicts over $\mathcal{Y}_{1:t}$ without task identity at inference, training at task $t$ without access to previous-task data.

**RanPAC.** For clarity, we write the model as $f_{\mathbf{W}} \circ g$, where $g(\cdot)$ is a frozen feature extractor and $f_{\mathbf{W}}$ denotes the linear classifier. Given an input $\mathbf{x}$, RanPAC [9] extracts a frozen backbone feature $\mathbf{z} = g(\mathbf{x}) \in \mathbb{R}^d$, and expands it via a fixed Gaussian random projection $\mathbf{P}_{\text{rand}} \in \mathbb{R}^{d \times M}$ ($M \gg d$) with a nonlinearity $\phi$:

$$\mathbf{h} = \phi(\mathbf{z}\mathbf{P}_{\text{rand}}) \in \mathbb{R}^M. \tag{1}$$

The linear classifier weights $\mathbf{W} \in \mathbb{R}^{M \times C}$ (where $C$ is the number of classes) is then derived analytically via a closed-form ridge-regression update.

### 2.2 The Proposed Eigen-Directed Random Projection Algorithm

Based on the observation in Fig. 1, we revisit strong classifier-based replay-free CIL methods in the medical setting. Using RanPAC [9] as a representative example (see Eq. (1)), we observe a consistent failure mode on fine-grained recognition: visually similar classes remain poorly separated after random feature expansion. As illustrated in Fig. 2(a), embeddings of confusable pathologies stay entangled in the projected space, which leads to insufficient decision margins.

**Theoretical Analysis.** To connect this phenomenon to the projection design, we follow [16,14] to analyze the expansion map from a margin perspective. Let $\gamma$ denote the instance margin of a linear classifier in the original feature space, and let $\gamma_{\text{new}}$ be the corresponding margin after applying a linear projection $\mathbf{P}$. Then, the post-projection margin admits the deterministic lower bound,

$$\gamma_{\text{new}} \geq \sigma_{\min}(\mathbf{P})\gamma, \tag{2}$$

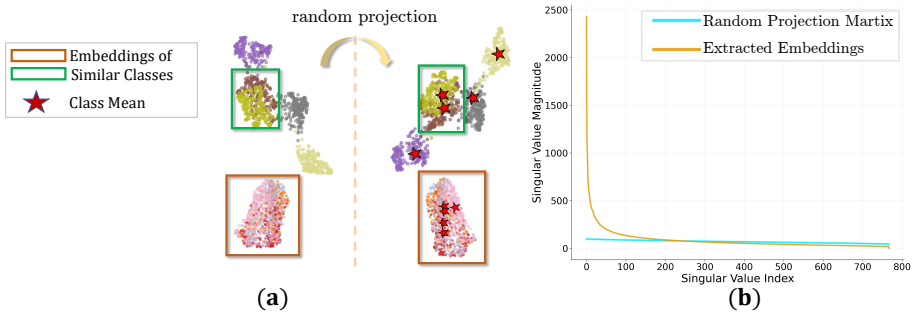



**Fig. 2.** **(a)** The t-SNE visualizations of the feature distribution shows that similar categories remain difficult to separate after random projections. **(b)** Comparison between the singular values of the random projection matrix and those of the embeddings covariance.

where $\sigma_{\min}(\mathbf{P})$ is the minimum singular value of $\mathbf{P}$. This follows directly from the definition of singular values: for any vector $\mathbf{u}$, $\|\mathbf{uP}\|_2 \geq \sigma_{\min}(\mathbf{P})\|\mathbf{u}\|_2$. After the projection, a pointwise activation $\rho$ preserves this bound up to a constant determined by its local Lipschitz behavior. Therefore, this bound implies that fine-grained separability can be bottlenecked by the weakest directions of the projection, which are not explicitly controlled by an isotropic random matrix.

**Eigen-Directed Projection.** Motivated by the margin bound, we first explore whether the random projection used in RanPAC is well matched to medical embeddings. Fig. 2(b) shows that medical features exhibit a strongly anisotropic spectrum, with a few dominant directions capturing most variation, whereas a Gaussian random matrix is isotropic. This mismatch suggests that purely random projection do not reliably strengthen the directions that matter most for separating confusable medical classes, motivating our eigen-directed projection aligned with the data principal directions.

We therefore move beyond isotropic randomness and construct a data-aware projection. Concretely, we maintain a small feature buffer $\mathbf{Z}_{\text{buf}} \in \mathbb{R}^{B\times d}$ (where $B$ denotes the buffer size) and perform spectral decomposition on its empirical correlation matrix to extract the principal eigenspace:

$$\mathbf{Z}_{\text{buf}}^{\top}\mathbf{Z}_{\text{buf}} \approx \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\top}, \tag{3}$$

where $\mathbf{U} \in \mathbb{R}^{d\times d}$ contains the orthogonal eigenvectors and $\mathbf{\Lambda}$ denotes the diagonal matrix of eigenvalues. We extract the sub-basis $\mathbf{U}_{\text{eig}} \in \mathbb{R}^{d\times K}$ corresponding to the top-$K$ eigenvalues to capture the dominant data manifold. Then, we construct an eigen-directed projection by aligning a random source matrix $\mathbf{R} \in \mathbb{R}^{d\times M_a}$ with this subspace:

$$\mathbf{P}_{\text{eig}} = \mathbf{U}_{\text{eig}}(\mathbf{U}_{\text{eig}}^{\top}\mathbf{R}) \in \mathbb{R}^{d\times M_a}. \tag{4}$$

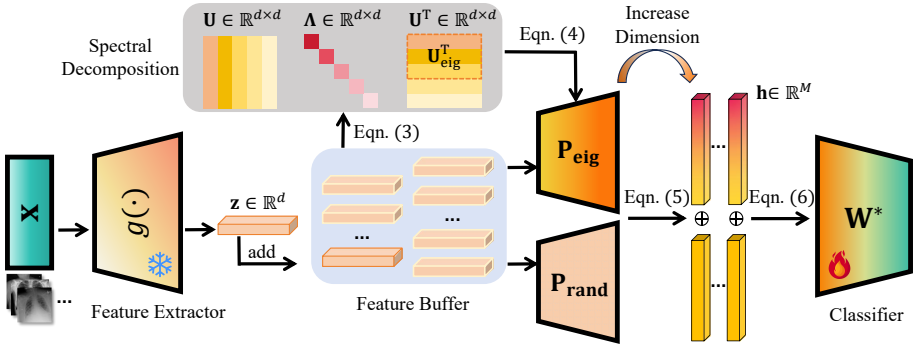



**Fig. 3.** The procedure for learning the classifier weights $\mathbf{W}^*$ during training. During testing, it simply requires calculating the cosine similarity between the obtained feature $\mathbf{h}$ and $\mathbf{W}^*$.

Since discriminative cues may also lie outside the dominant subspace, relying solely on truncation risks discarding these fine-grained details. To address this, we adopt a dual-view mapping that synergizes $\mathbf{P}_{\text{eig}}$ with a fixed random view $\mathbf{P}_{\text{rand}} \in \mathbb{R}^{d \times M_f}$, which acts as an isotropic safety net to preserve the spectral long tail. The final feature $\mathbf{h} \in \mathbb{R}^M$ (where $M = M_f + M_a$) is the concatenation:

$$\mathbf{h} = \phi([\mathbf{z}\mathbf{P}_{\text{rand}}, \mathbf{z}\mathbf{P}_{\text{eig}}]). \tag{5}$$

We then learn the linear classifier weights $\mathbf{W}^*$ via ridge regression. We construct the global feature matrix $\mathbf{H} \in \mathbb{R}^{N \times M}$ by stacking the dual-view features $\mathbf{h}$ of all available samples (including both the current task data and the historical buffer). The optimal weights are then derived via the closed-form solution:

$$\mathbf{W}^* = (\mathbf{H}^\top \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^\top \mathbf{Y}, \tag{6}$$

where $\mathbf{Y} \in \mathbb{R}^{N \times C}$ contains the corresponding one-hot labels. The overall pipeline is shown in Fig. 3.

**Remark.** Our method is driven by the margin collapse observed in classifier-based CIL. Through our theoretical analysis, it suggests that improving separability requires strengthening the projection, i.e., increasing $\sigma_{\min}(\mathbf{P})$, which motivates our eigen-directed design. Fig. 5 below will further empirically confirm that our proposed EDRP enlarges the observed margins.

## 3 Experiments

### 3.1 Experimental Setup

We evaluate on four medical imaging benchmarks. Skin8 [17] contains dermatoscopic images from 8 lesion categories and is split into 4 tasks with 2 classes per

task, featuring substantial class imbalance. MedMNIST-Sub [26,28] aggregates four diverse 2D datasets (Blood, OrganA, Path, and Tissue); we shuffle categories across datasets to mimic cross-modality continual learning in clinical practice, resulting in 4 tasks with 9 classes per task. We also evaluate on the Peripheral Blood Cell (PBC) [1] dataset (4 tasks, 2 classes each) and the challenging ChestX [20,10] dataset (3 tasks, 2 classes each). During training, all images are resized to $224 \times 224$ pixels. For all experiments, we employ ViT-B/16-IN1K [4] as the frozen feature extractor. In all relevant methods using projection, the projection dimension $M$ is set to $10,000$. For our method, we set $M$ to be evenly split: $M_f = M_a = M/2$. We store 80% of the embeddings for each class to enable more accurate eigenspace estimation. In our extended version, we will show in detail that the overhead associated with this type of storage is acceptable.

To ensure a fair comparison, following RanPAC [9], we apply the same first-session adaptation strategy. Let $a_{t,i}$ denote the accuracy on task $i$ after learning task $t$ out of $T$ total tasks. Models are evaluated using Last Accuracy ($Acc_{Last} = \frac{1}{T}\sum_{i=1}^{T} a_{T,i}$) and Average Accuracy ($Acc_{Avg} = \frac{1}{T}\sum_{t=1}^{T} A_t$, where $A_t = \frac{1}{t}\sum_{i=1}^{t} a_{t,i}$).

### 3.2 Main Results

**Performance on Standard Benchmarks.** Table 1 summarizes the performance on Skin8, MedMNIST-Sub, and PBC datasets. Consistent with Fig. 1, many CIL methods exhibit severe performance degradation, whereas classifier-based methods establish stronger baselines. Building upon RanPAC, our proposed EDRP significantly outperforms all competitors. For instance, on Skin8 dataset, our EDRP achieves 65.48% Last Accuracy and 76.31% Average Accuracy, outperforming LoRanPAC by a margin of nearly 3%. This superiority is consistent across MedMNIST-Sub and PBC.

**Performance on the Challenging ChestX Dataset.** ChestX is a particularly difficult CIL setting, where fine-grained categories and strong inter-class similarity lead to low accuracy for most methods. As shown in Table 2, under this challenging regime, recent CIL techniques degrade markedly, whereas our EDRP remains robust, achieving 29.76% Last Accuracy and 40.43% Average Accuracy. Moreover, the proposed EDRP demonstrates superior clinical robustness with minimal variance ($\pm$1.01%) compared to the unstable PROOF ($\pm$9.47%).

### 3.3 Ablation Studies and Other Analysis

**Effects of Different Components.** Table 3 highlights the progressive performance gains from our proposed modules. Starting from the baseline (without PR and EDP), incorporating the Eigen-Directed Projection (EDP) from Eqn. (4) already provides a substantial improvement, elevating $Acc_{Last}$ from 57.80% to 63.29%. This underscores the benefit of aligning features with the principal component space. Building upon this, the addition of Random Projection (RP) from Eqn. (5) further complements the framework, forming our complete EDRP which

**Table 1.** Class-incremental learning results on three datasets.

| Method | Skin8 | | MedMNIST-Sub | | PBC | |
|---|---|---|---|---|---|---|
| | $Acc_{Last}(\%)$ | $Acc_{Avg}(\%)$ | $Acc_{Last}(\%)$ | $Acc_{Avg}(\%)$ | $Acc_{Last}(\%)$ | $Acc_{Avg}(\%)$ |
| ACL [28] | 57.72±2.04 | 58.20±2.38 | 65.64±0.67 | 70.68±0.99 | 82.30±0.11 | 75.97±0.30 |
| CRCL [22] | 60.91±2.30 | 61.35±2.93 | 68.49±0.61 | 68.71±0.91 | 95.73±0.31 | 96.16±0.16 |
| RAPF [6] | 40.08±2.50 | 45.52±2.16 | 69.87±0.62 | 76.38±1.63 | 90.95±0.39 | 91.79±0.32 |
| PROOF [33] | 51.49±2.93 | 61.26±2.98 | 68.60±0.23 | 76.12±1.19 | 83.65±0.17 | 90.02±0.21 |
| ENGINE [30] | 50.29±2.55 | 59.95±2.36 | 63.96±0.86 | 74.81±1.51 | 85.07±0.32 | 90.88±0.25 |
| MG-CLIP [7] | 26.46±2.96 | 47.02±3.16 | 45.75±0.74 | 61.52±0.93 | 50.25±0.18 | 71.08±0.48 |
| CODA-Prompt [7] | 32.65±2.47 | 50.24±2.67 | 50.27±0.52 | 69.01±1.15 | 61.60±0.18 | 72.68±0.46 |
| InfLoRA [8] | 37.55±2.33 | 54.98±2.87 | 57.80±0.23 | 70.10±1.53 | 70.91±0.23 | 79.11±0.25 |
| SD-LoRA [24] | 20.41±2.79 | 54.13±3.46 | 54.47±0.70 | 67.87±1.44 | 65.14±0.29 | 76.04±0.27 |
| EASE [32] | 40.41±3.43 | 60.57±2.45 | 45.28±0.33 | 67.85±1.77 | 73.51±0.29 | 78.90±0.16 |
| SimpleCIL [29] | 39.64±3.21 | 56.79±2.09 | 61.33±0.84 | 72.10±1.81 | 80.56±0.77 | 82.10±0.74 |
| SLDA [5] | 61.37±2.44 | 73.02±2.43 | 78.97±0.29 | 80.59±1.05 | 95.99±0.13 | 96.46±0.28 |
| RanPAC [9] | 62.31±2.86 | 73.05±3.68 | 79.75±0.34 | 82.58±1.62 | 96.84±0.16 | 97.95±0.20 |
| LoRanPAC [11] | 62.66±2.31 | 74.31±2.65 | 78.93±0.31 | 82.52±0.98 | 96.81±0.14 | 97.87±0.18 |
| **Ours** | **65.48**±0.96 | **76.31**±0.94 | **80.77**±0.28 | **83.52**±1.51 | **97.17**±0.12 | **98.19**±0.42 |

**Table 2.** Performance on ChestX.

| Method | ChestX | |
|---|---|---|
| | $Acc_{Last}(\%)$ | $Acc_{Avg}(\%)$ |
| Joint Training | 55.51 | 55.51 |
| InfLoRA [8] | 15.18±2.73 | 31.86±3.21 |
| PROOF [33] | 21.41±9.47 | 36.49±6.06 |
| SLDA [5] | 18.14±8.12 | 25.12±2.19 |
| RanPAC [9] | 12.29±2.58 | 31.04±2.42 |
| LoRanPAC [9] | 26.06±7.22 | 38.67±4.72 |
| Ours | **29.76**±1.01 | **40.43**±2.12 |

**Table 3.** Ablation study.

| RP | EDP | $Acc_{Last}(\%)$ | $Acc_{Avg}(\%)$ |
|---|---|---|---|
| × | × | 57.80±1.10 | 69.94±1.89 |
| ✓ | × | 59.78±2.79 | 71.75±2.20 |
| × | ✓ | 63.29±0.64 | 74.27±2.09 |
| ✓ | ✓ | **65.48**±0.96 | **76.31**±0.94 |

reaches 65.48% Last and 76.31% Avg Accuracy. This incremental success further validates the effectiveness of our proposed EDRP algorithm.

**Robustness Across Principal Component Dimensions ($K$).** Fig. 4 illustrates the singular value distribution of different projection strategies. Compared to Random Projection used in RanPAC, our Eigen-Direction Projection significantly elevates the singular value spectrum across various dimensions $K$. Crucially, even as $K$ increases, the minimum singular values remain consistently higher than those of Random Projection, which directly corroborates our theoretical analysis in Sec. 2.2. On the other hand, Tabel 4 reveals an inverted-U performance trend peaking at $K = 200$, balancing margin expansion with detail preservation. Crucially, our EDRP consistently outperforms RanPAC across all $K$ values. This remarkable robustness, as corroborated by the elevated singular values in Fig. 4, eliminates the need for exhaustive hyperparameter tuning.

**Empirical Validation of Margin Enhancement.** To empirically validate our theoretical claim that maximizing the minimum singular value $\sigma_{\min}$ enhances the decision margin, we visualize the projected feature space on the Skin8 dataset.

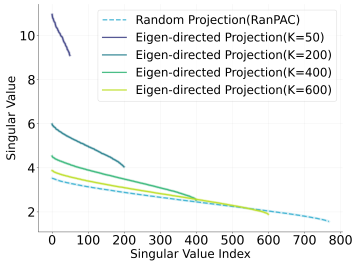
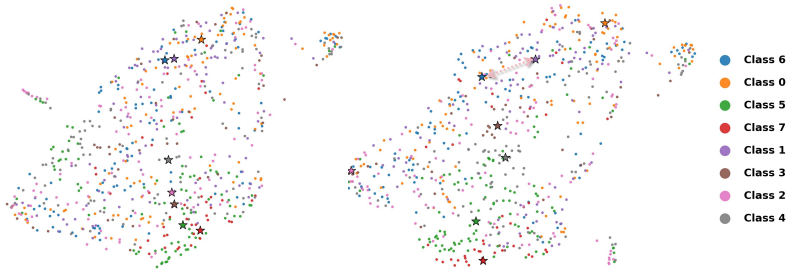



**Fig. 4.** Singular value comparison between two types of projections.

**Table 4.** Effect of the number of principal components.

| Top-$K$ | 50 | 100 | 200 | 300 |
|---|---|---|---|---|
| $Acc_{Last}(\%)$ | 64.67 | 65.97 | 66.67 | 65.85 |
| $Acc_{Avg}(\%)$ | 76.51 | 77.09 | 77.87 | 77.44 |

| Top-$K$ | 500 | 600 | 700 | w/o |
|---|---|---|---|---|
| $Acc_{Last}(\%)$ | 64.96 | 64.54 | 63.31 | 62.41 |
| $Acc_{Avg}(\%)$ | 76.55 | 75.83 | 74.79 | 73.76 |




(a) (b)

**Fig. 5.** Visualization of class embeddings and class means on the Skin8 under random projection (a) and our method (b). Star markers indicate class means.

As shown in Fig. 5(a), under the standard Gaussian random projection (used in RanPAC), class prototypes of visually similar lesions remain highly entangled. In contrast, our EDRP (Fig. 5(b)) effectively pushes these overlapping clusters apart. By deterministically aligning the projection basis with the data's principal directions, our method demonstrably expands the margin between hard negatives. This clear separation directly corroborates our theoretical foundation and explains the superior classification performance.

**Memory Consumption.** Table 5 validates our $O(Bd)$ memory efficiency. While baselines like RanPAC impose a static $>$ 411 MB overhead across datasets due to their $O(M^2)$ covariance matrices, our strategy requires only 40.30 MB on Skin8 and 120.60 MB on the largest MedMNIST-Sub. This substantial reduction makes our framework highly practical for resource-constrained clinical hardware.

## 4 Conclusion

In this paper, we extensively analyze existing pre-trained continual learning methods on medical images, revealing a general and severe performance degradation. Although analytic classifier-based approaches show promise for their strong

**Table 5.** Storage comparison (MB) on different datasets.

| **Methods** | **RanPAC** [9] | | **Ours** | |
|---|---|---|---|---|
| **Dataset** | MedMNIST | Skin8 | MedMNIST | Skin8 |
| **Storage (MB)** | 414.13 | 411.07 | 120.60 | 40.30 |

stability, we identify that they still struggle with fine-grained medical tasks due to feature entanglement under standard random projections. Specifically, the inherent inter-class similarity in medical datasets often leads to collapsed decision boundaries when using conventional Gaussian matrices. To address this, we propose Eigen-Directed Random Projection (EDRP), a training-free dual-view framework. It couples an eigen-directed random projection to deterministically expand the decision margin along principal dimensions with a fixed random projection to preserve long-tail details. Extensive experiments across diverse modalities including skin lesions and blood cells demonstrate that our approach achieves state-of-the-art accuracy while significantly reducing memory overhead.

**Acknowledgments.** This work was supported in part by the Ministry of Science and Technology of the People's Republic of China, Key Research and Development Program under Grant 2024YFA1012003, Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China under Grant JYB2025XDXM101, and China NSFC Project under Grant 62501466.

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

## References

1. 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)
2. Chen, H., Wang, Y., Fan, Y., Jiang, G., Hu, Q.: Reducing class-wise confusion for incremental learning with disentangled manifolds. In: Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR). pp. 10121–10130 (June 2025)
3. De Lange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., Slabaugh, G., Tuytelaars, T.: A continual learning survey: Defying forgetting in classification tasks. IEEE Transactions on pattern analysis and machine intelligence **44**(7), 3366–3385 (2021)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. ICLR (2021)
5. Hayes, T.L., Kanan, C.: Lifelong machine learning with deep streaming linear discriminant analysis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 220–221 (2020)
6. Huang, L., Cao, X., Lu, H., Liu, X.: Class-incremental learning with clip: Adaptive representation adjustment and parameter fusion. In: European Conference on Computer Vision. pp. 214–231. Springer (2024)

7. Huang, L., Cao, X., Lu, H., Meng, Y., Yang, F., Liu, X.: Mind the gap: Preserving and compensating for the modality gap in clip-based continual learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2025)
8. Liang, Y.S., Li, W.J.: Inflora: Interference-free low-rank adaptation for continual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23638–23647 (2024)
9. 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**, 12022–12053 (2023)
10. Menabue, M., Frascaroli, E., Boschini, M., Sangineto, E., Bonicelli, L., Porrello, A., Calderara, S.: Semantic residual prompts for continual learning. In: European Conference on Computer Vision. pp. 1–18. Springer (2024)
11. Peng, L., Elenter, J., Agterberg, J., Ribeiro, A., Vidal, R.: Loranpac: Low-rank random features and pre-trained models for bridging theory and practice in continual learning. arXiv preprint arXiv:2410.00645 (2024)
12. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision (2021), https://arxiv.org/abs/2103.00020
13. Smith, J.S., 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. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11909–11919 (2023)
14. Sokolic, J., Giryes, R., Sapiro, G., Rodrigues, M.R.D.: Robust large margin deep neural networks. IEEE Transactions on Signal Processing **65**(16), 4265–4280 (Aug 2017), http://dx.doi.org/10.1109/TSP.2017.2708039
15. Toldo, M., Ozay, M.: Bring evanescent representations to life in lifelong class incremental learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16732–16741 (2022)
16. Tran, Q., Tran, T.L., Doan, K., Tran, T., Phung, D., Than, K., Le, T.: Boosting multiple views for pretrained-based continual learning. In: The Thirteenth International Conference on Learning Representations (2025), https://openreview.net/forum?id=AZR4R3lw7y
17. 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), 180161 (2018)
18. Wang, L., Zhang, X., Su, H., Zhu, J.: A comprehensive survey of continual learning: Theory, method and application. IEEE transactions on pattern analysis and machine intelligence **46**(8), 5362–5383 (2024)
19. Wang, Q., Wang, R., Li, Y., Wei, D., Wang, H., Ma, K., Zheng, Y., Meng, D.: Relational experience replay: Continual learning by adaptively tuning task-wise relationship. IEEE Transactions on Multimedia **26**, 9683–9698 (2024)
20. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2097–2106 (2017)
21. Wei, X.S., Song, Y.Z., Mac Aodha, O., Wu, J., Peng, Y., Tang, J., Yang, J., Belongie, S.: Fine-grained image analysis with deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence **44**(12), 8927–8948 (2021)

22. Wu, X., Xu, Z., Lu, D., Sun, J., Liu, H., Shakil, S., Ma, J., Zheng, Y., Tong, R.: Conservative-radical complementary learning for class-incremental medical image analysis with pre-trained foundation models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) (2025)
23. Wu, X., Xu, Z., Tong, R.K.y.: Continual learning in medical image analysis: A survey. Computers in Biology and Medicine **182**, 109206 (2024)
24. Wu, Y., Piao, H., Huang, L.K., Wang, R., Li, W., Pfister, H., Meng, D., Ma, K., Wei, Y.: Sd-lora: Scalable decoupled low-rank adaptation for class incremental learning. arXiv preprint arXiv:2501.13198 (2025)
25. Wu, Y., Wang, H., Zhao, P., Zheng, Y., Wei, Y., Huang, L.K.: Mitigating catastrophic forgetting in online continual learning by modeling previous task interrelations via pareto optimization. In: Forty-first international conference on machine learning (2024)
26. 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)
27. Yang, J., Wu, Y., Cen, J., Huang, W., Wang, H., Zhang, J.: Continual learning for segment anything model adaptation. arXiv preprint arXiv:2412.06418 (2024)
28. Zhang, W., Huang, Y., Zhang, T., Zou, Q., Zheng, W.S., Wang, R.: Adapter learning in pretrained feature extractor for continual learning of diseases. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 68–78. Springer (2023)
29. 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 **133**(3), 1012–1032 (2025)
30. Zhou, D.W., Li, K.W., Ning, J., Ye, H.J., Zhang, L., Zhan, D.C.: External knowledge injection for clip-based class-incremental learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3314–3325 (2025)
31. Zhou, D.W., Sun, H.L., Ning, J., Ye, H.J., Zhan, D.C.: Continual learning with pre-trained models: A survey. arXiv preprint arXiv:2401.16386 (2024)
32. Zhou, D.W., Sun, H.L., Ye, H.J., Zhan, D.C.: Expandable subspace ensemble for pre-trained model-based class-incremental learning. In: CVPR (2024)
33. Zhou, D.W., Zhang, Y., Wang, Y., Ning, J., Ye, H.J., Zhan, D.C., Liu, Z.: Learning without forgetting for vision-language models. IEEE Transactions on Pattern Analysis and Machine Intelligence (2025)