ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10890364

AIDC: Benchmark for Analytical Learning in Incremental Disease Classification

Rongchang Zhao
School of Computer Science
Central South University
Changsha, China
zhaorc@csu.edu.cn

Jianyu Qi
School of Computer Science
Central South University
Changsha, China
qijianyu@csu.edu.cn

Rui Li
School of Computer Science
Central South University
Changsha, China
rui.li@csu.edu.cn

Zhijie Zheng
School of Traffic and Transportation Engineering
Central South University
Changsha, China
zhengzhijie@csu.edu.cn

Jian Zhang
School of Computer Science
Central South University
Changsha, China
jianzhang@csu.edu.cn

Jiaxu Li*
School of Computer Science
Central South University
Changsha, China
lijiaxu@csu.edu.cn

Abstract—**Class Incremental Learning (CIL) aims to enable models to continuously learn new categories while retaining previous classification abilities. In medical scenarios, where new disease categories frequently emerge, CIL becomes crucial. Traditional CIL approaches often face "catastrophic forgetting". Analytical Class Incremental Learning (ACIL) offers an analytical (i.e., closed-form) linear solution that does not depend on conventional replay or regularization techniques, thereby mitigating forgetting and addressing privacy concerns, making it suitable for medical datasets. However, few studies have explored the problem of knowledge forgetting in CIL for medical data with ACIL. Based on the latest research, we systematically study this problem for the first time. Specifically, we present a benchmark named AIDC (Analytical Incremental Disease Classification), comparing ACIL against five established CIL methods across three medical datasets. Results show that ACIL achieves notably higher average classification accuracy and exhibits better anti-forgetting capabilities compared to traditional methods.**

Index Terms—**analytic learning, incremental learning, medical data classification**

I. Introduction

Conventional machine learning models typically learn by capturing static data, but real-world data is frequently in stream format [1]. The interest for incremental learning research has been fueled by the expectation that models, like organisms, may learn and adapt to this continually changing dynamic input. However, incremental learning suffers from a serious catastrophic forgetting problem, which is also known as the "stability-plasticity dilemma" in neural system [2], where directly using new data to optimize the network will erase previous knowledge, causing the model to experience irreversible performance degradation. Therefore, how to effectively resist catastrophic forgetting becomes the core issue in building the incremental learning model [3].

As a typical representative of incremental learning, **class incremental learning (CIL)** enables the learner to incorporate the knowledge of new classes incrementally and build a universal classifier among all seen classes [1]. It becomes increasingly important as it provides a system that can continuously learn from new categories of data in an ever-changing world. And there are two main categories to solve catastrophic forgetting in CIL: replay-based CIL and exemplar-free CIL.

Replay-based CIL. ICaRL [4] is the pioneering replay-based approach in CIL that combats catastrophic forgetting by maintaining a limited set of historical data. Chaudhry et al. [5] structures the replay mechanism as a dual-level optimization process and retained full predictions for certain anchor points from previous tasks. De Lange and Tuytelaars [6] implemented data replay within prototype networks [7], using the exemplars as pseudo-prototypes to evaluate embeddings. GR [8] is the first to suggest employing generative adversarial networks within the context of CIL. DDGR [9] adopts a diffusion model for generation and instructs the generation of samples by computing instruction operators via a classifier. However, due to the fact that the exemplar set retains only a small fraction of the training dataset, replay-based CIL may suffer the overfitting problem. In addition, it faces data storage and data privacy issues. For methods based on generative models, model performance is subject to the quality of real data, and training generative models introduces additional computing and storage overhead.

Exemplar-free CIL. We mainly introduce two types of CIL incremental learning methods that do not require exemplars: regularization-based CIL and model correction-based CIL. In addition, there are some methods based on dynamic networks [10]–[12]. The above methods are often implemented with the help of knowledge distillation [13]–[15].

1) Regularization-based CIL: Regularization-based CIL imposes some constraints on the objective function to prevent forgetting. As the first work to address parameter regularization, EWC [16] estimates the importance with a Fisher information matrix and constrains those weights. Later, KFAC [17]

[*] Corresponding author

Authorized licensed use limited to: Central South University. Downloaded on December 25,2025 at 03:35:31 UTC from IEEE Xplore. Restrictions apply.

employs a Kronecker factorization technique to enhance the approximation of the Fisher information matrix. However, the process of estimating parameter importance necessitates storing matrices that are the same size as the network itself, which poses a challenge in terms of linearly increasing memory requirements as more tasks are learned.

2) Model Correction-based CIL: Model correction-based CIL aims to address recency bias, which means that incremental training usually trains more data for new classes [18]. LUCIR [19] suggests employing a cosine classifier to minimize this bias towards recent targets. WA [20] builds upon this by normalizing the weights following each optimization step and incorporating weight clipping to ensure that the predicted probability is proportional to the classifier weight. The model correction-based approach attempts to reduce the inductive bias of the CIL model by investigating the specific elements that lead to catastrophic forgetting. Future research should delve deeper into additional factors and potential remedies for catastrophic forgetting.

As an advanced class incremental learning method, **analytic class incremental learning (ACIL)** was originally inspired by analytical learning (AL) [21]. Analytical learning, also known as pseudoinverse learning [22], is used to address the limitations of the back-propagation (BP) method, such as gradient vanishing or exploding, divergence during iteration, and long training time requiring a large number of iterations. Analytical learning requires learning weights using the entire dataset at once, so it may encounter the problem of insufficient internal demand. It can be alleviated by block-wise recursive Moore-Penrose learning (BRMP) [21]. Coincidentally, BRMP uses a recursive method for learning, which can maintain the influence of previous exemplars while streaming new exemplars to update weights, which well meets the needs of class incremental learning. By combining analytical learning with its recursive formulation, Zhuang et al. [23] built an ACIL system whose classifier uses an algorithm similar to concatenated recursive least squares (C-RLS) to train to generate closed-form solutions [24]. It can effectively preserve historical exemplars without revisiting them.

Class Incremental Learning for Medical Data. CIL requires the model to continually learn to differentiate between an increasing number of classes over time and evaluate all observed classes at test time, which is essential in the medical field [1]. However, traditional CIL methods often show more serious catastrophic forgetting problem on medical data, making them unable to meet medical needs. This is because medical data is more dependent on professional knowledge and has more sensitive data privacy issues.

ACIL for Medical Data. Since the introduction of ACIL, its performance has been enhanced across various domains, including few-shot CL [25], federated learning [26], reinforcement learning [27], and generalized CIL [28]. However, few studies have explored the problem of knowledge forgetting in CIL based on the latest ACIL for medical data. To bridge this gap, we systematically study this problem for the first time based on ACIL. Specifically, we introduced ACIL and its two variants, DS-AL [29] and AIR [24]. In addition, we compared five advanced incremental learning methods, including LwF [30], iCaRL [4], EWC [16], MAS [31] and EEIL [32]. By testing and comparing the average classification accuracy and forgetting of these methods on three medical disease datasets, we introduce a benchmark for CIL on medical data.

II. Methods

A. Class Incremental Learning Definition

The network undergoes continuous training for T tasks, where the training data for each task consists of different classes. Let the training and testing datasets for task t ($t = 1, \ldots, T$) be represented as $\mathcal{D}_t^{\text{train}} = \{\mathbf{Z}_t^{\text{train}}, \mathbf{Y}_t^{\text{train}}\}$ and $\mathcal{D}_t^{\text{test}} = \{\mathbf{Z}_t^{\text{test}}, \mathbf{Y}_t^{\text{test}}\}$, respectively. Here, $\mathbf{Z}_t \in \mathbb{R}^{N_t \times c \times w \times h}$ (e.g., N_t images with dimensions $c \times w \times h$) and $\mathbf{Y}_t^{\text{train}} \in \mathbb{R}^{N_t \times d_{y_t}}$ (where task t has d_{y_t} classes) are the stacked input and label (one-hot) tensors. Moreover, $\mathbf{Z}_i \cap \mathbf{Z}_j = \emptyset$ for $i \neq j$, indicating that the input sets for different tasks are disjoint.

B. Analytic Class Incremental Learning Methods

ACIL (Analytical Class Incremental Learning [23]) integrates analytical learning to train classifiers using concatenated recursive least squares (C-RLS), mitigating the issue of catastrophic forgetting and avoiding reliance on historical data. Compared to traditional BP-based methods, ACIL offers significant improvements in both time and cost efficiency, particularly when privacy concerns in medical data prevent the use of historical information. Since ACIL requires only one training epoch during subsequent CIL following the pretraining on task 1, the performance of learning new classes heavily depends on the initial quality of the backbone on task 1. ACIL comprises three main components: BP-based training, AL-based re-alignment and AL-based CIL. As depicted in Figure 1, ACIL begins by utilizing a BP-based method to train a regular classification network on the base training set (task 1), which typically involves multiple training epochs. The network structure can be ResNet [33] or ViT [34] backbone combined with a conventional classifier. Once training is complete, the backbone weights are saved and frozen for later use with the analytic classifier during retraining. Following the BP-based training, ACIL moves into the AL-based re-alignment phase using least squares (LS) to determine the optimal solution, which is then saved. After completing the AL-based re-alignment, ACIL continues the AL-based CIL process using C-RLS algorithm, allowing an equivalence between the CIL and its joint learning counterpart. This process is driven by the analytic classifier, the correctness of which is detailed in [23]. Based on ACIL, here are two enhanced variants, which we will introduce below.

DS-AL (Dual-Stream Analytical Learning [29]) enhances the performance of ACIL by introducing a compensation mechanism. Since the backbone is frozen during the incremental learning process, training only the classification head may be insufficient. DS-AL improves this by comprising a main stream that provides an analytical linear solution and a

Authorized licensed use limited to: Central South University. Downloaded on December 25,2025 at 03:35:31 UTC from IEEE Xplore. Restrictions apply.

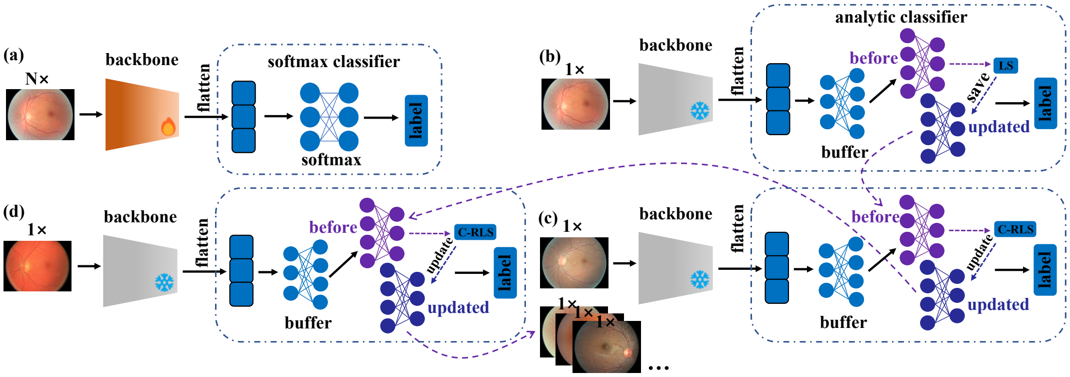

Fig. 1. The three main components of ACIL: **(a)** BP-based training, which uses a backbone and a softmax classifier for multiple epochs of training. **(b)** AL-based re-alignment, it replaces the previous classifier with an analytical classifier and aligns it with the backbone. The buffer is a high-dimensional expansion layer. The features will be calculated using the least squares (LS) method to obtain the optimal solution and saved. **(c) and (d)** AL-based CIL, using the aligned network to learn new classes using the C-RLS algorithm, with only one epoch of training for each new class.

compensation stream designed to tackle the inherent under-fitting issue caused by linear mapping. The adjustment of the compensation mechanism requires optimization across different datasets. The compensation stream operates through a dual-activation compensation (DAC) module, which reactivates the embedding with an activation function different from that of the main stream. It compensates for under-fitting by projecting the embedding onto the null space of the main stream's linear mapping.

AIR (Analytical Imbalance Rectifier [24]) addresses data-imbalanced class-incremental learning by integrating a re-weighting factor into the loss function, ensuring equal class contribution. The analytical reweighting module (ARM) calculates this factor for each class, effectively handling imbalanced training data. In the AIR paper, this factor is simply the inverse of the number of samples within each class.

C. Five Classical Class Incremental Learning Methods

LwF (Learning without Forgetting [30]) tackles the problem of catastrophic forgetting using knowledge distillation. LwF does not require storing old task data but instead leverages the output of previous tasks as soft labels to guide the learning of new tasks. By maintaining consistency between the outputs of new and old tasks, LwF preserves knowledge from earlier tasks. It is particularly useful in situations where old task data cannot be stored, effectively reducing catastrophic forgetting.

iCaRL (Incremental Classifier and Representation Learning [4]) addresses the problem of catastrophic forgetting through a sample replay strategy. Unlike LwF, which relies on knowledge distillation, iCaRL periodically revisits representative samples from previous tasks to retain knowledge while learning new tasks. However, depending on a limited set of samples can lead to class imbalance, and over time, the accumulation of errors may degrade the model's overall performance.

EWC (Elastic Weight Consolidation [16]) is a regularization-based approach aimed at protecting important parameters from previous tasks. It uses the Fisher information matrix to assess which parameters are crucial for earlier tasks and penalizes significant changes to these parameters. However, EWC assumes that the loss function in parameter space is locally quadratic, which may reduce its effectiveness when tasks vary significantly.

MAS (Memory Aware Synapses [31]) builds on the idea of regularizing important parameters, like EWC, but instead measures parameter importance based on their sensitivity to output activation. By focusing on controlling parameter updates that are critical to maintaining memory of old tasks, MAS offers more robust performance, especially when there are large differences between new and old tasks.

EEIL (End-to-End Incremental Learning [32]) addresses a specific issue within incremental learning: task-time bias, where models tend to favor recently learned tasks. EEIL introduces a balanced training step at the end of each phase, ensuring equal representation across all classes. Additionally, it balances the loss between old and new tasks, preventing overfitting to the new data and maintaining consistency between past and present tasks.

III. EXPERIMENTS

Datasets. We conducted our experiments using three datasets from the MedMNIST collection [35] for multi-class disease classification: BloodMNIST [36], PathMNIST [37], and OrganaMNIST [38], as used in previous work [39]. All images were normalized and rescaled to 28×28 pixels to enable faster computation and evaluation. These datasets contain 8, 9, and 11 distinct classes, respectively, and were divided into 4 tasks, ensuring the classes are non-overlapping across tasks. The detailed dataset specifications are provided in Table I.

Evaluation Metrics. To quantitatively assess the performance of class incremental learning methods, we adopt two key metrics commonly used in prior approaches: average accuracy and average forgetting. *Average accuracy* is calculated by averaging the accuracy of all previously encountered tasks, including the current task after learning the current task t. It is defined as $\mathrm{Acc}_t = \frac{1}{t}\sum_{i=1}^{t}\mathrm{Acc}_{t,i}$, where $\mathrm{Acc}_{t,i}$ represents the accuracy on task i after learning task t. *Average Forgetting* is measured to capture how much performance degrades on previous tasks after learning a new task t. It is computed as $F_t = \frac{1}{t-1}\sum_{i=1}^{t-1}\left(\max_{j\in\{1,\ldots,t-1\}}\mathrm{Acc}_{j,i} - \mathrm{Acc}_{t,i}\right)$. At task t,

Authorized licensed use limited to: Central South University. Downloaded on December 25,2025 at 03:35:31 UTC from IEEE Xplore. Restrictions apply.

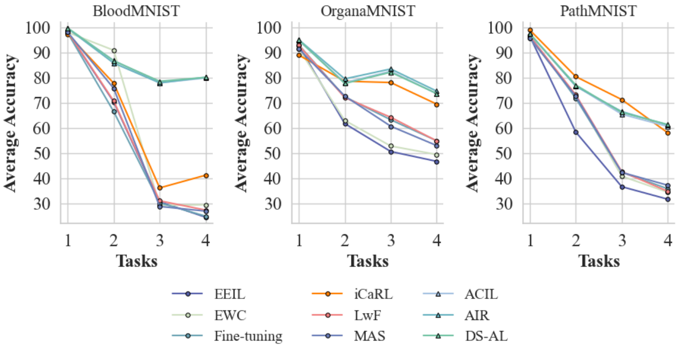

Fig. 2. The AIDC benchmark results for three datasets in terms of average accuracy and average forgetting at each task show that ACIL methods outperform the five baseline methods, particularly on the BloodMNIST dataset.

the forgetting on task i is defined as the maximum difference between the highest accuracy previously achieved on task i and the accuracy on task i after learning task t.

Implementation Details. We employ ResNet18 [33] as the backbone for feature extraction, coupled with task-specific fully connected layers serving as classifiers, to train all methods across the datasets. We follow previous work in [39], each experiment is conducted five times using identical hyperparameters, with different seed values (0, 1, 2, 3, 4) for each run, and the average results are reported. The models are optimized using stochastic gradient descent (SGD) with a batch size of 32 images, trained over 200 epochs, with early stopping applied to mitigate overfitting. To address potential gradient explosion, we apply gradient clipping, capping the maximum gradient value at 10. For replay-based methods, we adopt a herding [40] strategy to select samples, maintaining a fixed memory size of 200. For all ACIL methods, we set the gamma penalty term to 1, and use a random buffer layer with ReLU activation, projecting the extracted features to 8192 dimensions. For DS-AL, we set the compensation ratio to 0.6. Additionally, we establish a lower bound by fine-tuning the model on the current task without applying any specific continual learning strategy.

TABLE I
WE ADHERE TO THE DATA SPLITS AS DEFINED BY MEDMNIST AND IMPLEMENT A CLASS-PER-TASK DIVISION TO FACILITATE CONTINUAL LEARNING, CONSISTENT WITH PRIOR RESEARCH [39].

Dataset	Train	Val	Test	Class	Classes per Task
BloodMNIST	11,959	1,712	3,421	8	[2, 2, 2, 2]
OrganaMNIST	34,561	6,491	17,778	11	[3, 3, 3, 2]
PathMNIST	89,996	10,004	7,180	9	[3, 2, 2, 2]

IV. RESULTS AND DISCUSSION

Results. As shown in Table II and Figure 2, iCaRL outperforms the other four non-ACIL methods, which is consistent with previous work [39]. The three ACIL methods achieve the best results across the three datasets, with AIR standing out as the most prominent. On the BloodMNIST dataset, the performance of the ACIL methods is particularly notable. Specifically, in terms of average accuracy, AIR exceeds the best performing classical method (iCaRL) by 38.95%, and in terms of average forgetting, ACIL exceeds the best performing classical method (iCaRL) by 28.19%.

TABLE II
THE AIDC BENCHMARK RESULTS FOR THE ACIL METHODS AND FIVE PROVIDED METHODS ARE EVALUATED IN TERMS OF BOTH AVERAGE ACCURACY AND AVERAGE FORGETTING AT LAST.

	Class Incremental Learning					
Method	**BloodMNIST**		**PathMNIST**		**OrganaMNIST**	
	Acc ↑	Forg ↓	Acc ↑	Forg ↓	Acc ↑	Forg ↓
Fine-tuning	24.87	93.88	54.87	50.2	35.78	77.77
EWC [16]	24.41	66.41	49.38	54.58	34.40	76.76
MAS [31]	43.94	69.43	34.22	74.98	44.99	55.13
LwF [30]	27.29	91.71	54.83	50.1	34.76	79.15
EEIL [32]	24.46	94.86	46.68	61.4	31.62	79.04
iCaRL [4]	41.17	37.83	69.44	15.85	58.05	17.83
ACIL [23]	79.84	**9.64**	73.55	9.06	60.55	11.56
DS-AL [29]	80.07	11.58	73.67	10.09	**61.33**	14.77
AIR [24]	**80.12**	11.74	**74.61**	**8.42**	61.15	**9.73**

Discussion. The superior performance of AIR can be explained by the fact that the three datasets used are imbalanced. For example, the class ratio of BloodMNIST is approximately $7\colon 18\colon 9\colon 17\colon 7\colon 8\colon 19\colon 14$. AIR addresses this issue by incorporating a re-weighting factor into the loss function, which balances the contribution of each class, thereby producing an unbiased optimal classifier. Additionally, the excellent performance of ACIL methods on the BloodMNIST dataset can be attributed to the fact that the features learned by the backbone in task 1 are transferable across all categories in BloodMNIST. As a result, learning only the classification head in subsequent tasks is sufficient for achieving strong results. Moreover, since ACIL methods can mathematically prove that incremental learning in subsequent tasks is equivalent to joint learning, they are better able to mitigate catastrophic forgetting, leading to overall improved results.

V. CONCLUSION

In this paper, we proposed AIDC (Analytical Learning in Incremental Disease Classification), a benchmark based on ACIL methods to systematically study the problem of knowledge forgetting of CIL in the medical field. By comparing and analyzing with five classic CIL methods on three medical datasets, we found that ACIL methods lead CIL in the medical field in terms of both average accuracy and forgetting. In addition, ACIL methods do not need to store old exemplars, and learning of each new class only requires one epoch. These advantages further show that ACIL has great potential in promoting CIL of disease classification, and has the opportunity to become a new generation of CIL framework for disease classification in the future.

VI. ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (62372474), the 111 Project (B18059). We are grateful for resources from the High Performance Computing Center of Central South University.

Authorized licensed use limited to: Central South University. Downloaded on December 25,2025 at 03:35:31 UTC from IEEE Xplore. Restrictions apply.

References

[1] D.-W. Zhou, Q.-W. Wang, Z.-H. Qi, H.-J. Ye, D.-C. Zhan, and Z. Liu, "Class-incremental learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[2] S. T. Grossberg, *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control*, vol. 70. Springer Science & Business Media, 2012.

[3] L. Wang, X. Zhang, H. Su, and J. Zhu, "A comprehensive survey of continual learning: theory, method and application," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[4] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "icarl: Incremental classifier and representation learning," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 2001–2010, 2017.

[5] A. Chaudhry, A. Gordo, P. Dokania, P. Torr, and D. Lopez-Paz, "Using hindsight to anchor past knowledge in continual learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, pp. 6993–7001, 2021.

[6] M. De Lange and T. Tuytelaars, "Continual prototype evolution: Learning online from non-stationary data streams," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 8250–8259, 2021.

[7] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *Advances in neural information processing systems*, vol. 30, 2017.

[8] H. Shin, J. K. Lee, J. Kim, and J. Kim, "Continual learning with deep generative replay," *Advances in neural information processing systems*, vol. 30, 2017.

[9] R. Gao and W. Liu, "Ddgr: Continual learning with deep diffusion-based generative replay," in *International Conference on Machine Learning*, pp. 10744–10763, PMLR, 2023.

[10] J. Xie, S. Yan, and X. He, "General incremental learning with domain-aware categorical representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14351–14360, 2022.

[11] A. Douillard, A. Ramé, G. Couairon, and M. Cord, "Dytox: Transformers for continual learning with dynamic token expansion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9285–9295, 2022.

[12] Y. Wang, Z. Huang, and X. Hong, "S-prompts learning with pre-trained transformers: An occam's razor for domain incremental learning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 5682–5695, 2022.

[13] K. Lee, K. Lee, J. Shin, and H. Lee, "Overcoming catastrophic forgetting with unlabeled data in the wild," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 312–321, 2019.

[14] J. Zhang, J. Zhang, S. Ghosh, D. Li, S. Tasci, L. Heck, H. Zhang, and C.-C. J. Kuo, "Class-incremental learning via deep model consolidation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 1131–1140, 2020.

[15] F.-Y. Wang, D.-W. Zhou, H.-J. Ye, and D.-C. Zhan, "Foster: Feature boosting and compression for class-incremental learning," in *European conference on computer vision*, pp. 398–414, Springer, 2022.

[16] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the national academy of sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[17] J. Lee, H. G. Hong, D. Joo, and J. Kim, "Continual learning with extended kronecker-factored approximate curvature," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9001–9010, 2020.

[18] J. Leo and J. Kalita, "Survey of continuous deep learning methods and techniques used for incremental learning," *Neurocomputing*, vol. 582, p. 127545, 2024.

[19] S. Hou, X. Pan, C. C. Loy, Z. Wang, and D. Lin, "Learning a unified classifier incrementally via rebalancing," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 831–839, 2019.

[20] B. Zhao, X. Xiao, G. Gan, B. Zhang, and S.-T. Xia, "Maintaining discrimination and fairness in class incremental learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13208–13217, 2020.

[21] H. Zhuang, Z. Lin, and K.-A. Toh, "Blockwise recursive moore–penrose inverse for network learning," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 5, pp. 3237–3250, 2021.

[22] P. Guo, M. R. Lyu, and N. Mastorakis, "Pseudoinverse learning algorithm for feedforward neural networks," *Advances in Neural Networks and Applications*, vol. 1, no. 321-326, 2001.

[23] H. Zhuang, Z. Weng, H. Wei, R. Xie, K.-A. Toh, and Z. Lin, "Acil: Analytic class-incremental learning with absolute memorization and privacy protection," *Advances in Neural Information Processing Systems*, vol. 35, pp. 11602–11614, 2022.

[24] D. Fang, Y. Zhu, R. Fang, C. Chen, Z. Zeng, and H. Zhuang, "Air: Analytic imbalance rectifier for continual learning," *arXiv preprint arXiv:2408.10349*, 2024.

[25] H. Zhuang, Z. Weng, R. He, Z. Lin, and Z. Zeng, "Gkeal: Gaussian kernel embedded analytic learning for few-shot class incremental task," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7746–7755, 2023.

[26] H. Zhuang, R. He, K. Tong, D. Fang, H. Sun, H. Li, T. Chen, and Z. Zeng, "Analytic federated learning," *arXiv preprint arXiv:2405.16240*, 2024.

[27] Z. Liu, C. Du, W. S. Lee, and M. Lin, "Locality sensitive sparse encoding for learning world models online," *arXiv preprint arXiv:2401.13034*, 2024.

[28] R. Aljundi, M. Lin, B. Goujaud, and Y. Bengio, "Gradient based sample selection for online continual learning," *Advances in neural information processing systems*, vol. 32, 2019.

[29] H. Zhuang, R. He, K. Tong, Z. Zeng, C. Chen, and Z. Lin, "Ds-al: A dual-stream analytic learning for exemplar-free class-incremental learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 17237–17244, 2024.

[30] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2935–2947, 2017.

[31] F. Zenke, B. Poole, and S. Ganguli, "Continual learning through synaptic intelligence," in *International conference on machine learning*, pp. 3987–3995, PMLR, 2017.

[32] F. M. Castro, M. J. Marín-Jiménez, N. Guil, C. Schmid, and K. Alahari, "End-to-end incremental learning," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 233–248, 2018.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[34] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[35] J. Yang, R. Shi, and B. Ni, "Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 191–195, IEEE, 2021.

[36] A. Acevedo, A. Merino González, E. S. Alférez Baquero, Á. Molina Borrás, L. Boldú Nebot, and J. Rodellar Benedé, "A dataset of microscopic peripheral blood cell images for development of automatic recognition systems," *Data in brief*, vol. 30, no. article 105474, 2020.

[37] J. N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C.-A. Weis, T. Gaiser, A. Marx, N. A. Valous, D. Ferber, *et al.*, "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study," *PLoS medicine*, vol. 16, no. 1, p. e1002730, 2019.

[38] P. Bilic, P. Christ, H. B. Li, E. Vorontsov, A. Ben-Cohen, G. Kaissis, A. Szeskin, C. Jacobs, G. E. H. Mamani, G. Chartrand, *et al.*, "The liver tumor segmentation benchmark (lits)," *Medical Image Analysis*, vol. 84, p. 102680, 2023.

[39] M. M. Derakhshani, I. Najdenkoska, T. van Sonsbeek, X. Zhen, D. Mahapatra, M. Worring, and C. G. Snoek, "Lifelonger: A benchmark for continual disease classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 314–324, Springer, 2022.

[40] M. Welling, "Herding dynamical weights to learn," in *Proceedings of the 26th annual international conference on machine learning*, pp. 1121–1128, 2009.

Authorized licensed use limited to: Central South University. Downloaded on December 25,2025 at 03:35:31 UTC from IEEE Xplore. Restrictions apply.