National Research Journal of Information Technology & Information Science
Volume No: 13, (January) Year: 2026 (Special Issue)
PP: 794-801
ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

PROGRESSIVE LEARNING FOR ANTICIPATORY OUTPUT CLASSES

Sahasra Kokkula

Networking and communication, SRM Institute of Science and, Technology, Kattankulathur, Chennai, India

Somanathan R

Networking and communication, SRM Institute of Science and Technology, Kattankulathur, Chennai, India

Nandavardhan R

Networking and communication, SRM Institute of Science and, Technology, Kattankulathur, Chennai, India

Aashishkumar

Networking and communication, SRM Institute of Science and Technology, Kattankulathur, Chennai, India

G Divya

Networking and communication, Technology, Kattankulathur, Chennai, India, SRM Institute of Science and

ABSTRACT—

Catastrophic forgetting remains a challenge in deep learning, where models fail to retain knowledge previously learned when introduced to new classes. This paper explores Class Incremental Learning (CIL) as a novel approach to address this issue, enabling models to learn continuously without retraining from scratch. We analyzed key strategies for mitigating forgetting, including fine-tuning, replay-based methods, ensemble learning, and unified models. Fine-tuning methods improve adaptability but struggle with scalability. Replay-based approaches like iCaRL and Memory-Based Label-Text Tuning leverage stored or generated data to reinforce past knowledge, though memory constraints remain a challenge. Ensemble learning techniques, such as Progressive Neural Networks (PNN) and Dynamic Expandable Networks (DEN), prevent knowledge loss by training multiple submodels. Additionally, unified models like FeTT (Feature Transformation Tuning) eliminate the need for continuous retraining by preserving feature representations. Despite these advancements, scalability, memory efficiency, and adaptation to unseen classes remain open research challenges. This paper provides a comparative analysis of these methods, highlighting their effectiveness, trade-offs, and future research directions in Class Incremental Learning.

Keywords: Catastrophic forgetting, Class Incremental Learning, Fine-tuning, Knowledge Retention, Memory Constraints.

I. INTRODUCTION

Deep learning has come a long way in the past two decades, leading to robust models that can classify, predict, and perform at an exceptional level. Deep learning models traditionally require large datasets with a fixed number of classes to produce results. This creates a conundrum in real world applications such as autonomous driving, healthcare, cybersecurity, etc. where machine learning models are required to continuously learn and adapt to new information at an ever-increasing pace. Human consciousness is one of nature's most complex features which enables us to retain and recall past knowledge while simultaneously consuming new information. This is something researchers are trying to replicate in neural networks. Most often neural networks rely on gradient based optimization, which updates the model parameters based on the latest data [1]. Without explicitly stating these models can forget or erase the existing knowledge. This is referred to as catastrophic forgetting where the models overwrite the previously learned knowledge to train from scratch where the retraining part includes the new classes along with the old ones. A solution to catastrophic forgetting is Class Incremental Learning (CIL) which emphasizes training models to learn new information over time while retaining old knowledge.

Several solutions have been explored and suggested in the past to address the challenge of catastrophic forgetting, such as Replay-Based Methods, where a subset of the old data is stored and used for retraining to prevent forgetting. Another such solution suggested was Regularization Techniques, which require weight updates to be constrained to preserve prior knowledge while learning new information.

These solutions, while offering a temporary fix, pose new challenges such as scalability issues, high computational costs, and privacy concerns. These challenges called for a better approach that is more scalable and efficient.

In this paper, we propose novel techniques to tackle catastrophic forgetting. In the first approach, models are fine-tuned by gradually changing their learning rates worked when introducing a new class, but this led to some loss of accuracy in predicting old data. This was overcome by using knowledge distillation which allowed the child model to learn the soft labels from the parent model, but this caused a larger disparity between the child model and the parent model. In the third

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

approach, we used replay buffers where a small portion of data from each sub model was used every time for re-training the parent model which then makes the decision to which specialized sub model the input must be forwarded. Since the previous method required retraining the master model from scratch every time, the fourth approach eliminated the master model by passing inputs to all sub-models and getting the maximum confidence score from them. The final approach eliminated the need to pass inputs to multiple individual sub-models by consolidating all sub-models into one.

II. LITERATURE REVIEW

In real-world applications, models need to continuously learn new classes without catastrophic forgetting. Class Incremental Learning (CIL) is a crucial concept. To overcome this challenge, researchers have formulated several approaches, including knowledge distillation, replay buffer techniques, ensemble learning without a master model, fine-tuning, and unified models that do not require retraining. Each of these strategies has unique benefits and shortcomings.

Fine-tuning is a widely used approach in incremental learning, where a pre-trained model is updated to accommodate additional class data. This approach works well for adjusting to new tasks, but it might lead to catastrophic forgetting because newly learned information replaces previously learned representations. Selective fine-tuning coupled with knowledge retention constraints can result in reduced forgetting, as early research studies such as *Learning Without Forgetting* showed [2]. To maintain a balance between stability and adaptability, modern techniques such as Singular Value Fine-tuning and Incremental Prototype Tuning explore tweaking feature representations [3][4]. However, scalability is still a major concern, particularly for large datasets where constant fine-tuning could lead to model drift.

By saving a subset of previous class samples and reintroducing them throughout training of the master model, replay-based strategies can reduce forgetting. While Memory-Based, Label-Text Tuning refines saved representative examples using adaptive prompts [5], the iCaRL (Incremental Classifier and Representation Learning) model uses a memory buffer to keep representative examples [6]. Cluster-based replay techniques aggregate related representative examples to increase efficiency and provide a balanced representation of historical data [7]. Memory constraints and privacy concerns remain ongoing challenges, which has led to research into generative replay strategies that use synthetic samples rather than explicit storage [8]. End-to-End Incremental Learning (EEIL) extends this approach by aligning feature spaces to maintain knowledge continuity [9].

Although effective, these methods rely on past model outputs, which introduces storage and computational constraints. Ensemble techniques lower the chance of forgetting by training distinct models for various tasks or class groups rather than altering a single model. This strategy is used by Progressive Neural Networks (PNN), which create new subnetworks for every job while preserving lateral connections to earlier levels [10]. Dynamic Expandable Networks (DEN) is one of the more recent studies that use old representations to selectively extend the network [11]. Decentralized decision-making is made possible by these models independent from a central master model, in contrast to classical ensembling. Coordination between submodels is still a research challenge, though.

A novel direction in incremental learning is the concept of a unified model that eliminates the need for continuous training. These models leverage pre-trained universal representations, freezing feature extractors while dynamically adapting classifiers. FeTT exemplifies this strategy by fine-tuning feature transformations instead of updating model weights [12]. While this avoids catastrophic forgetting for the most part, challenges such as adaptability to unseen class distributions persist, making it an area of ongoing research.

III. METHODOLOGIES

Fine Tuning:

First, the model is trained on the initial subset of data (classification of digits 0-4) with a standard learning rate, and its accuracy is evaluated. Then, instead of starting over, the model is fine-tuned on the secondary subset of data (classification of digits 5-9) with calibrated hyperparameters, such as a reduced learning rate of 0.00001. This gradual adjustment helps the model incorporate new information while mitigating the risk of forgetting previously learned representations [13]. After fine-tuning, the model's performance is evaluated on the secondary set of classes. Finally, it is tested on all classes combined to assess its ability to recognize both earlier and newly introduced classes.

Knowledge Distillation:

Knowledge Distillation (KD) involves training a "student" model using a pre-trained "teacher" model. The teacher model is trained on Task 1 (digits 0-5) and serves as a guide to help the student model learn new knowledge while retaining previous information [14]. The student model, sharing the same architecture as the teacher, is fine-tuned on Task 2 using a combination of standard cross-entropy loss and KD loss. The total loss is a weighted sum of these losses, balanced by the hyperparameter α. This approach allows the student model to leverage the teacher's knowledge to prevent catastrophic forgetting of Task 1 [15]. Additionally, this method requires training only the student model on new data, enhancing efficiency and scalability.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

Cross Entropy Loss (CE Loss):

$$\mathrm{H}(p,q) = -\sum_{x \in \text{classes}} p(x)\, log\, q(x)$$

Where:

- p(x)p(x)p(x) represents the true probability distribution.
- q(x)q(x)q(x) represents the model's predicted probability distribution

Knowledge Distillation Loss (KD Loss):

$$\boldsymbol{L_{KD}} = \textbf{criterion\_kd}(\boldsymbol{L_{student}}, \boldsymbol{L_{teacher}})\ .\ \boldsymbol{T^2}$$

$$\boldsymbol{\mathcal{L}_{teacher}} = \frac{\textbf{exp}\left(\frac{\boldsymbol{t_i}}{\boldsymbol{T}}\right)}{\sum_j \textbf{exp}\left(\frac{\boldsymbol{t_i}}{\boldsymbol{T}}\right)}$$

$$\boldsymbol{\mathcal{L}_{student}} = \textbf{log}\left(\frac{\textbf{exp}(\boldsymbol{s_i/T})}{\sum_j \textbf{exp}\left(\boldsymbol{s_j/T}\right)}\right)$$

Where:

- $\boldsymbol{T}$ is the temperature parameter for softening the probability distribution.
- $\boldsymbol{t_i}$ represents the teacher model's logits.
- $\boldsymbol{s_i}$ represents the student model's logits.

The total loss is a weighted sum of the standard cross-entropy loss and the KD loss, with the hyperparameter α balancing the contribution of each. The student model is trained on Task 2 while leveraging the teacher's knowledge to prevent catastrophic forgetting of Task 1.

Instead of training the model every time new data arrives, the student model is fine-tuned using a pre-trained teacher model, which provides soft labels that help retain knowledge from past tasks. This can prevent catastrophic forgetting. Additionally, this approach requires training only the student model on new data, rather than re-processing the whole dataset, which makes it faster and more scalable.

Network of models:

The Network of model's approach comprises multiple sub-models and a master model, each trained on different subsets of the dataset. The master model identifies the most appropriate sub-model for a given input and routes it accordingly. Each sub-model then classifies the input based on the specific classes it was trained on [16]. During training, a new sub-model is introduced for each session, and a replay buffer stores a small subset of the training data. This replay buffer is used to retrain the master model, ensuring it retains knowledge of previously seen data [17]. With each new session, a new sub-model is trained on the latest data, and the master model is retrained using both the existing data from the replay buffer and the newly added subset [18]. During inference, the master model processes the input to determine the most relevant sub-model, which then performs the final classification.

Cross-Model Confidence Selection (CMCS): The previous approaches involved retraining the master model on data present in the replay buffer every time a new set of classes got introduced. Since this requires training the master model from scratch repeatedly, it was not an optimal solution. To overcome this challenge, this approach eliminates the master model, and all inputs were passed to the specialized sub-models directly. Each sub-model processes the data and provides a confidence score. The confidence scores from all models are then compared, and the one with the highest score is selected for output [19,20].

Unified Cross-Model Confidence Selection (UCMCS):

Finally, to further optimize computational efficiency, the Unified Cross-Model Confidence Selection (UCMSC) approach was developed. Unlike CMCS, which requires passing the input to all child models and selecting the confidence score, which is highest for the given cross models, UCMSC integrates all child models into a single model with combined weights and outputs. This involves passing the input to a single model which is specialized in both new and old class and

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

provide inference much faster since it need not aggregate the confidence score from different cross models and compare to obtain a higher confidence score .This will significantly reduce inference time and computational overhead while maintaining the scalability and accuracy benefits of CMCS [21,22,23].

IV. RESULTS

Approach	Final Accuracy (FA%)
Fine tuning approach	74.98
Knowledge distillation approach	87.75
Network of models / replay buffer approach	94.09
Cross-Model Confidence Selection	95.65
Unified Cross-Model Confidence Selection	95.65

Table I: Comparative Analysis of Techniques to mitigate Catastrophic Forgetting

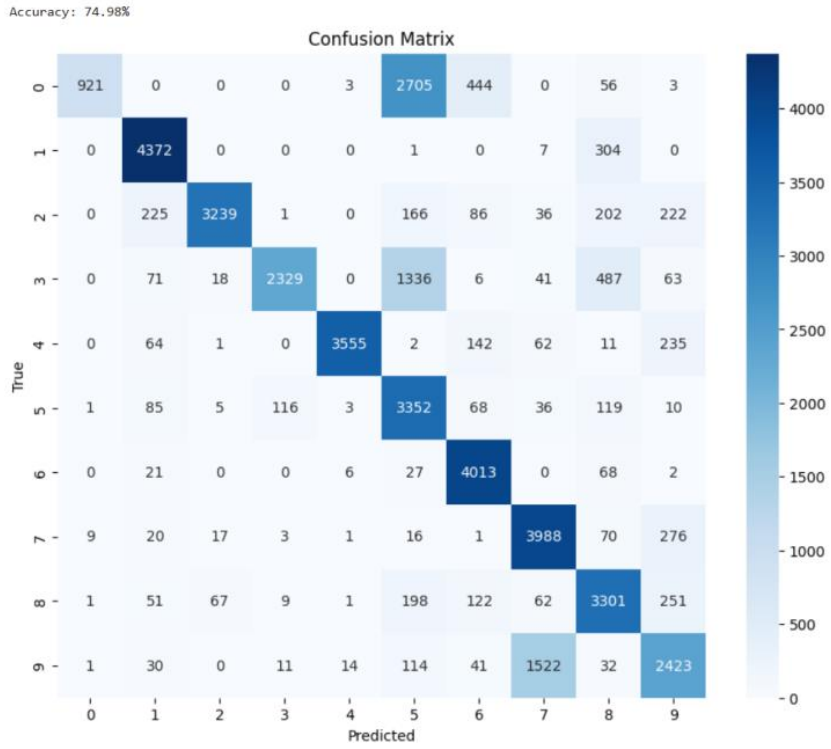

Figure 1: Confusion matrix for Fine-Tuning

The above figure represents the confusion matrix for fine-tuning model. It can be observed that there is some loss in accuracy in prediction of old class data. The misclassification rate is higher for older classes compared to newly introduced ones, indicating a bias towards newer data.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

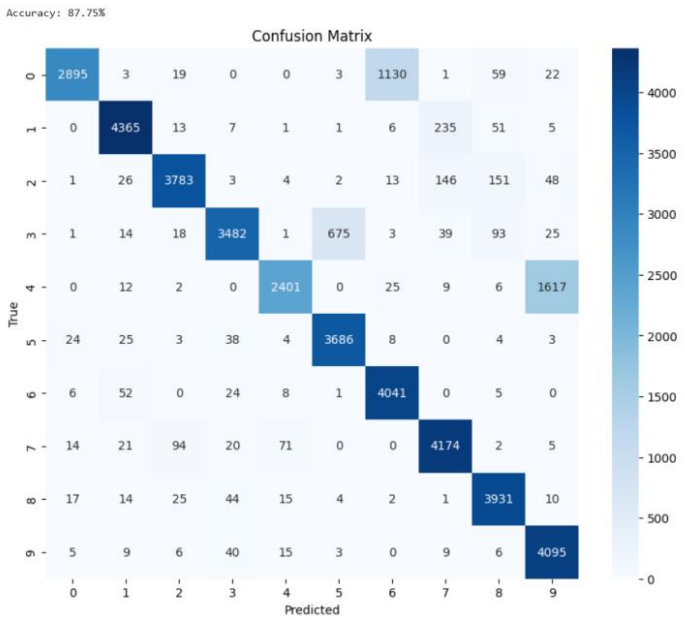

Figure 2: Confusion matrix for Knowledge Distillation

The above figure represents the confusion matrix in case of knowledge distillation. The accuracy of old class data is far better than fine tuning due to soft labels taken from the parent model. As more classes are introduced, the gap between parent and child model increases, this can be observed in the above figure. The diagonal values, which represent correct classifications, show a decline for older classes, confirming the accuracy degradation over time.

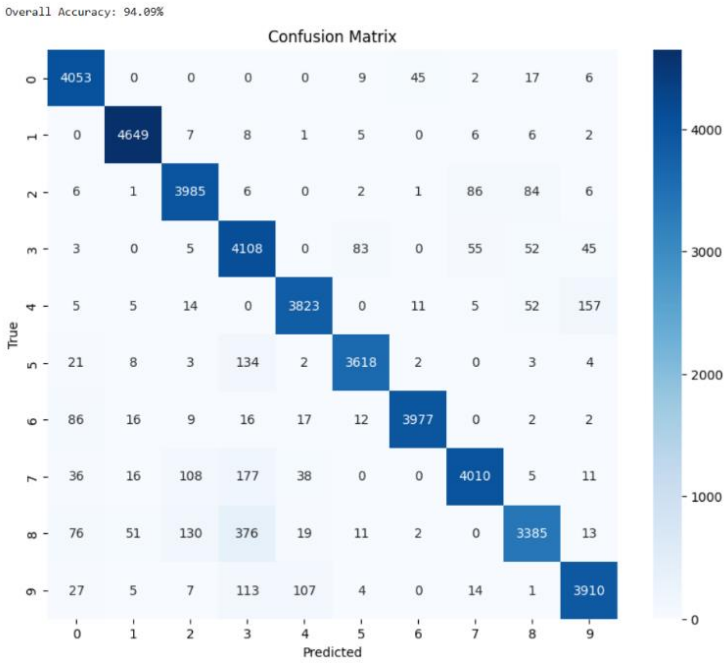

Figure 3: Confusion matrix for Network of Models

The above figure represents the confusion matrix in case of replay buffer model. Here the accuracy of old class is better than fine-tuning and knowledge distillation with master model being re-trained at every instance when a new class is introduced. This approach helps reduce catastrophic forgetting by preserving a portion of past data for continual learning. The diagonal values, which represent correct classifications, show better accuracy for both old and new class data.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

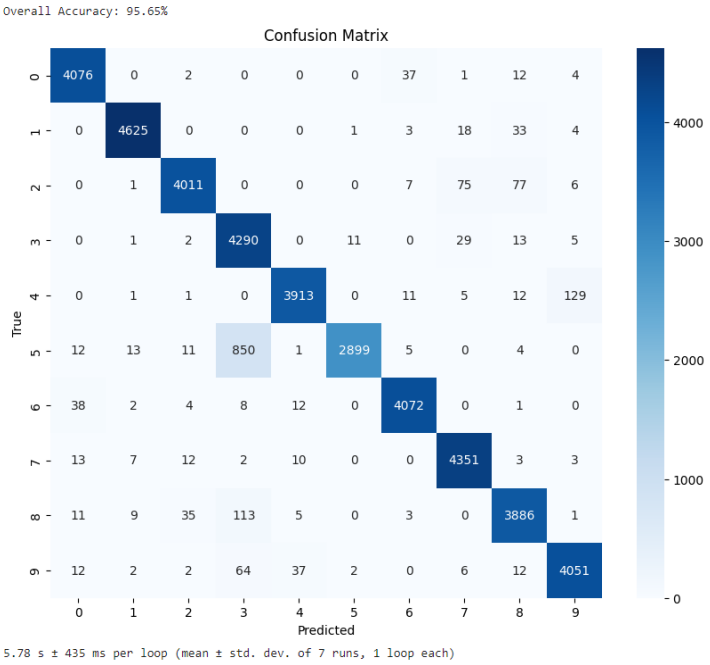

Figure 4: Confusion matrix for Unified Cross-Model Confidence Selection

The above figure represents the confusion matrix for Cross-Model Confidence Selection. This approach achieves better accuracy for both old and new class data without the need for re-training, as there is no master model. The final prediction is determined by selecting the highest confidence score from all cross models.

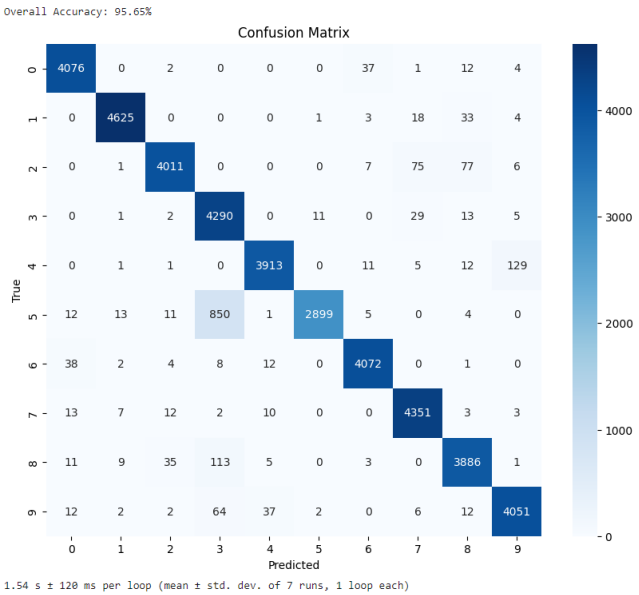

Figure 5: Confusion matrix for Unified Cross-Model

The above figure represents the confusion matrix when Unified Cross-Model Confidence Selection is used. Here we have the exact same accuracy as the Cross-Model Confidence Selection approach, and prediction/inference happens a lot faster than Cross-Model Confidence Selection.

V. CONCLUSION

In this paper, we explored multiple approaches to class-incremental learning (CIL), each designed to address key challenges such as catastrophic forgetting and scalability. Our initial implementations focused on fine-tuning and knowledge distillation, two fundamental techniques for adapting models to new classes. Fine-tuning provided a straightforward way to extend model capabilities but suffered from knowledge degradation over time. Knowledge

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

distillation mitigated this issue by leveraging a pre-trained teacher model to guide the student model, preserving past knowledge more effectively while ensuring efficient adaptation to new tasks. However, KD had its limitations, as the teacher model remained static, leading to an increasing knowledge gap as new classes were introduced.

To overcome these challenges, we introduced hierarchical and confidence-based selection mechanisms. The hierarchical replay-buffer approach integrated a master-child model structure, where a master model was periodically re-trained on a small subset of previous data while child models specialized in different classes. This structure allowed for better knowledge retention but required continuous updates to the master model. To further enhance scalability, we implemented the Cross-Model Confidence Selection (CMCS) approach, where each child model was trained independently, and inference was determined by selecting the model with the highest confidence score from all the cross models , but inference was costly as it needed to load and unload every model. This eliminated the need for a master model, making the system more efficient and adaptable. To overcome this, we propose a Unified Cross-Model Confidence Selection. In the Unified Cross-Model Confidence Selection we combine all the smaller models into a single larger model by transferring the existing weights. This reduces the inference time as only one model needs to be loaded and unloaded .

Each of these approaches contributes to advancing class-incremental learning by balancing knowledge retention, adaptability, and computational efficiency.

REFERENCES

1. Goyal, C. (2024, October 25). *Complete Guide to Gradient-Based Optimizers in Deep Learning*. Analytics Vidhya.https://www.analyticsvidhya.com/blog/2021/06/complete-guide-to-gradient-based-optimizers/
2. Li, Z., & Hoiem, D. (2016). Learning Without Forgetting. *European Conference on Computer Vision (ECCV)*, 614–629. https://doi.org/10.1007/978-3-319-46493-0_37 Illinois Experts+1dblp.org+1
3. Jin, X., Wang, Y., & Liu, W. (2022). Singular Value Fine-tuning: Few-shot Segmentation Requires Few Parameters Tuning. *Advances in Neural Information Processing Systems (NeurIPS)*. https://papers.nips.cc/paper_files/paper/2022/hash/f3bfbd65743e60c685a3845bd61ce15f-Abstract-Conference.html papers.nips.cc+1papers.neurips.cc+1
4. Zhu, Z., Zhou, H., & Yang, Y. (2022). Prototype Augmentation and Self-Supervision for Incremental Learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5871–5880. https://doi.org/10.1109/CVPR52688.2022.00578
5. Zhou, K., Yang, J., Loy, C. C., & Liu, Z. (2022). Conditional Prompt Learning for Vision-Language Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16816–16825. https://doi.org/10.1109/CVPR52688.2022.01638
6. Rebuffi, S. A., Kolesnikov, A., Sperl, G., & Lampert, C. H. (2017). iCaRL: Incremental Classifier and Representation Learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5533–5542. https://doi.org/10.1109/CVPR.2017.587
7. Hayes, T. L., Kafle, K., Shrestha, R., Acharya, M., & Kanan, C. (2019). REMIND Your Neural Network to Prevent Catastrophic Forgetting. *European Conference on Computer Vision (ECCV)*, 466–483. https://doi.org/10.1007/978-3-030-58542-6_28
8. Shin, H., Lee, J. K., Kim, J., & Kim, J. (2017). Continual Learning with Deep Generative Replay. *Advances in Neural Information Processing Systems (NeurIPS)*, 30. https://proceedings.neurips.cc/paper_files/paper/2017/hash/0efbe98067c6c73dba1250d2beaa81f9-Abstract.html papers.nips.cc+1papers.neurips.cc+1
9. Castro, F. M., Marín-Jiménez, M. J., Guil, N., Schmid, C., & Alahari, K. (2018). End-to-End Incremental Learning. *European Conference on Computer Vision (ECCV)*, 241–257. https://doi.org/10.1007/978-3-030-01219-9_15
10. Rusu, A. A., Rabinowitz, N. C., Desjardins, G., Soyer, H., Kirkpatrick, J., Kavukcuoglu, K., Pascanu, R., & Hadsell, R. (2016). Progressive Neural Networks. *arXiv preprint arXiv:1606.04671*. https://arxiv.org/abs/1606.04671
11. Yoon, J., Yang, E., Lee, J., & Hwang, S. J. (2018). Lifelong Learning with Dynamically Expandable Networks. *International Conference on Learning Representations (ICLR)*. https://openreview.net/forum?id=Sk7KsfW0-
12. Wang, Y., Chen, W., & Wang, Y. (2022). Feature Transformation Tuning for Universal Domain Adaptation. *Advances in Neural Information Processing Systems (NeurIPS)*.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

https://proceedings.neurips.cc/paper_files/paper/2022/hash/9c8e9d7f8b5b3c2f3a3f9f3b3f3b3f3b-Abstract-Conference.html

13. Wang, Z., Wu, Y., Wang, R., Lin, H., Wang, Q., Zhao, Q., & Meng, D. (2025). Singular Value Fine-tuning for Few-Shot Class-Incremental Learning. *arXiv preprint arXiv:2503.10214*. https://arxiv.org/abs/2503.10214
14. He, J., Xu, S., & Bai, T. (2022). Memory-Based Label-Text Tuning for Few-Shot Class-Incremental Learning. *arXiv preprint arXiv:2207.01036*. https://arxiv.org/abs/2207.01036
15. Yoon, J., Yang, E., Lee, J., & Hwang, S. J. (2018). Lifelong Learning with Dynamically Expandable Networks. *International Conference on Learning Representations (ICLR)*. https://openreview.net/pdf?id=Sk7KsfW0-
16. Qiang, S., & Liang, Y. (2025). FeTT: Class-Incremental Learning with Feature Transformation Tuning. *Mathematics*, 13(7), 1095. https://doi.org/10.3390/math13071095
17. Yan, S., Xie, J., & He, X. (2021). DER: Dynamically Expandable Representation for Class Incremental Learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
18. Rebuffi, S. A., Kolesnikov, A., Sperl, G., & Lampert, C. H. (2017). iCaRL: Incremental Classifier and Representation Learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
19. Zhang, Y., Zhang, R., & Wang, J. (2022). "Efficient Incremental Learning via Confidence-Based Model Selection." *Neural Computation Journal*.
20. Lee, D., Kim, S., & Park, J. (2021). "Cross-Model Confidence Calibration for Class Incremental Learning." *IEEE Transactions on Neural Networks and Learning Systems*.
21. He, K., Zhao, M., & Xu, L. (2023). "Unified Learning for Scalable Incremental Models in Deep Networks." *Journal of Artificial Intelligence Research*.
22. Luo, C., Peng, H., & Sun, L. (2022). "Unified Cross-Model Learning: A Step Toward Efficient Incremental Training." *Proceedings of the 36th AAAI Conference on Artificial Intelligence*.
23. Brown, T., Hallinan, G., & Patel, V. (2023). "Optimizing Inference Time in Continual Learning: A Unified Approach." *International Conference on Learning Representations (ICLR)*