Characterizing Continual Learning Scenarios for Tumor Classification in Histopathology Images

Veena Kaustaban[1], Qinle Ba, Ipshita Bhattacharya[1*], Nahil Sobh[1], Satarupa Mukherjee[1], Jim Martin[1], Mohammad Saleh Miri[1], Christoph Guetter[1], and Amal Chaturvedi[1]

[1]Roche Sequencing Solutions, Santa Clara, CA, USA. e-mail: qinle.ba@roche.com

Overview

Deep learning has achieved unprecedented performance in solving complex problems in digital pathology (DP) based analysis. There are two challenges in model development for DP applications. First, the increasing volume of data to be analyzed. The increasingly wide adoption of DP along with development of new assays with, for example, the emerging multiplexing technologies, produce a near continuous stream of data. Second, evolving data distributions for modeling, such as varying image digitization conditions and onset of new diseases. Such ever-evolving image data call for iterative model development and model update even after model deployment. Continual learning (CL) aims to alleviate catastrophic forgetting without necessitating iterative model updates with significant resource requirements.

Here, we perform a systematic study on continual learning for digital pathology:

(1) We identify continual learning scenarios that are practical for DP-based analysis

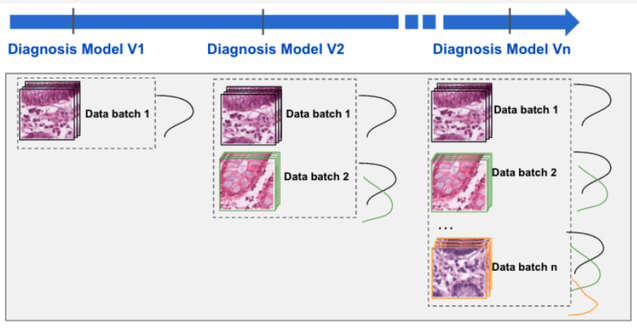
(2) We establish a dataset with augmented H&E images, simulating data streams from multiple data sources

(3) We systematically characterized the performance of recent CL methods

(4) We explored the feasibility of continual learning for multiple tumor types and online continual learning for DP

NEED FOR CONTINUAL LEARNING (CL) IN DP

Efficiently Adapt Diagnosis Solutions to Evolving Patient Data

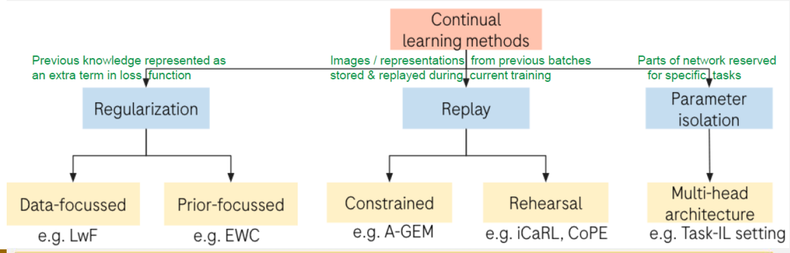


- Conventional methods require iterative training and updates with evolving patient data distributions.
- Continual learning makes deep learning models more flexible and adaptive to new data without training from scratch.

Methods

CONTINUAL LEARNING METHODS & EVALUATION

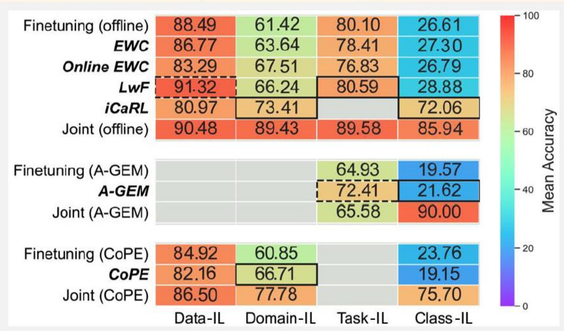
- **We tested methods from three broad categories and report accuracy averaged over test sets of multiple repeats at each experience, including the last experience, i.e. end of training.**
- **All methods were compared against 2 baselines: joint training (upper-bound) and finetuning (lower bound).**



Results

SUMMARY OF CL METHODS COMPARED TO BASELINES

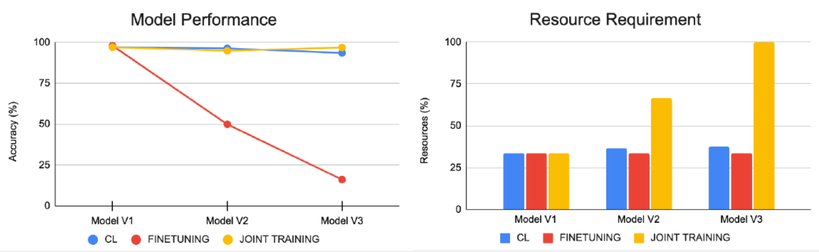
- Columns arranged by difficulty: from easiest Data-IL to the most challenging class-IL
- CL performance most effective for the challenging class-IL scenario
- iCaRL has the best overall performance across scenarios.



Methods

CL PROVIDES PROMISING PERFORMANCE WITH FEWER RESOURCES

- **Joint training provides best performance as the model is iteratively trained with all the available data, thus requiring more data storage and compute resources at each iteration.**
- **Finetuning uses less resources with a compromise in model performance.**
- **CL provides promising performance with fewer resources.**



SIMULATING DOMAIN SHIFTED DATA STREAMS

- **We generated an augmented colorectal cancer (CRC) dataset comprised of multiple subsets, each with distinct colors and stain intensities, simulating commonly observed domain shifts in practice.**

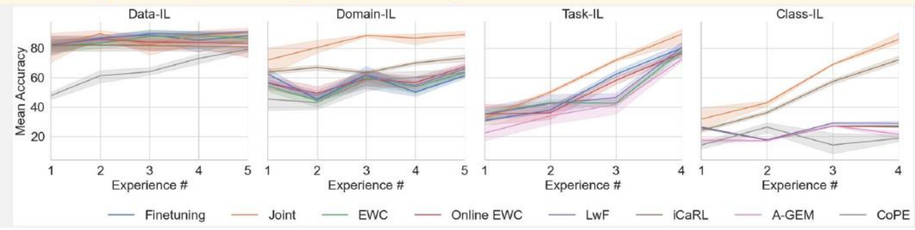
Domain	Augmentation Description	Simulated Data Source	Augmentation Setting
Domain 2	Increased stain intensity	Change in stainer, staining protocol or reagent concentration	• Eosin intensity was increased with a scaling factor sampled from [1.75, 2.75] • Hematoxylin (HTX) intensity scaling factor sampled from [1.5, 2.0]
Domain 3	Decreased eosin stain intensity	Variation in staining or use of an older slide with fading stain	• Eosin intensity scaling factor sampled from [0.4, 2.75]
Domain 4	Change of hue	Change in reagent manufacturer, scanner, stainer	• Eosin hue scaling factor sampled from [-0.05, -0.03] • Hematoxylin hue scaling factor sampled from [0.05,0.08]
Domain 5	Change of hue and saturation	Change in reagent manufacturer, scanner, stainer	• Hue was changed for eosin ([0.03, 0.05]) • Saturation was increased for both stains (eosin by [1.2, 1.4], Hematoxylin by [1.1, 1.3])

Results

CL PERFORMANCE ACROSS EXPERIENCES

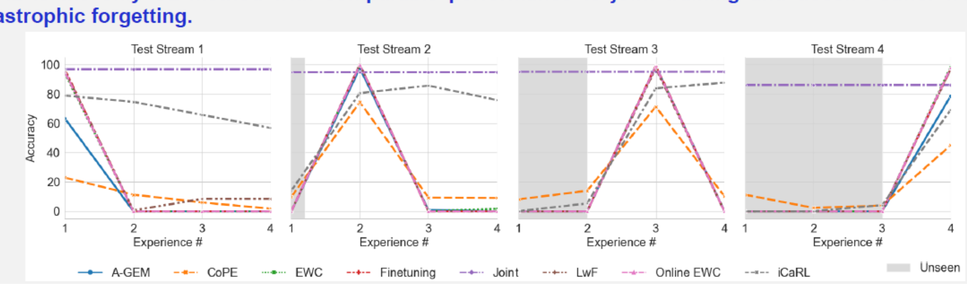
CL Methods across Scenarios against Offline Baselines

- CL performance across experiences shows NO catastrophic forgetting in Data-, Task-IL scenarios
- iCaRL comparable to joint training in Class-IL setting.



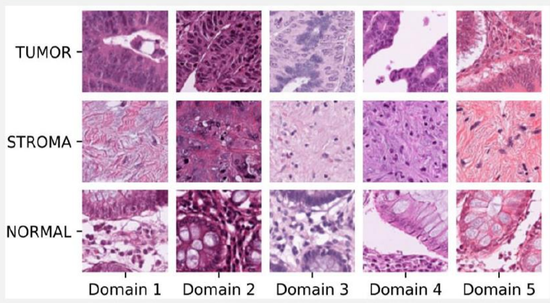
Performance of CL Methods in Class-IL Scenario

iCaRL is the only CL method with comparable performance to joint training and doesn't suffer from catastrophic forgetting.



Methods

AUGMENTED COLORECTAL CANCER (CRC) DATASET



- Example Augmented CRC images.
- For each of the 5 domains, we show 1 example image per class for 3 out of the 9 classes.

CONTINUAL LEARNING SCENARIOS IN DP

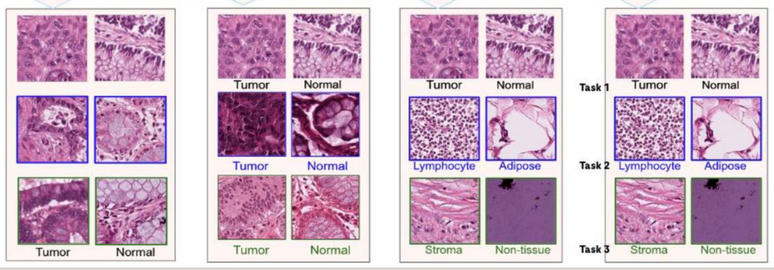
Four CL scenarios were identified for DP applications based on how newer streams of data different from previous streams and how data from different augmented domains are distributed between data batches. Shown below are illustrations for CL scenarios with two classes - tumor and normal.

Data-IL: each data batch contains equal number of images from *all* domains

Domain-IL: each data batch contains equal number of images from *one* domain

Class-IL: each data batch contains images from new classes from *all* domains

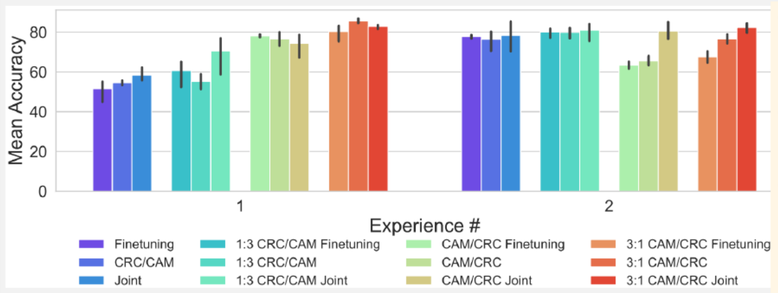
Task-IL: each data batch contains images from one task from new classes from *all* domains



Results

CONTINUALLY LEARNING FROM MULTIPLE TUMOR TYPES

We evaluated the performance of LwF in Domain-IL with dramatic domain shift of CRC data in one experience and breast tissue (PatchCam) in a 2nd experience



- We experimented with training order and sample size of the two tumor datasets.
- Best training order: PatchCam first followed by CRC
- Using 3 times more PatchCam data improved accuracy by 11%.

Conclusions

- We propose evaluation approaches and domain-shifted dataset as CL benchmark
- CL methods are computationally efficient, taking only ~28% of the runtime as joint training.
- iCaRL, a replay-based method, was most effective across scenarios.
- Surprisingly, Task-IL scenarios may not be as easy to learn for DP as reported elsewhere.

References

1. Van de Ven, G.M., Tolias, A.S.: Three scenarios for continual learning. arXiv preprint arXiv:1904.07734 (2019)
2. De Lange, M., Aljundi, R., Masana, M., et al. (2022). A Continual Learning Survey: Defying Forgetting in Classification Tasks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44, 3366-3385.
3. 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)