Dynamic AI in the Clinical Open World

DAICOW

MICCAI 2023

The 1st MICCAI tutorial on dynamic and continual learning for AI in medical imaging

OCTOBER 12th, 2023

Summary

Have you ever worked with data from five, or even ten years ago? You probably noticed how different it is from more recent cases. If you did not, *your model definitely did*.

In medical imaging, there are many factors that cause the data distribution to change over time, which is reflected in *abrupt performance drops* in deep learning models. We may reduce this effect through domain adaptation and improved generalization, but a point will inevitably come where we need to *update our model with new data*. And when that happens, we may no longer have access to all training data used in the past.

Continual learning addresses precisely this question by developing strategies that acquire new information without losing previous knowledge. This opens up attractive possibilities, such as extending the lifespan of medical software solutions and leveraging large amounts of multi-institutional data without the need for federated protocols. Yet deploying continual learning solutions comes with a number of *technical and legal challenges* and requires *additional quality monitoring*.

This tutorial covers potential sources of domain shift, continual learning metrics and strategies, and regulatory guidelines. You will be assigned a group working on a particular medical imaging task, and each group will take the role of a company looking to release an AI solution that learns over time.

By combining expert presentations on key topics and practical assignments where participants can engage with the subject hands-on, we hope to convey the core opportunities and challenges of releasing lifelong learning solutions.

Part 1: Data drift in medical imaging

Common sources of domain shift and their effect on model performance

Jayashree Kalpathy-Cramer

University of Colorado and Harvard Medical School

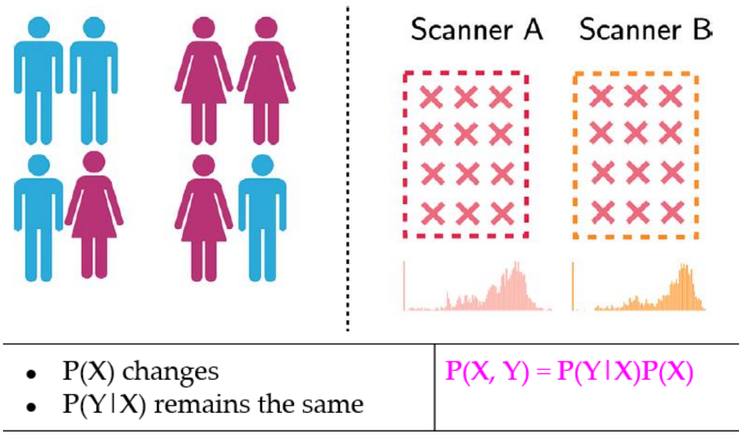
Jayashree Kalpathy-Cramer leads the Artificial Medical Intelligence Division at the University of Colorado ophthalmology department. She has previously co-directed the Quantitative Translational Imaging in Medicine lab and the Center for Machine Learning at the Athinoula A. Martinos Center at Harvard Medical School. In addition to developing novel machine learning algorithms, she is actively engaged in applying these to clinical problems in radiology, oncology and ophthalmology. She was funded through NIH to develop quantitative imaging methods in cancer and is the PI of an NSF-funded project to develop and apply algorithms to build diagnostic tools in ophthalmology. Research from this work has resulted in a deep-learning based algorithm for disease diagnosis and response assessment that is currently being evaluated at several clinics and screening trials in the US and India.

It is becoming increasingly easy to create an medical imaging ML algorithm today. But it continues to be difficult to create an ML algorithm that is:

- Not brittle
- “Self aware”
- Explainable
- Unbiased/Fair

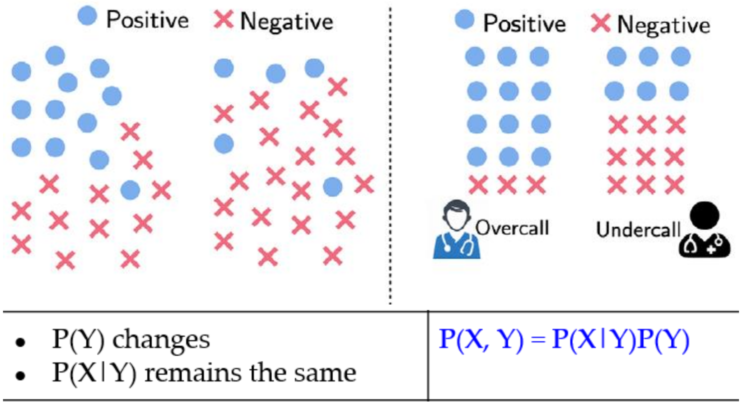
Drifts can be common in medical imaging. These include:

Covariate Shift: distribution of the input data changes. Occurs when the cohort selected to train the model is not representative of the target population. The shift in patient population demographics such as age and sex or the differences in image acquisition such as resolution, hardware, modalities, sequences, or use of contrast material may also include covariate shift.

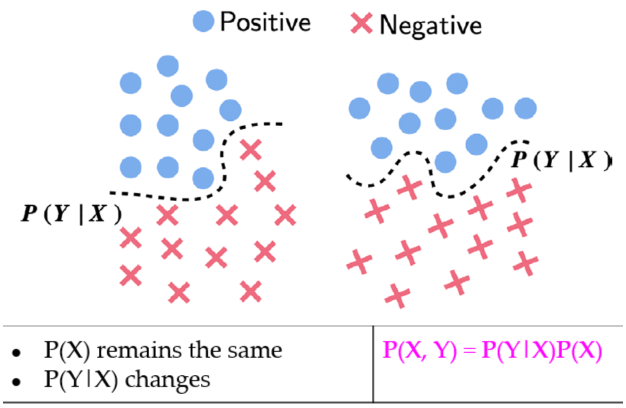


Covariate shift. Gupta, Kumar et. al, Radiographics, 2023.

Label Shift: probability of the target label changes. A label shift can manifest through differences in the disease prevalence, characteristics, or severity (e.g. during the different phases and variants of the COVID-19 pandemic). Furthermore, annotations may vary among experts (radiologists), creating label shift.



Label shift. Gupta, Kumar et. al, Radiographics, 2023.



Concept drift. Gupta, Kumar et. al, Radiographics, 2023.

Concept drift: occurs when the distribution of the input remains the same, but the conditional probability of the target label for any given input changes. It is characterised by changes in the statistical properties of the target variables after the deployment of the model. For e.g. For example, changing the diabetes diagnostic criteria for fasting plasma glucose may change the meaning of automatically derived labels.

Part 2: Continual Learning

Pillars of Forgetting & Lifelong Evaluation

Martin Mundt

Technical University of Darmstadt and hessian.AI

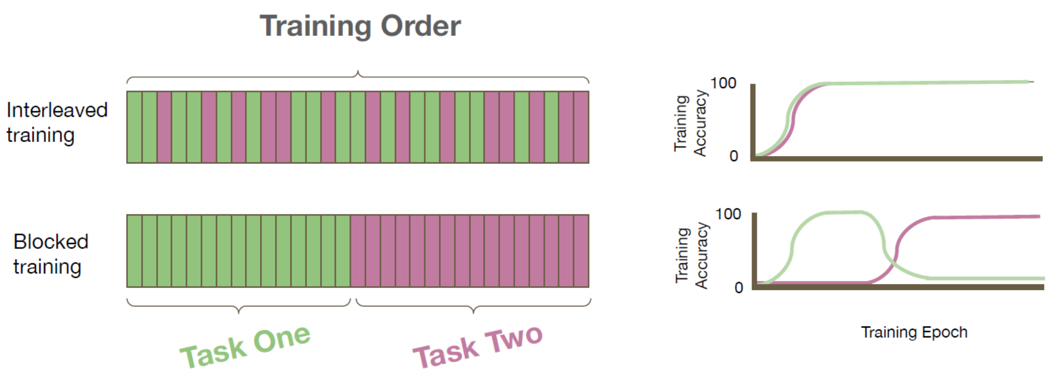
Dr. Martin Mundt is a hessian.AI junior research group leader of the Open World Lifelong Learning lab, where the focus is to create robust systems that learn continually in an open-ended world. He is also a board member of directors at the non-profit ContinualAI organization for the 2022-2024 election term. He obtained his PhD in continual deep learning at the Goethe University Frankfurt, for which he received the best thesis in natural sciences award.

If humans & animals learn continually, why shouldn't our machines? At the least, lifelong learning may be one pathway to more human-like intelligence At the most, it's one pathway towards stronger artificial intelligence.

In the meantime, lifelong learning has direct benefits towards improving AI systems across research & real-world deployment:

- Efficiency and Scalability
- Fairness, Privacy & Security
- Robustness and Accuracy

Despite the achievements of many AI systems, few, if any, truly can learn continually over time. Among other causes, this is due to relying on narrow, fixed models, lacking robustness. Connectionist models fail to learn sequentially. When they are trained with a sequence of task one after the other, they tend to adapt too strongly to the last seen distribution and *catastrophically forget* previous knowledge.



The sequential learning problem. Adapted from Flesch et al, 2022.

This problem does not only affect deep neural networks. While most commonly associated with deep learning, catastrophic interference applies to a much broader class of algorithms, including SVMs and linear regression.

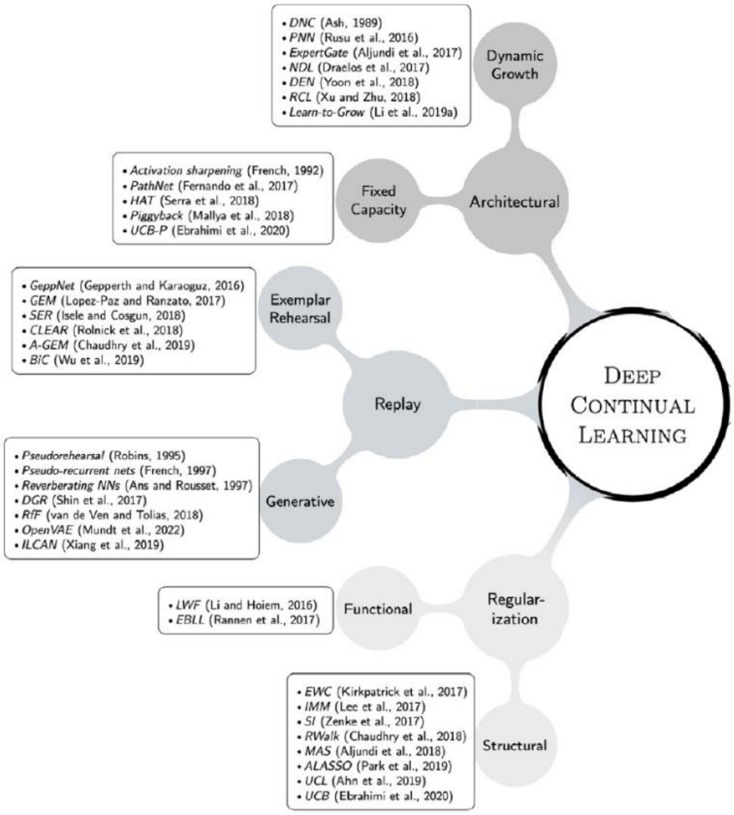
How can be train models so they capture new information without forgetting previous knowledge? Some strategies are:

Replay/rehearsal: Save samples of each tasks' data in (external) memory buffer and progressively replace parts of memory buffer with new examples (may violate privacy constraints!).

Generative (pseudo-) reply: Don't memorize samples direc-tly; instead, memorize their exemplars and replay generated samples instead.

Regularization: Don't greedily optimize for a new task, preserve the (important) old weights by penalizing updates to already learned parameters.

Architectural: Change the macro or micro architecture of the network, e.g. add new parameters or prevent repre-sentations from overlapping.



Overview of deep continual learning methods. Mundt et al., "Wholistic View of Continual Learning with Deep Neural Networks", Neural Networks 160, 2023.

These strategies all have advantages and disadvantages. So, how do we evaluate lifelong learning? Some metrics are:

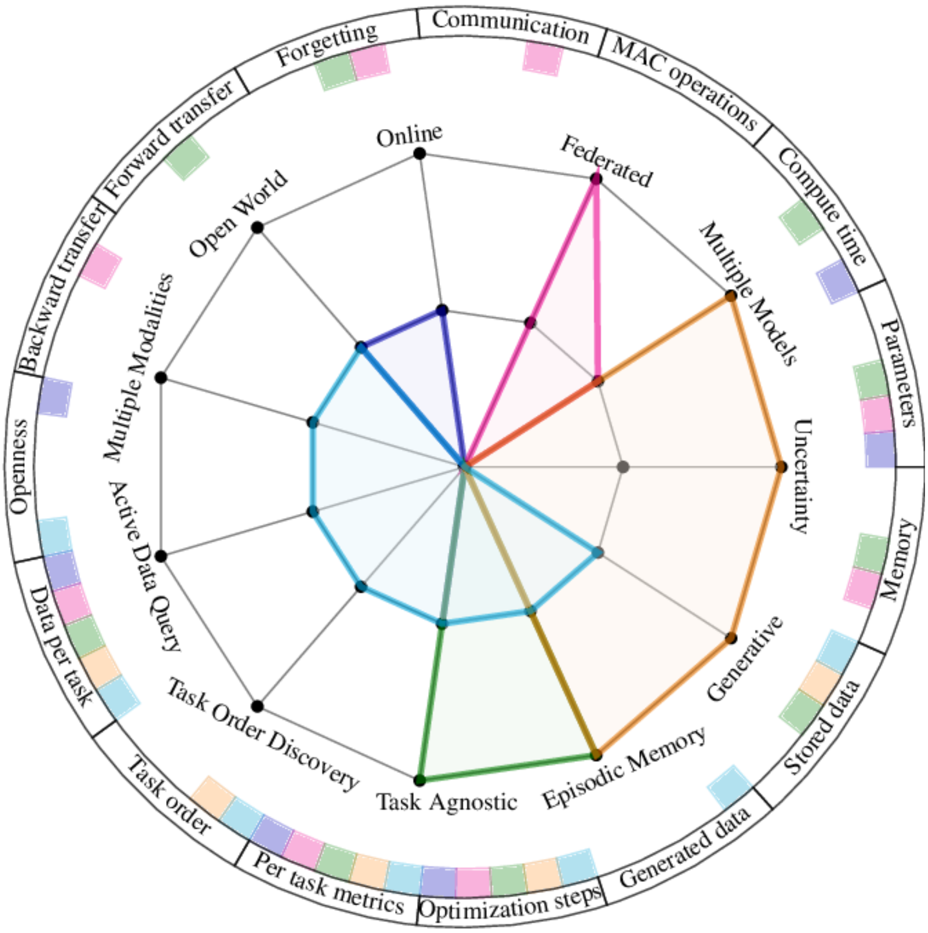
- **Per-task average loss:** average of all tasks up to the present point in time → Measure present overall performance
- **Forward transfer:** influence of a learning task on future tasks
- **Backward transfer:** influence of a task on previous tasks; negative = forgetting, positive = retrospective improvement

We also need to consider other factors, such as storage capacity, memory, size and computational requirements.

So, how do we compare & draw conclusions with various metrics + set-ups?

The differences between machine learning paradigms with continuous components can be nuanced. Key aspects often reside in how we evaluate Each paradigm seems to have a particular preference (potentially neglecting other important factors)

Distinct applications warrant the existence of numerous scenarios, but as far as possible, we must promote comparability!



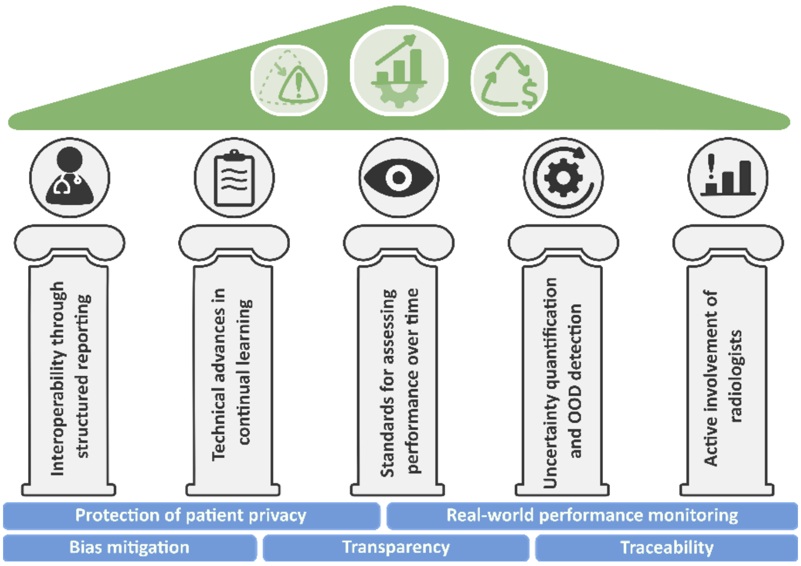
Continual learning desiderata. Mundt et al, "CLEVA-Compass: A Continual Learning Evaluation Assessment Compass to Promote Research Transparency and Comparability", ICLR 2022.

Part 3: Regulatory Aspects of Continual Learning

Current landscape in the USA and EU and a look into the future

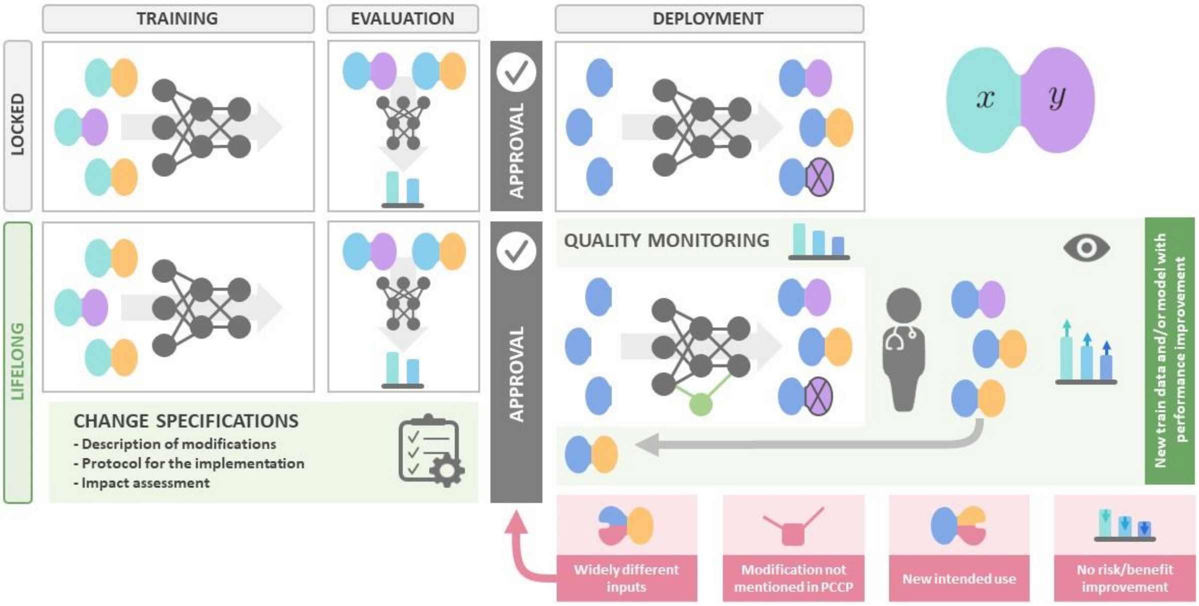
We have explored technical mechanisms to adapt deep learning models. However, AI-powered *Software as a Medical Device* (SaMD) must be approved by the responsible regulatory entities in order to be used in clinical practice. Updating the model often requires re-certification, which is a lengthy and expensive process.

Fortunately, regulatory bodies around the world are moving away from the static learning paradigm and towards a **lifecycle regulatory protocol**. In particular, directives currently in force or under consideration in the USA and European Union address systems that *continue to learn* after deployment. These include the *harmonized rules for Artificial Intelligence* (AI Act) proposal for European member states and the FDA's *draft guidance in drafting a Predetermined Change Control Plan* (PCCP) in the United States.



From bottom to top: (1) Requirements set by recent regulatory efforts in EU and USA, (2) building blocks for effectively designing dynamic systems, and (3) potential benefits in the form of risk reduction, performance increase, and sustainable business models.

Though different in certain aspects, both pose similar requirements and follow the same concept: when seeking approval for a SaMD, manufacturers can submit documentation on the planned changes that aim to improve the risk/benefit ratio of the product without affecting its intended use. As long as future modifications follow the specifications outlined in the protocol and fulfil the base requirements, they can be rolled out without re-approval.



Approval process for a deterministic locked AI system (top) vs. a lifelong learning system (bottom). In the second case, a description of the planned modifications and protocols outlying how they will be implemented and monitored are submitted as additional documentation during initial clearance. Modifications following these specifications, such as re-training the model with additional training data, do not require re-approval.

Thank you for participating!

We hope that the tutorial helped you adapt your workflow for continual training and quality monitoring! Follow us on **@ContinualMedAI** on Twitter/X for future updates, and **join our Slack workspace** if you wish to continue the conversation.

We hope you enjoy the rest of your time in Vancouver!

Camila Gonzalez

Stanford University

Anirban Mukhopadhyay

Technical University of Darmstadt

Maria Zuluaga

EURECOM, 3IA Institute Côte d'Azur and King's College London

Ilkay Öksüz

Istanbul Technical University and King's College London

Niklas Babendererde

Technical University of Darmstadt

Khrystyna Faryna

Radboud UMC

Magda Paschali

Stanford University

John Kalkhof

Technical University of Darmstadt

Moritz Fuchs

Technical University of Darmstadt