Online Learning for Multi-Label Medical Images

Andrea Dunn Beltran, Cameron Wolfe, Anastasios Kyrillidis, Dept. of Computer Science, Rice University, *Dept. of Computer Science, UNC
asdunnbe@ad.unc.edu crw13@rice.edu anastasios@rice.edu

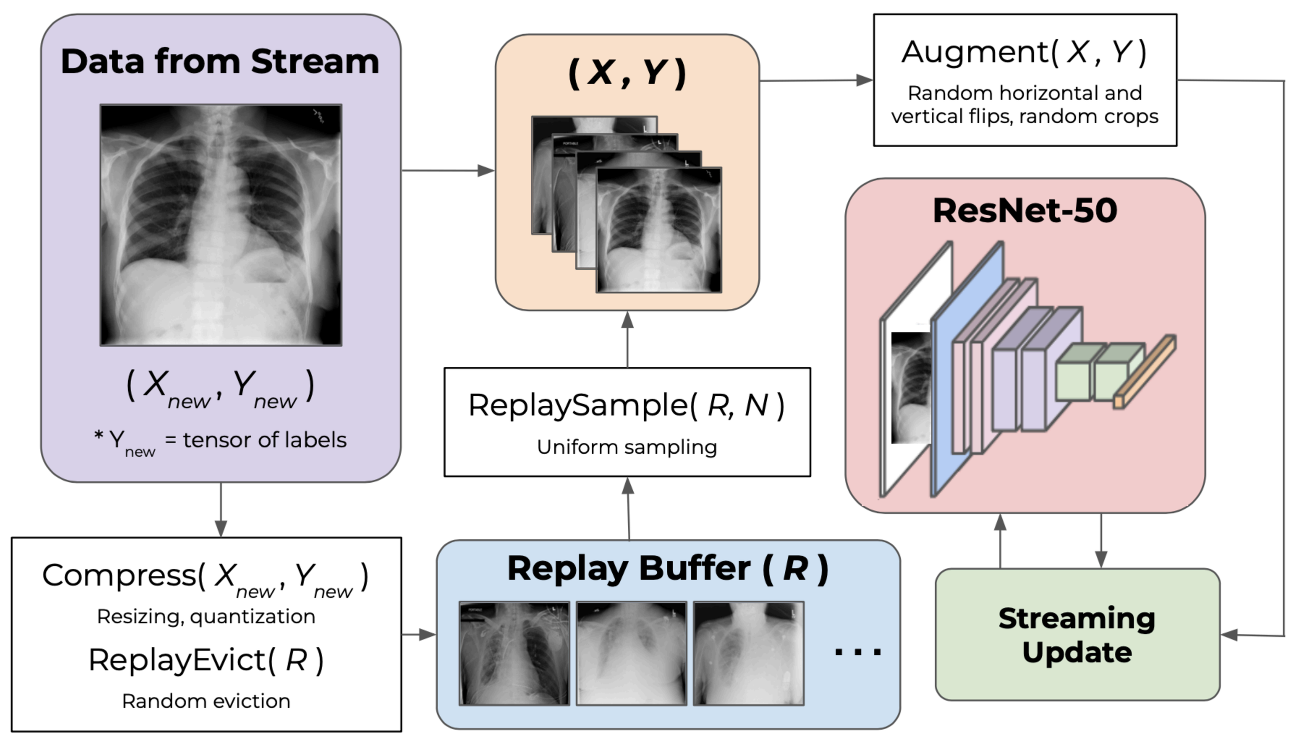
Background & Motivation

What is Online Learning?

- **Online Learning** is a machine learning method in which the full dataset is never available to the model at once and the data becomes available in a sequential order
- It is useful when…
  - It is computationally impractical to train over the entire dataset
  - An algorithm must adapt dynamically to new patterns
  - Data generated as a function of time
- We must address **Catastrophic Forgetting,** when an online model 'forgets' how to classify previous data as it is exposed to new data.

The CSSL Algorithm

- In the **CSSL** Algorithm, a unique data example from the *data stream* is compressed then sent to the *replay buffer*. If the replay buffer if full, an example is deleted
- The sample is added to the *collection of replay samples* along with a random sample from the replay buffer of size *N*
- The collection is then *augmented* and the streaming update changes the learner parameters of the initiated model



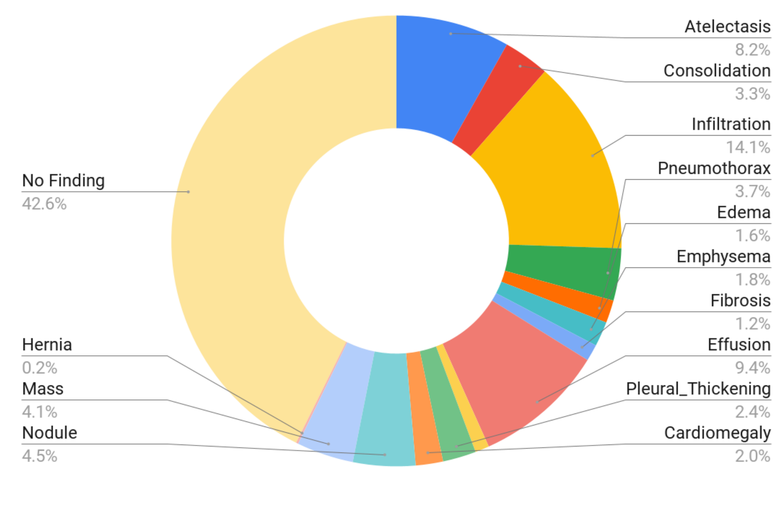
The Reality of Medical Data

- Medical Image sets are typically **small**, take a long time to accumulate, and require an expert to annotate.
- Inconsistent between sources
- **High Fidelity** is especially important in this setting
- If we can predict an image's label before it reaches the expert, it will **speed up** the labeling process. The model will learn from mistakes waste of the doctors' time
- Offline training wastes computation and does not **dynamically** learn well in contrast to Online training.

Methodology & Contribution

NIH Chest X-Ray Dataset

- Collection of **112,000+** x-ray images
- Over **60,000** 'No Finding' examples and **50,000** images with at least 2 labels
- X-Rays are only 1 channel so the **ResNet50** model must be changed
- Used to explore the effects of vastly **unbalanced** data

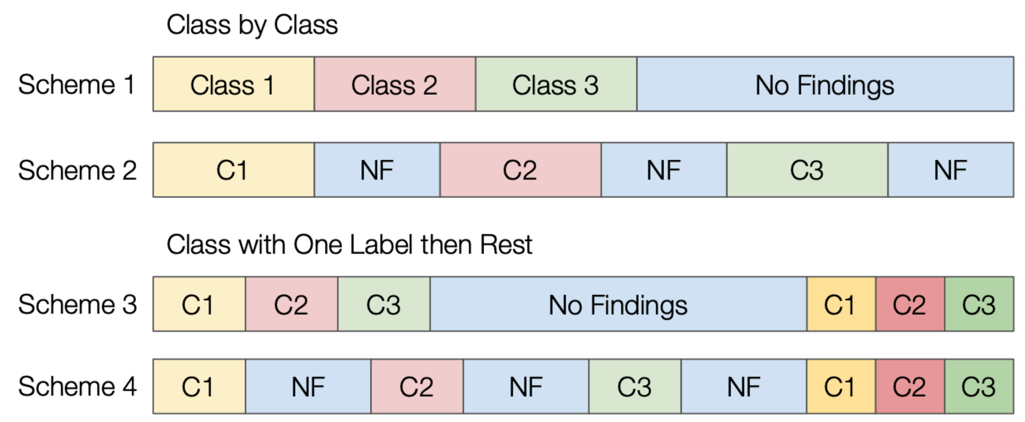


Changes to CSSL

- To apply CSSL to medical data, we must use traditional medical image data augmentation (horizontal and vertical flips, random crops, rotation)
- Consider how to pass "classes" when they are not mutually exclusive
  - How do we treat 'No Finding' examples
- How to handle learning rates given that there are no clear classes

Index Schemes

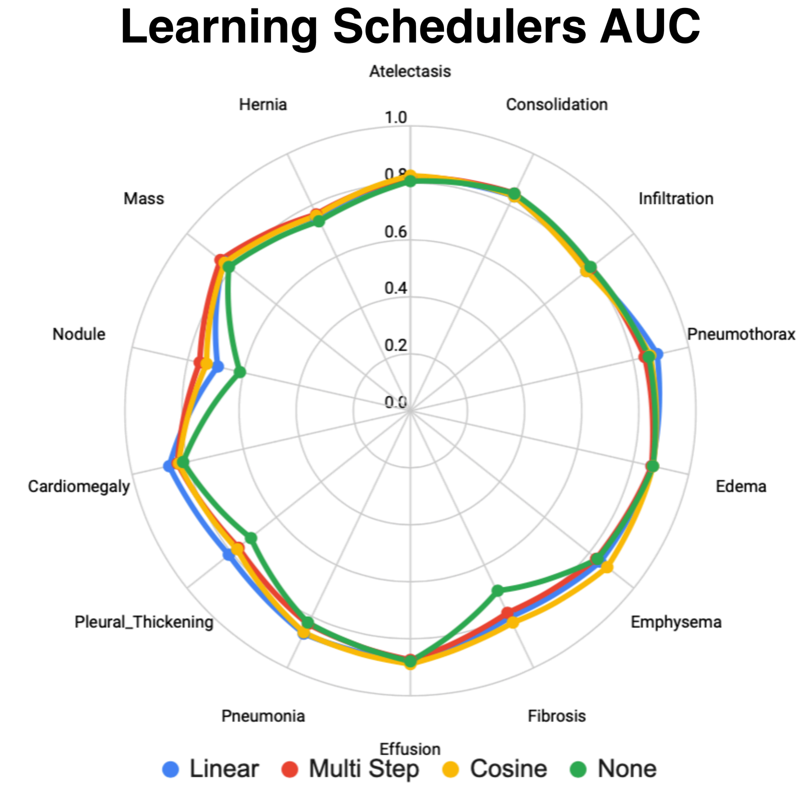
- In the case of **Multi-label Classification**, classes are not mutually exclusive so training is less intuitive
- Index schemes are used to fed the model "classes" at a time
- Schemes 1 and 2: Classes of pathologies as passed together as a group without revisiting previous data
- Schemes 3 and 4: Each class with only one label is passed first then multi-labels are passed in
- Schemes 2 and 4: Distribute the 'No Finding' examples



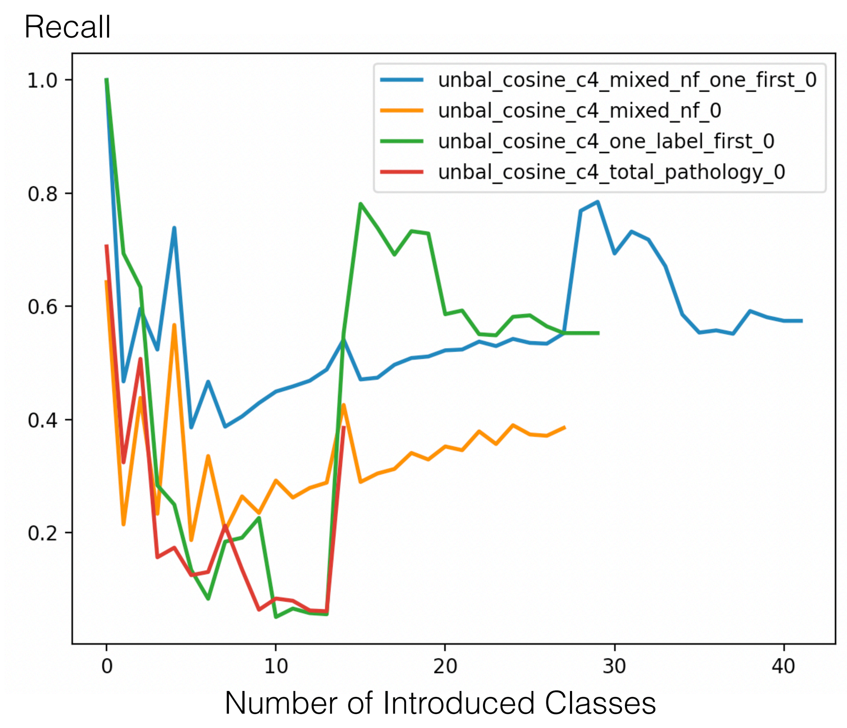
Results & Conclusions

Baselines

- Implement baselines with different learning schedulers
- Does not consider 'No Findings' as a pathology
- Area under **ROC** curve: measure across all possible thresholds with parameters (True Positive Rate, False Positive Rate)
- Baselines were more refined however the CSSL model is still in preliminary testing.

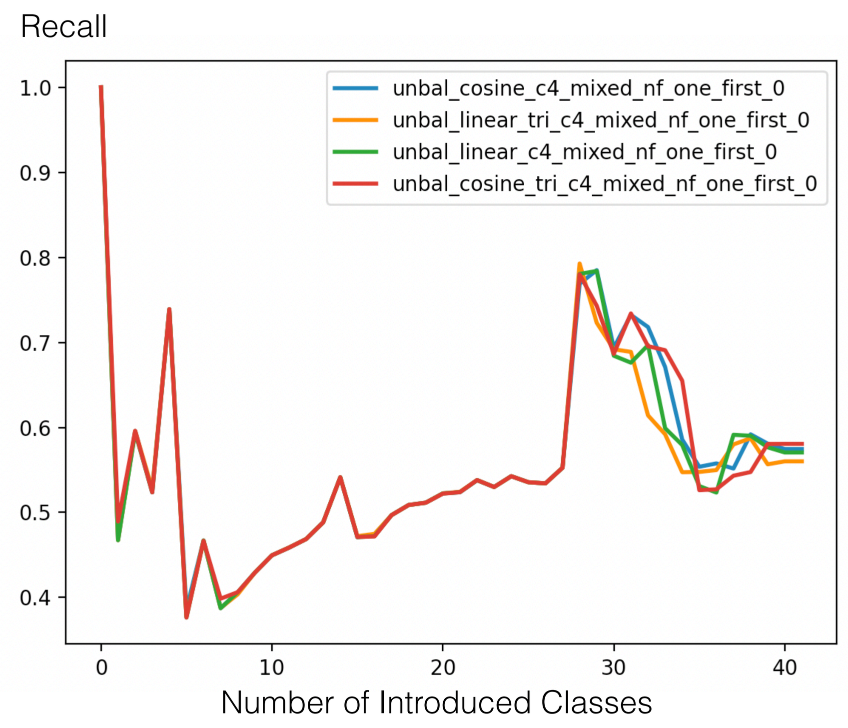


Index

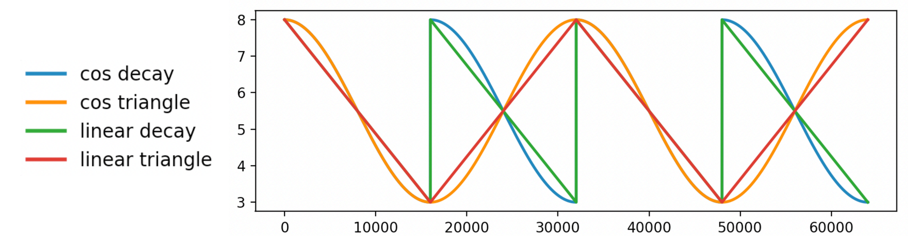


- After being exposed 'No Finding' classes, there is a **sharp spike** in recall
- By separating the 'No Finding' class, recall increases steadily

Learning Rate Schedulers



- Example of learning rates below



Future Work

Challenges and Limitations

- Open sourced data is hard to come by and inconsistent
- Since CNNs are blackbox methods and bugs are difficult to catch you must start simple and build up
- Only for one channel X-rays
- Online learning was worse performance than offline learning

New Directions

- Medical images as typically unbalance: how does this effects online learning scenarios
- Exploit the algorithms dynamic nature for multi-task learning
- Random index scheme with un-random replay eject
- Segmentation Tasks
- Incorporate more metadata for multi-nodal learning

Acknowledgements

Thank you to Dr. Anastasios Kyrillidisand Cameron Wolfe at Optimalab in Dept. Computer Science Rice University for their guidance.

Dept. Computer Science Rice University for sponsoring the program

The QR code includes relevant sources and links