cI[2] Platform for Rapid Development and Clinical Translation of ML Models

James R. Hawkins, Maryam Vareth, Marisa Lafontaine, Melanie Morrison, Janine Lupo, Jane Wang, Spencer Behr, Khadija, Wyatt Tellis, Emma Bahroos, Peter B. Storey, Jed Chan, Enrique Menendez, Pablo Damasceno, Eugene Ozhinsky, Victor Cheng, Valentina Pedoia, Justin Krogue, Ahmed Harouni*, Andriy Myronenko*, Mona Flores*, John Mongon, Marram P. Olson, Jason C. Crane, Chris Hess, Sharmila Majumdar

[1]*Department of Radiology and Biomedical Imaging, UCSF*

[2]*NVIDIA*

Highlights: The UCSF Center for Intelligent Imaging and NVIDIA have partnered to develop a framework to support rapid development of ML models and their clinical deployment. This platform will enable RBI clinicians to leverage the most advanced analytics in order to provide the best possible clinical care for our patients.

Introduction

In recent years, there has been wide-spread adoption of machine learning (ML) techniques in image processing, specifically with the use of convolutional neural networks and deep learning. As ML toolkits and algorithms are open-sourced, a key differentiator in model development is access to large quantities of quality, well-labelled data. Technology companies making rapid advancements in computer vision have built their models on vast amounts of consumer and personal data, but don't have access to private medical records. Researchers at UCSF have access to a trove of medical data, and the expertise to label and classify it, but have lacked the institutional infrastructure to take advantage of this. The Department of Radiology and Biomedical Imaging's Center for Intelligent Imaging (cI[2]) has partnered with NVIDIA to develop a framework to facilitate:

1. Streamlined ML model training using NVIDIA's Clara Train SDK [1]
2. Deployment of an ML inference service, connected to the clinical PACS, via Clara Deploy SDK

cI[2]'s Computational Core, in collaboration with Scientific Computing Services (SCS), the Majumdar/Pedoia/Link Lab, the Lupo Lab, the Quantitative Image Processing Center (QUIPC), clinical radiologists, the PACS group, and data scientists from NVIDIA, is building out and piloting this framework with 3 proof-of-concept projects. These POC's demonstrate the full ML model development lifecycle: from acquiring and preparing data, to model training, to running inference within a PACS connected application, to receiving feedback from radiologists that can be used to refine the models.

Methods

Clara Train SDK: The Clara Train SDK is a set of tools, built around the TensorFlow [2] machine learning platform, optimized by NVIDIA to maximize performance on their GPU's. It provides access to a set of pre-trained models and model architectures, and easy to use scripts and configuration files for training models from scratch, transfer-learning, and inference. Clara Train is accessed via Docker [3] containers, and in order to use it in our shared research and HPC environments, we have converted the Docker images to Singularity [4].

Clara Deploy SDK: The Clara Deploy SDK is a Docker container-based framework for deploying inference pipelines. It is a Kubernetes [5] based system, designed to receive DICOM images and execute inference pipelines. Imaging and scalar classification results can be push back into a PACS system, and the API's can be integrated into custom image viewing/analysis applications.

The 3 POC's: The first 2 projects (Brain Tumor Segmentation [6], Liver Segmentation for Surgical Planning of Transplant Cases) use Clara Train with NVIDIA's pre-configured model architectures [7], for model development using UCSF data. The third project focuses on development of a clinically integrated inference service for providing hip fracture detections [8] based on an MSK model developed outside of the Clara framework.

Images can be sent from scanners or PACS for inference with the resulting models. Model results will be pushed to an XNAT [9] research database and displayed for radiologist review and feedback via an application launched directly from our PACS workstations.

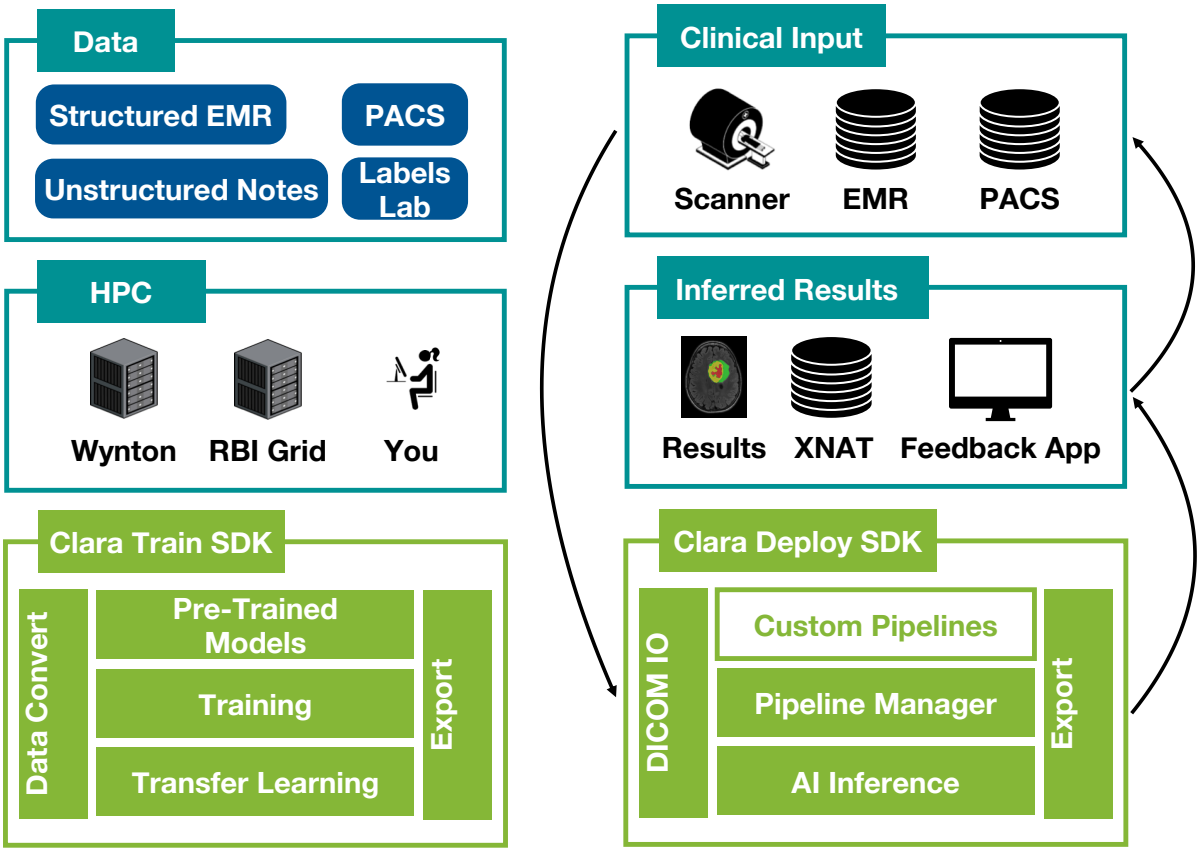

Figure 1: A framework for supporting ML training and clinical inference on UCSF infrastructure

References

1. NVIDIA CLARA Platform, https://developer.nvidia.com/clara-medical-imaging
2. TensorFlow, https://www.tensorflow.org
3. Docker, https://www.docker.com
4. Singularity, https://sylabs.io/docs/
5. Kubernetes, https://kubernetes.io
6. Myronenko, A., 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization, https://arxiv.org/abs/1810.11654
7. NIVIDIA Medical Model Library, https://ngc.nvidia.com/catalog/models
8. Krogue, J.D., Cheng, K.V., Hwang, K.M., Toogood, P., Meinberg, E.G., Geiger, E.J., Zaid, M., McGill, K.C., Patel, R., Sohn, J.H., Wright, A., Darger, B.F., Padrez, K.A., Ozhinsky, E., Majumdar, S., Pedoia, V (in submission). Automatic Hip Fracture Identification and Functional Subclassification with Deep Learning. Submitted to Radiology: Artificial Intelligence.
9. XNAT, https://www.xnat.org