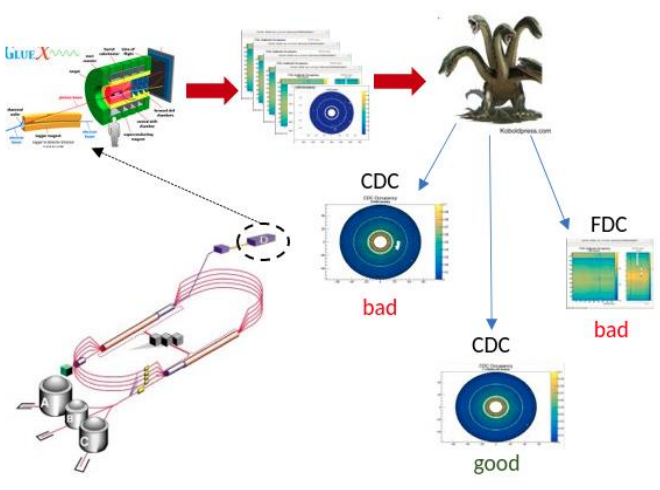
HYDRA

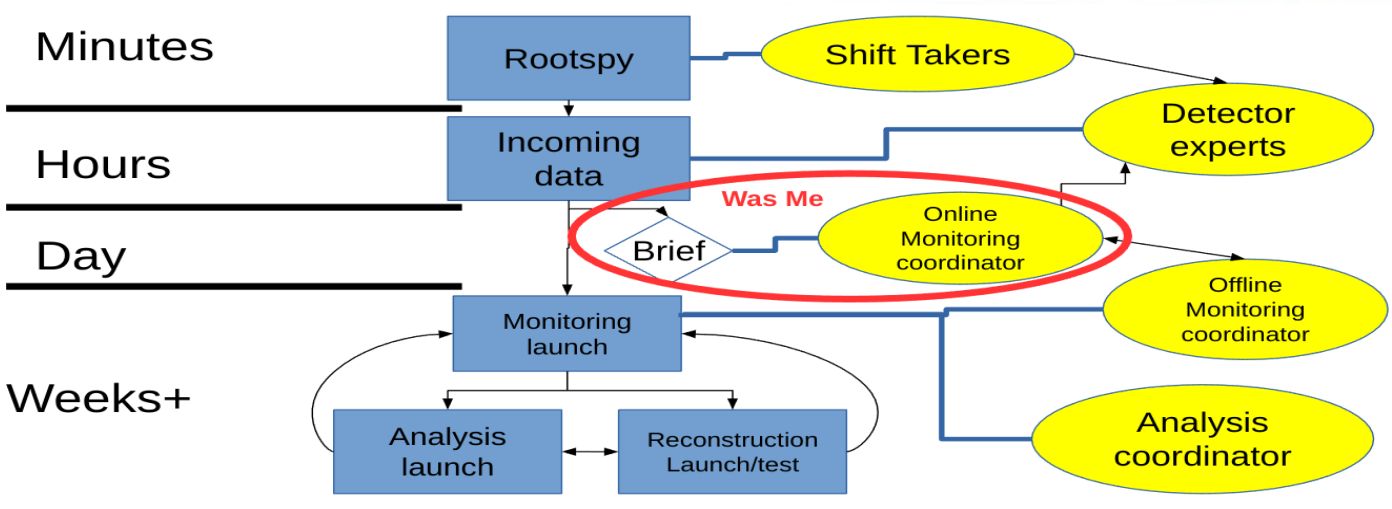
EPSCI Experimental Physics Software and Computing Infrastructure

Jefferson Lab — Thomas Jefferson National Accelerator Facility

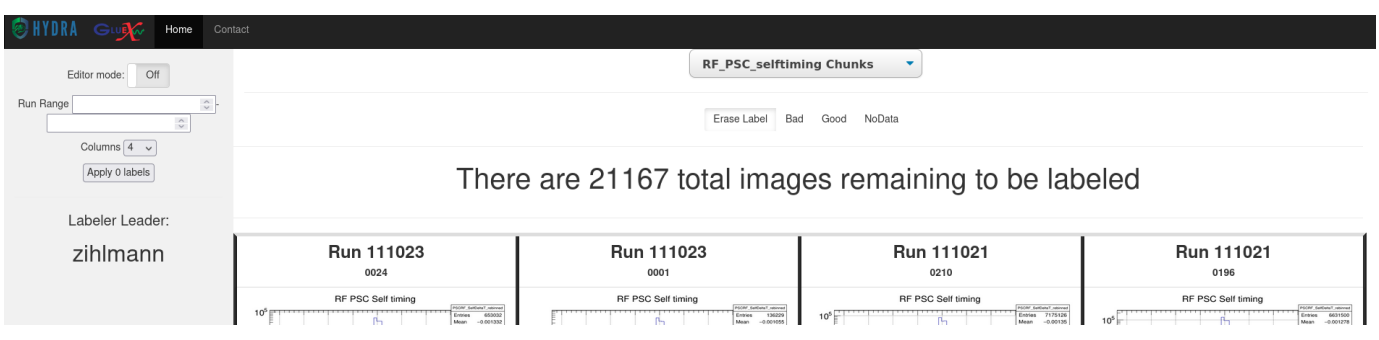
Thomas Britton, David Lawrence, Torri Jeske, Kishansingh Rajput

Hydra is an extensible framework for training and managing AI/ML for near real time monitoring that aims to relieve the shift crew and online monitoring coordinator of the tedious parts of data quality monitoring. It watches incoming data for signs of problems and flags them. Backed up by a comprehensive database it utilizes a web based front-end for efficient labeling as well as viewing the status from anywhere in the world.





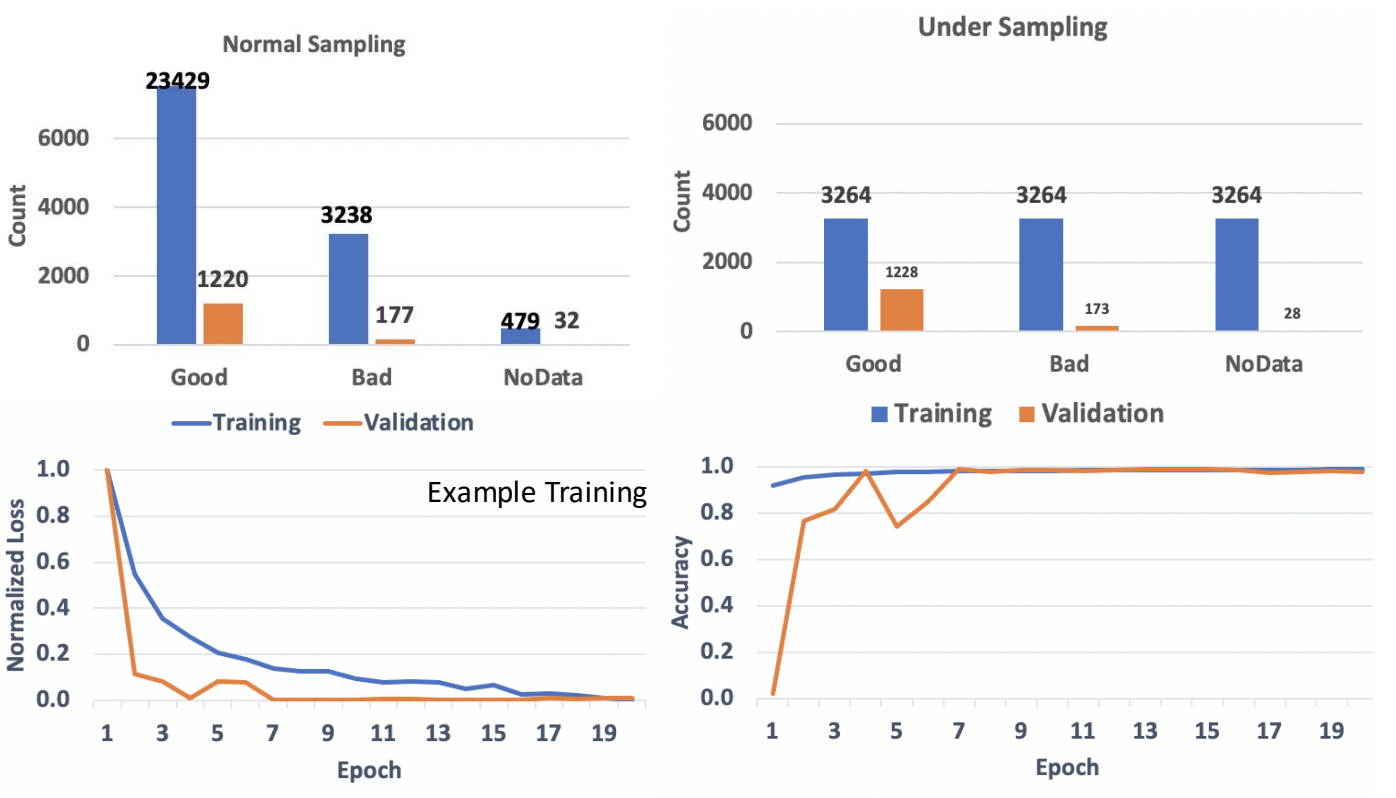
Web Based Labeling and Push Button Training:



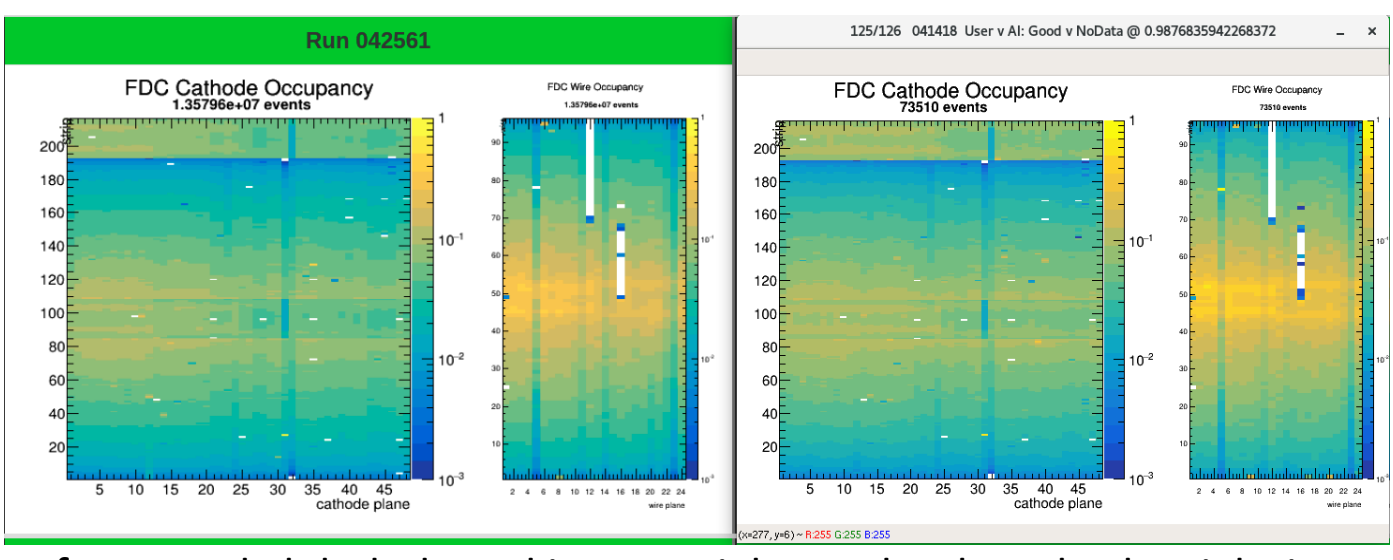
Web app for labeling allows labelers to label hundreds of images very quickly (~ 1 minute). Labels are stored in a database for use in training and model validation.

Optimized Training Time:

Training time is optimized by balancing number of images with each label. Thus, we use all instances of bad behavior and a sub-sample of good behavior. Training is virtually push button; most models based on Google's Inception v3 [1].

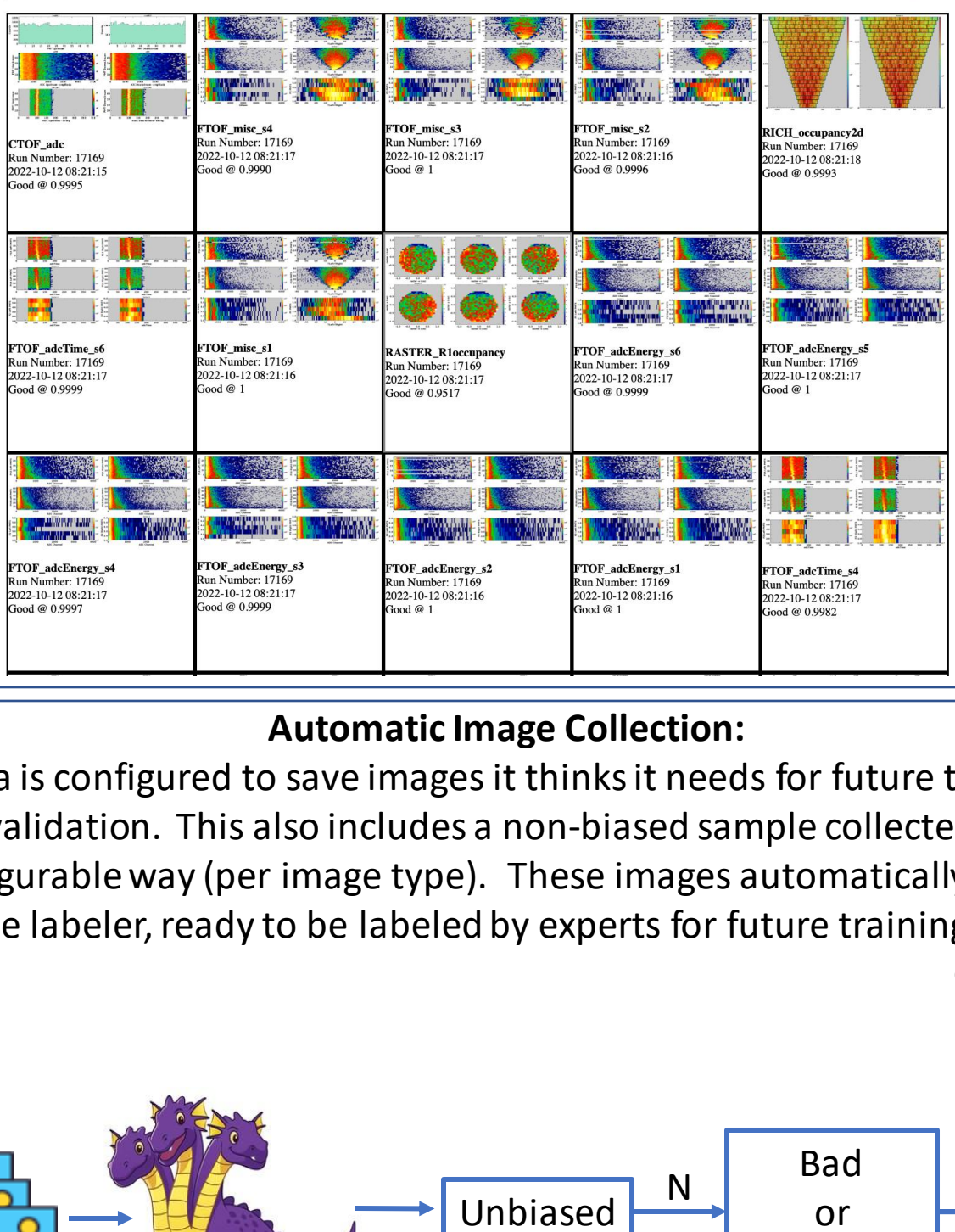


Each Model is Validated Before Being Put Into Production.



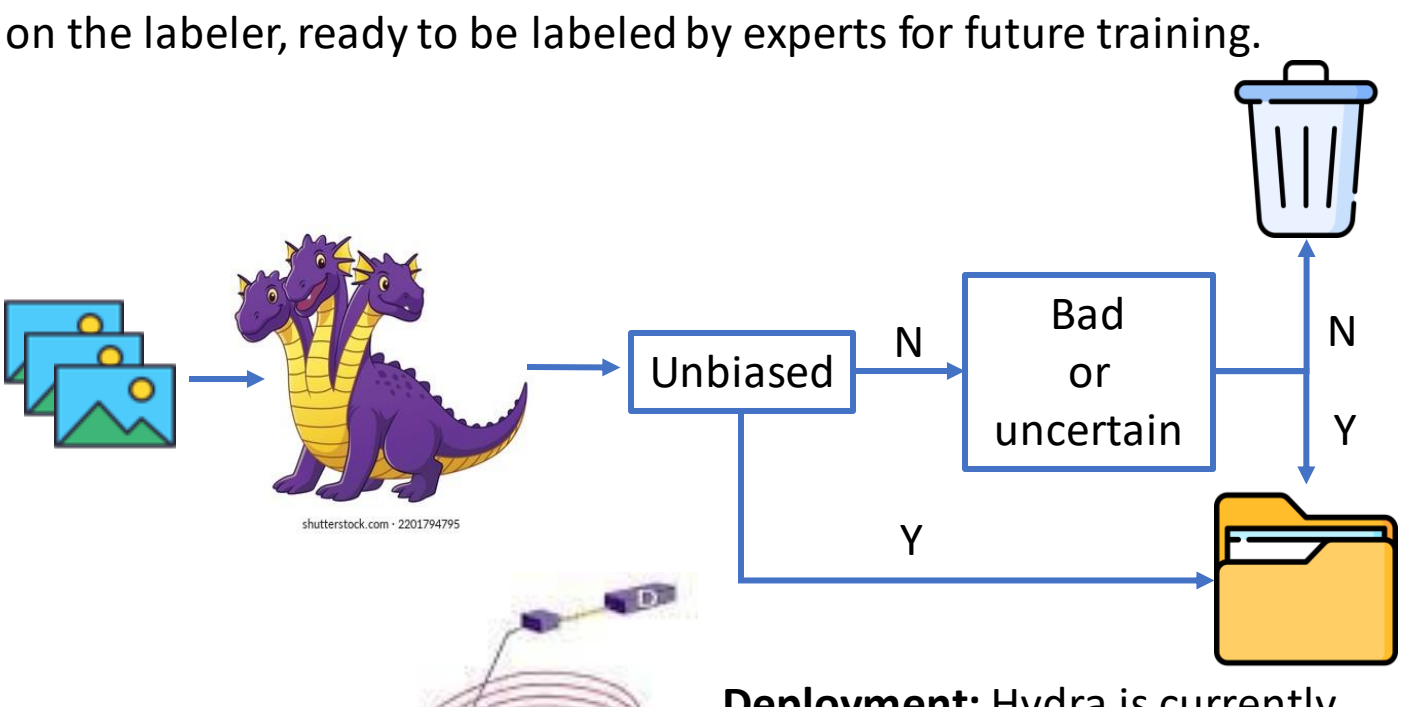
Left: Properly labeled good image. Right: Hydra thought the right image did not contain enough statistics and was confident in the NoData label. Super-human performance! Expert stated Events < 100,000 => no data; Hydra was right!

Near Real-time Hydra Analysis Available via Webpage to Users



Automatic Image Collection:

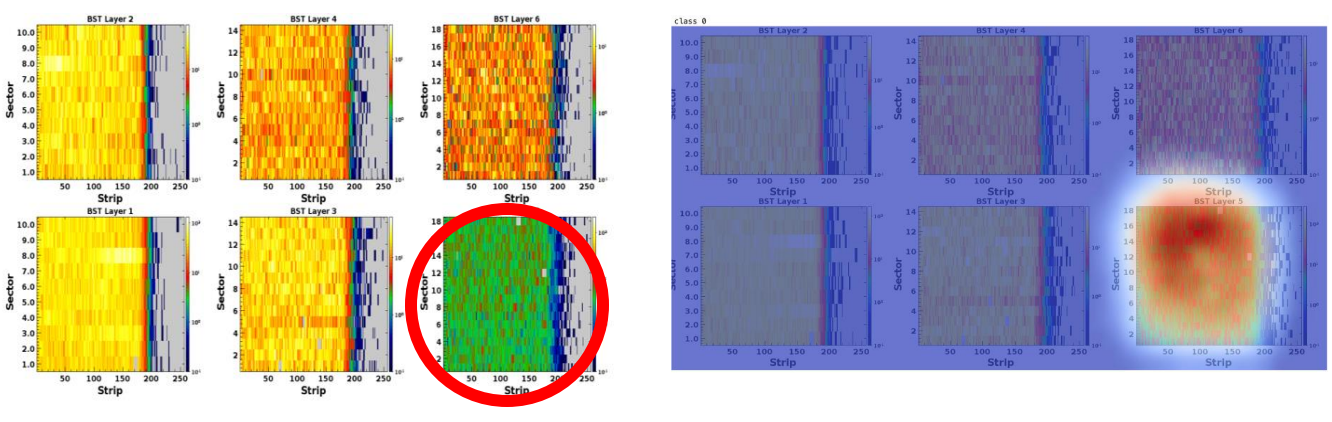
Hydra is configured to save images it thinks it needs for future training and validation. This also includes a non-biased sample collected in a configurable way (per image type). These images automatically appear on the labeler, ready to be labeled by experts for future training.



Deployment: Hydra is currently deployed in 3 of the 4 experimental halls at Jefferson Lab, with plans to deploy to the 4th very soon. The code is generic enough such that deployment to other Halls is straightforward and requires little in the way of infrastructure.

Upcoming Feature: Hydra + GradCAM

Can we identify regions in an image important for determining whether it is "Good" or "Bad"?



This example from the Barrell Silicon Vertex Tracker located in Hall B. The detector expert labeled (and Hydra classified the left image) as bad due to low occupancy in a single sector (red circle). Right: Heat map indicates the same region as important to Hydra for labeling the image as "Bad". GradCAM [2] should provide enhanced diagnostics to shift crews.

Contact: tbritton@jlab.org

Thanks to Diana McSpadden for being our proxy