DATA SCIENCE INSTITUTE®
AMERICAN COLLEGE OF RADIOLOGY

ACR AI-Lab – MGH Pilot

Romane Gauriau
Sr Machine Learning Scientist, MGH & BWH Center for Clinical Data Science (CCDS)

Goal of the pilot

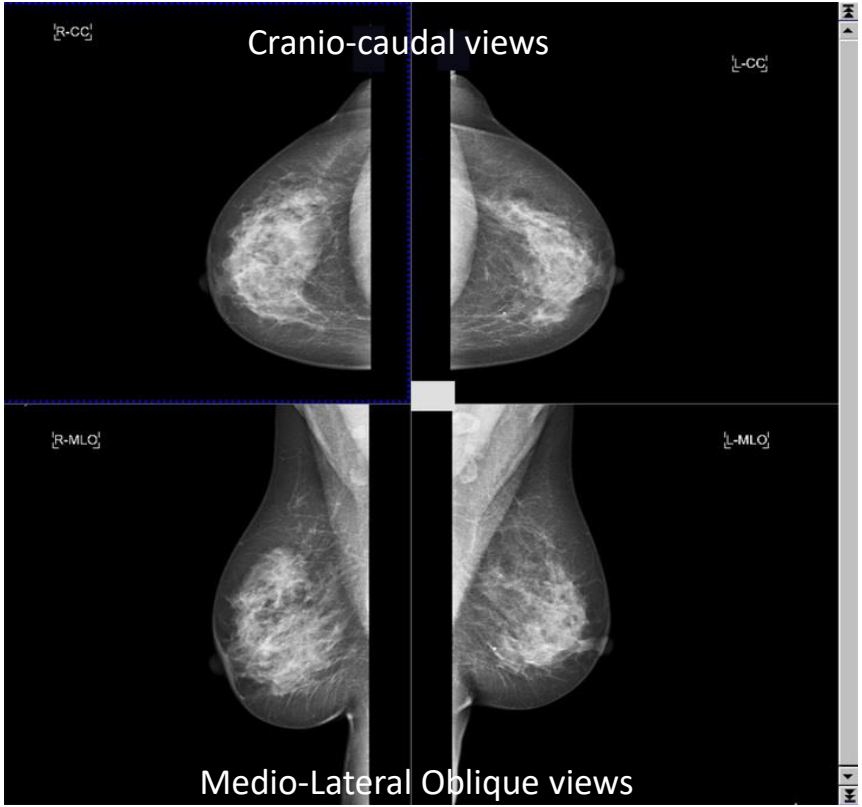
- The Primary Aim of this study is to determine whether the software can be successfully deployed at multiple institutions to allow them to download an AI model developed at another institution, add new data from their own institution, update the model, and share the results.
- Secondary Aim of this study is to compare the difference in performance of a Breast Density model before and after the model has been updated with images from the pilot sites.

Program plan

Task	Owner	Institution
IRB submission	Ali Pourvaziri	CCDS
Online Lab notebook	Ali Pourvaziri	CCDS
Provision Hardware	Adam McCarthy	CCDS
Enable user access	Adam McCarthy	CCDS
Install SW	Stephanie Bossong, Nir Neumark	ACR, CCDS
Identify data set, 2D Mammography Studies	Nir Neumark, Jayashree Kalpathy-Cramer	CCDS, Martinos
Request data pull	Min Yun	CCDS
Pull Data	Sean Doyle	CCDS
Select series and annotate	Nir Neumark, Ken Chang, Jayashree Kalpathy-Cramer	CCDS Martinos
Run model	Romane Gauriau	CCDS
Collate results	Romane Gauriau	CCDS

Use case – Breast Density in 2D mammography

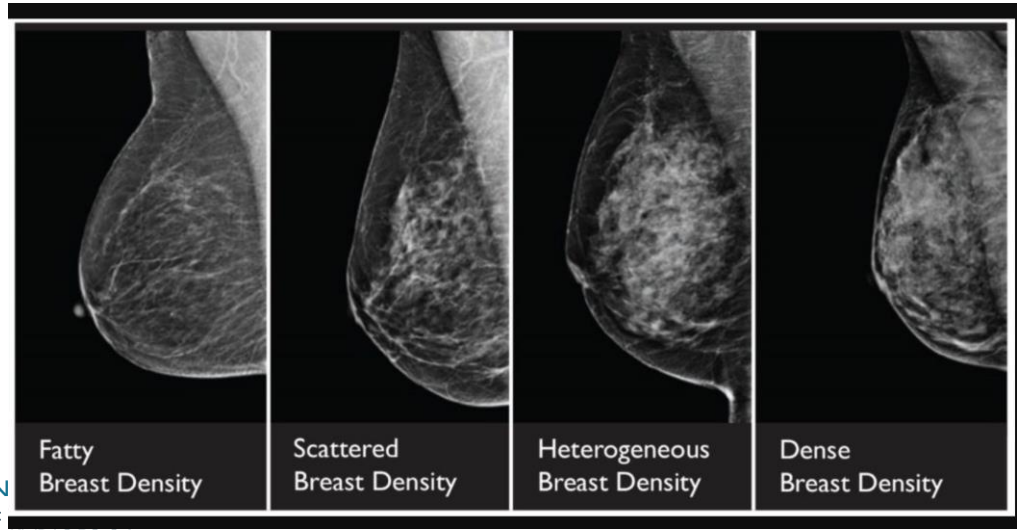
- Mammogram:
  low-dose X-ray of the breasts
- Used as a diagnostic
  and screening tool
- Early tumors can be detected
- Different views with different
  angles are performed



DATA SCIENCE INSTITUTE®
AMERICAN COLLEGE OF RADIOLOGY

Use case – Breast Density in 2D mammography

- Each examination was interpreted by a radiologist
- The mammograms were annotated using ACR BI-RADS breast density lexicon (fatty, scattered, heterogeneously dense, dense)



DATA SCIENCE IN
AMERICAN COLLEGE OF

Pilot script

Off the platform

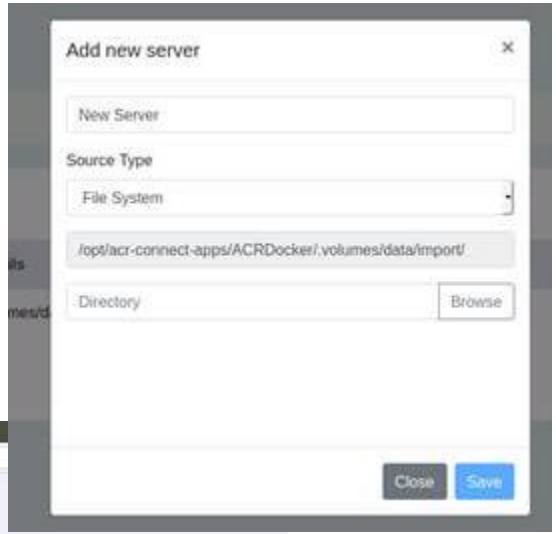
- Data pull
- Cohort creation
- Create excel file with list of studies and annotations
- Copy data on the local machine where the platform is installed

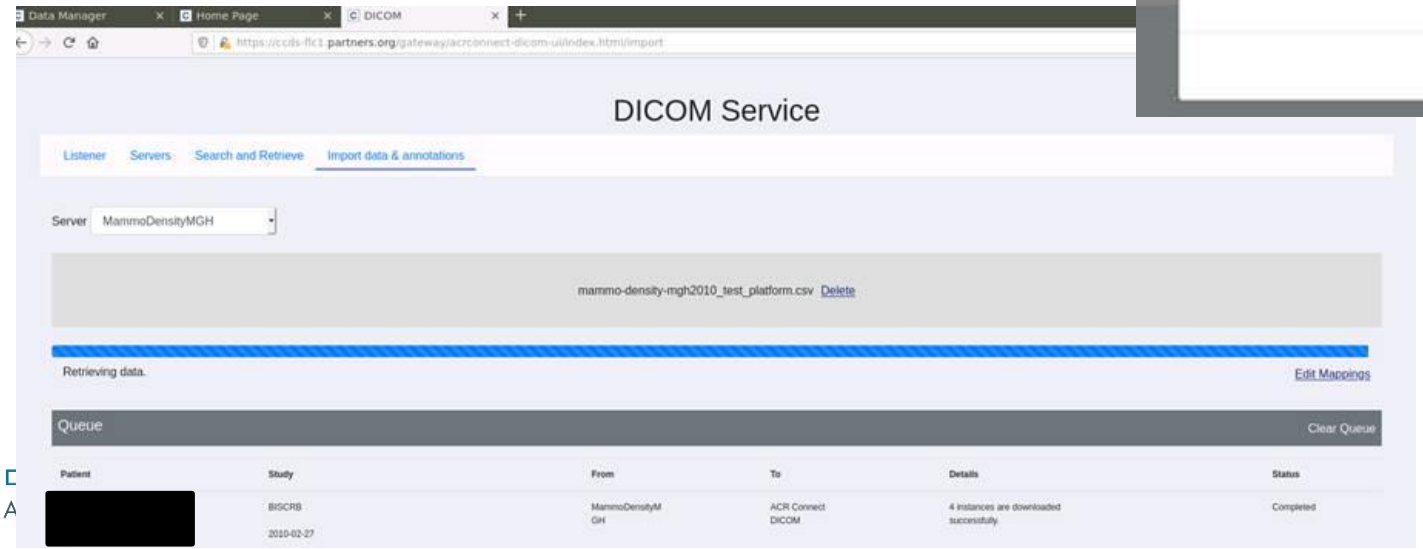
On the platform (Graphical Use Interface)

- **Import DICOM studies in the platform**
- **Prepare dataset**
- Evaluate ACR model
- Train new model on MGH data
- Evaluate new model on MGH data

DICOM import

- Copy the data on the local machine
- Configure the location on the platform
- Load excel file with the list of studies (+annotations) to create the cohort



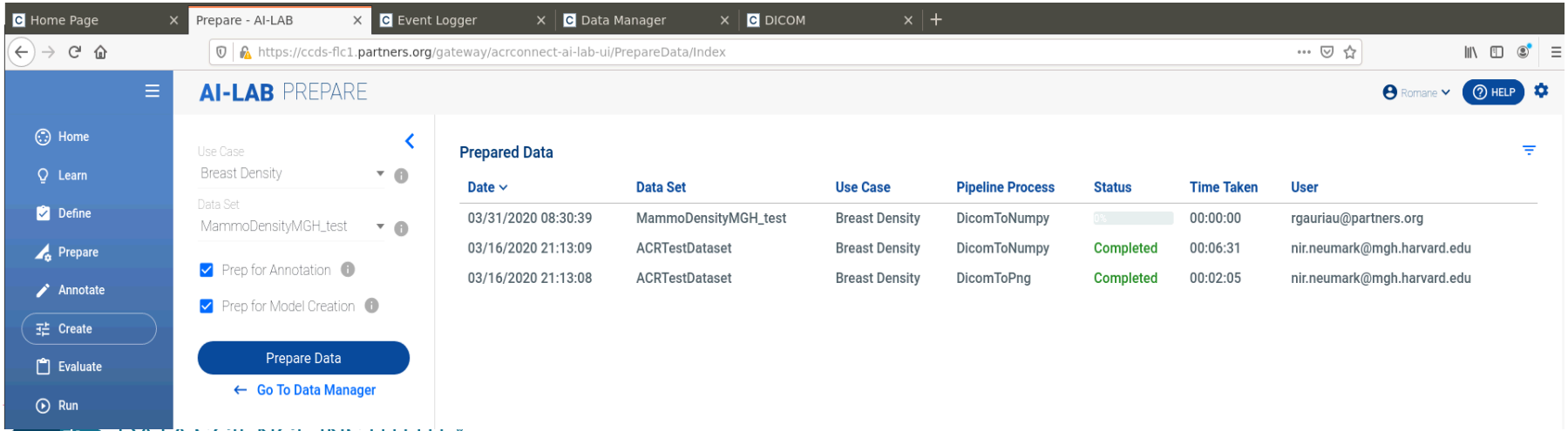


DICOM import

- Easy to setup
- Relatively fast (240 studies loaded in X min)
- Limitations found:
  - Requires enough disk space on the local machine
  - Multithreading was causing issues when loading many studies (>1000 studies)

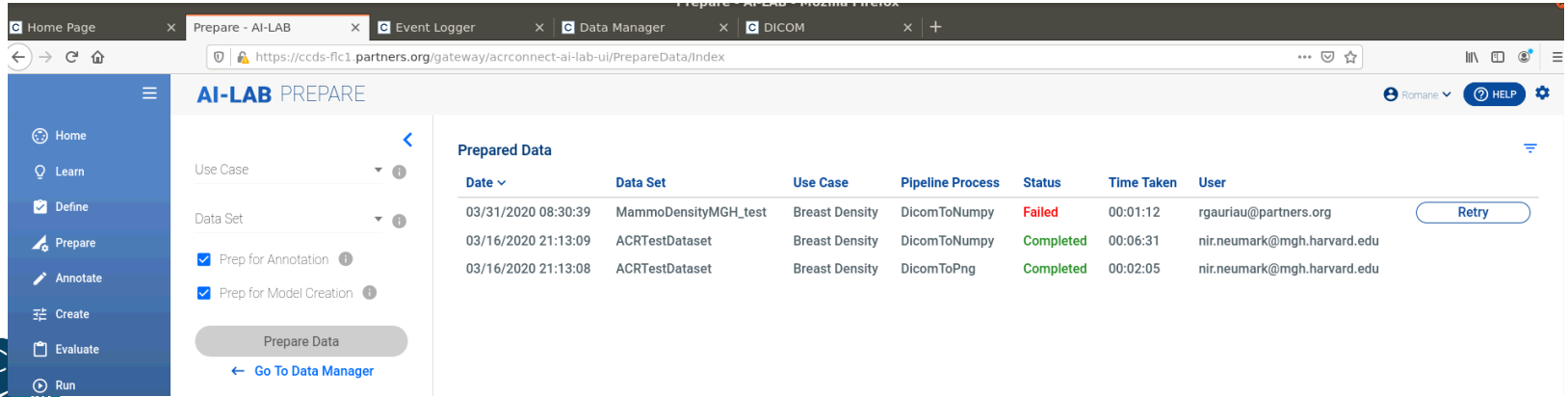
Data preparation

- Conversion to PNG and to numpy
- For the annotations and for the models



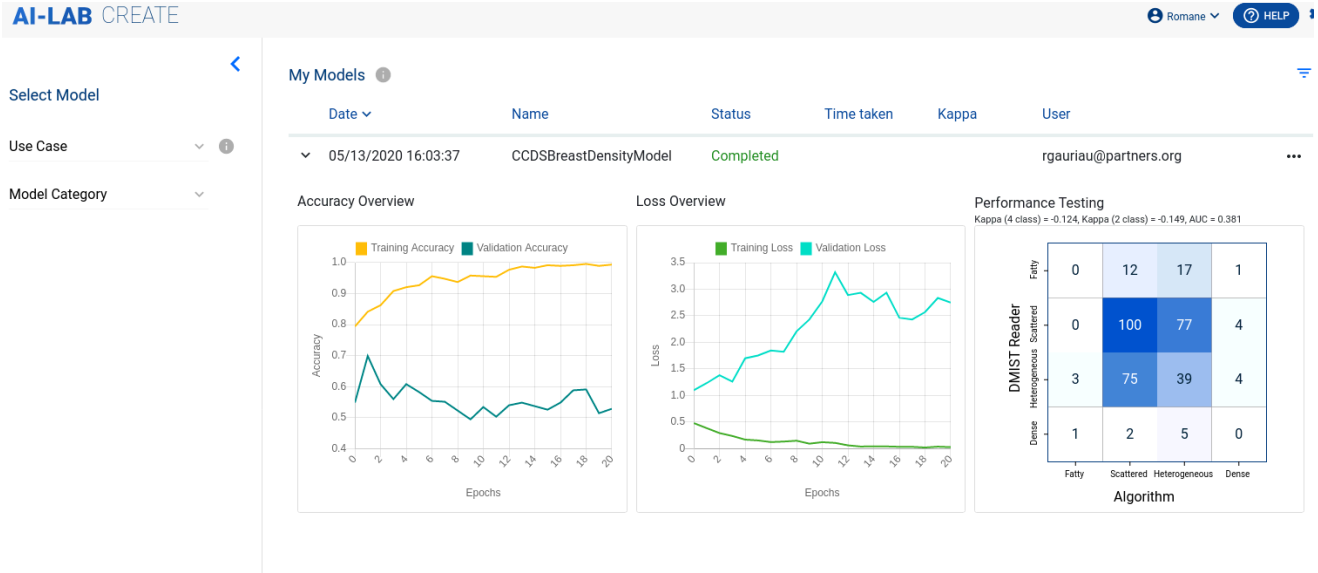
Data preparation

- User-friendly
- Easy to do
- Ran into problems because of some DICOM files that were not images (DICOM SR files had not been filtered from the raw DICOM studies)



Model training and evaluation

- Straightforward
- Fast
- User-friendly



DATA SCIENCE INSTITUTE®
AMERICAN COLLEGE OF RADIOLOGY

Conclusion

- User-friendly
- Well adapted for
  - well defined use cases
  - retrain existing models on new data
  - machine learning beginners
- Some limitations:
  - Requires precise data cleaning/curation of the raw DICOM studies
  - … but problem easy to fix

- Very promising tool for the democratization of AI in clinical environment

Acknowledgements

- **CCDS**: Nir Neumark, Shaimaa Sharaf, Ali Pourvaziri, Adam McCarthy, Min Yun, Sean Doyle, Ram Naidu
- **MGH**: Tarik Alkasab
- **Martinos Center**: Jayashree Kalpathy-Cramer, Ken Chang
- **ACR**: Stephanie Bossong, Laura Coombs, Kris Glassmir, Deepak Kattil Veettil

DATA SCIENCE INSTITUTE®
AMERICAN COLLEGE OF RADIOLOGY