Machine learning for Nephrocage

2nd NephroCAGE Symposium
Montréal, QC, Canada
16 August 2022

Dr. Mozhgan Bayat
Research Associate at NephroCAGE
Hasso Plattner Institute for Digital Engineering gGmbH

NEPHROCAGE
German-Canadian consortium on AI for improved kidney transplantation outcome
2nd International NephroCAGE Symposium, Aug 16, 2022

HPI Hasso Plattner Institut
Digital Engineering • Universität Potsdam
KIT Karlsruher Institut für Technologie
UBC The University of British Columbia
Université de Montréal
Polytechnique Montréal Université d'ingénierie
GenomeQuébec
PIRCHE
CHARITÉ Universitätsmedizin Berlin
Centre universitaire de santé McGill
McGill University Health Centre
CHUM
GenomeCanada
Genome BritishColumbia
Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

Agenda

- Motivation behind AI for Kidney transplantation
- State of Art in this field
- Clinical prediction models
- Future work and Conclusion

Why AI in kidney transplantation?

- Despite excellent short term graft survival rates, long-term graft survival has remained a challenge.
- The excess risk of developing cancer after kidney transplantation is approx. 5x higher.
- Kidney transplantation clinics are facing rich data which enable them to be more efficient, operate with greater insight and improves recipients' life

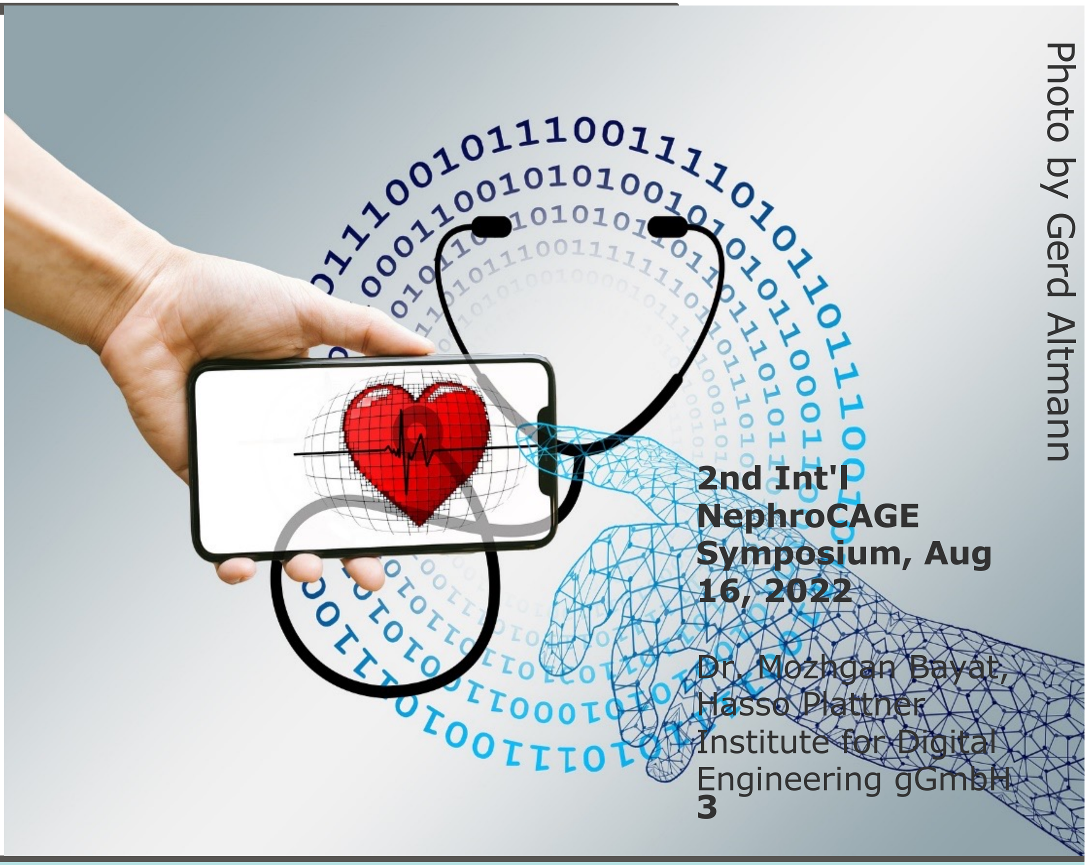
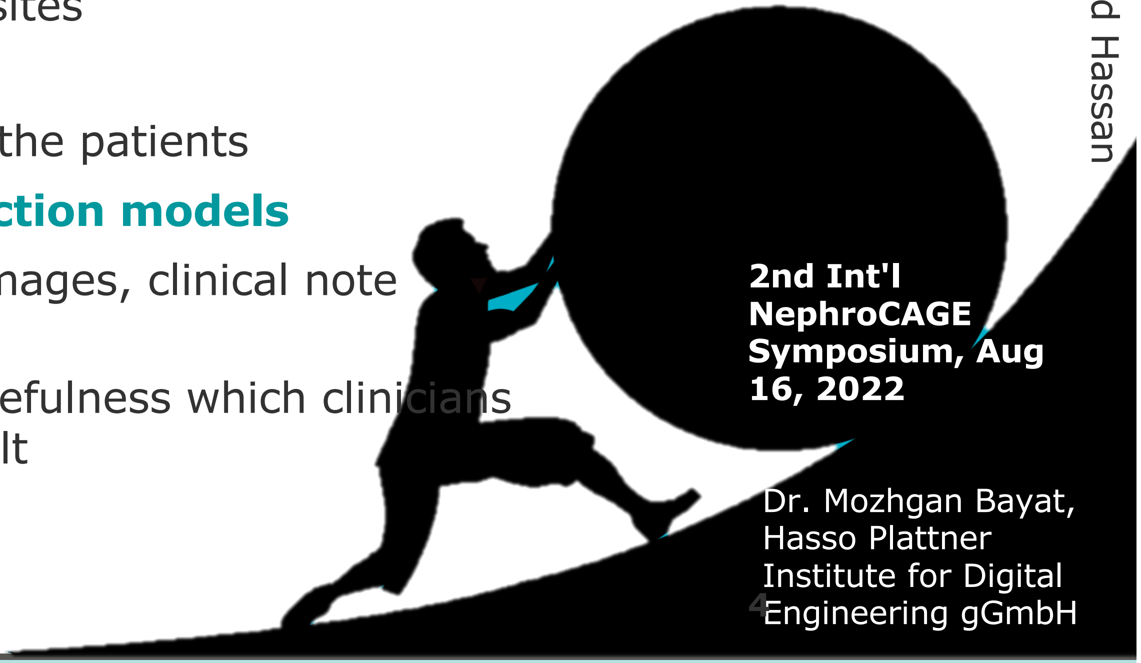


Photo by Gerd Altmann

Research in applying state of art in AI to predict outcomes of kidney transplant

- **Challenges in providing dataset for training**
  - Unavailable of large set of labeled data kidney transplant data
  - Harmonizing the dataset from multiple sites
  - Large dimensional of dataset
  - Keep data secure and protect privacy of the patients
- **Challenges in providing clinical prediction models**
  - Multi model data such as tabular data, images, clinical note , HLA data, etc
  - Providing interpretable CPM with high usefulness which clinicians can take actions upon on predicted result



2nd Int'l NephroCAGE Symposium, Aug 16, 2022

Dr. Mozhgan Bayat, Hasso Plattner Institute for Digital Engineering gGmbH

Photo by mohamed Hassan

State of art

- Statistical methods such as Cox regression [1]
- Bayesian Modeling [2]
- Machine learning: support vector machine(SVM),Tree-based method such as Random Forest(RF) and Xgboost [3]
- Deep learning are mainly used in classification of pathology result [4]

---

[1] Loupy, Alexandre, et al. "Prediction system for risk of allograft loss in patients receiving kidney transplants: international derivation and validation study." *bmj* 366 (2019)

[2] Raynaud, Marc, et al. "Dynamic prediction of renal survival among deeply phenotyped kidney transplant recipients using artificial intelligence: an observational, international, multicohort study." *The Lancet Digital Health* 3.12 (2021): e795-e805

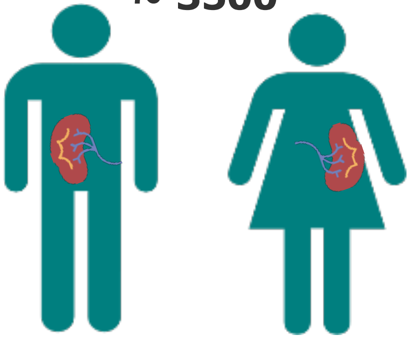
[3] Shaikhina, Torgyn, et al. "Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation." *Biomedical Signal Processing and Control* 52 (2019): 456-462

[4] Hermsen, Meyke, et al. "Deep learning–based histopathologic assessment of kidney tissue." *Journal of the American Society of Nephrology* 30.10 (2019): 1968-1979.

Patients’ characteristics

Number of Recipients

~ 3500



~ 2000 **~ 1500**

	Graft lost	Death
0-1 yr	83	116
1-3 yrs	87	146
3-5 yrs	103	166
5-8 yrs	132	255

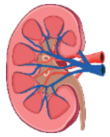
% of survival after 8 yrs	
88%	81%

2nd Int'l NephroCAGE Symposium, Aug 16, 2022

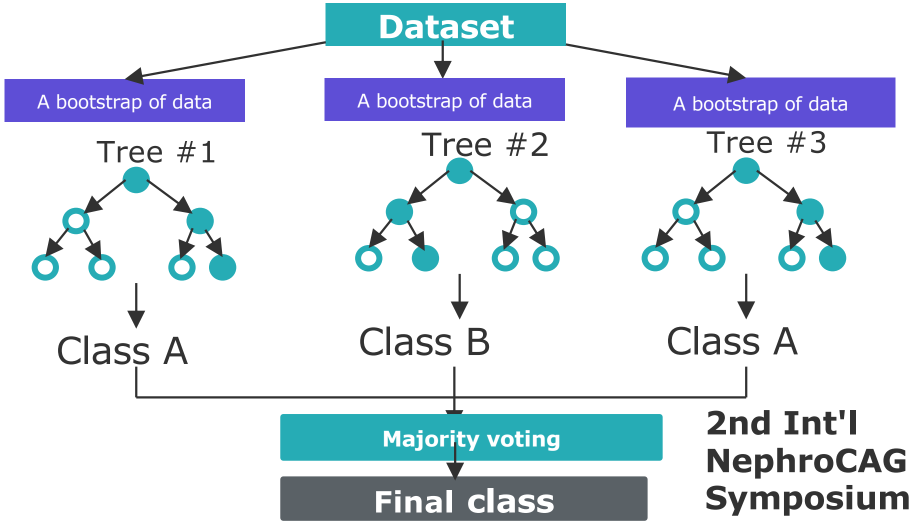
Dr. Mozhgan Bayat, Hasso Plattner Institute for Digital Engineering gGmbH

Clinical prediction models

Long term recipient survival



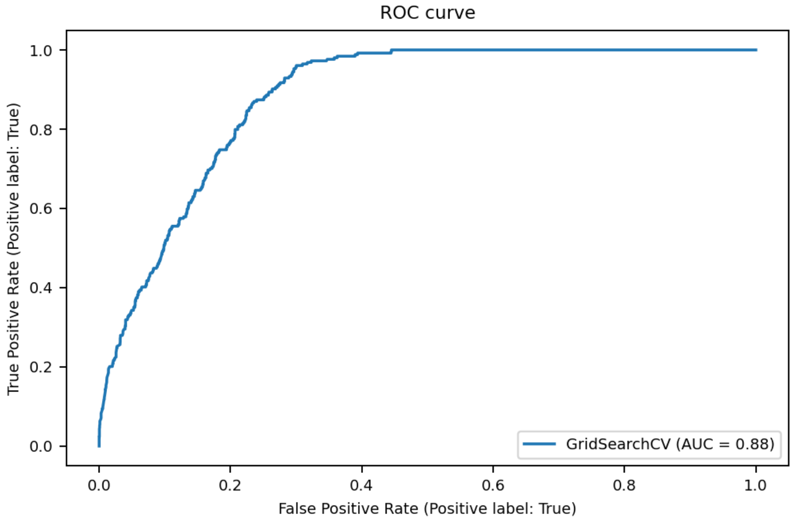
Long term kidney allograft failure

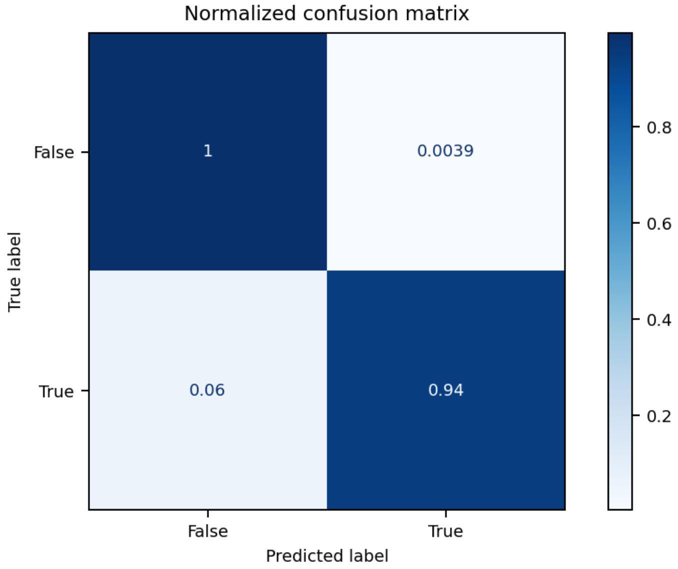


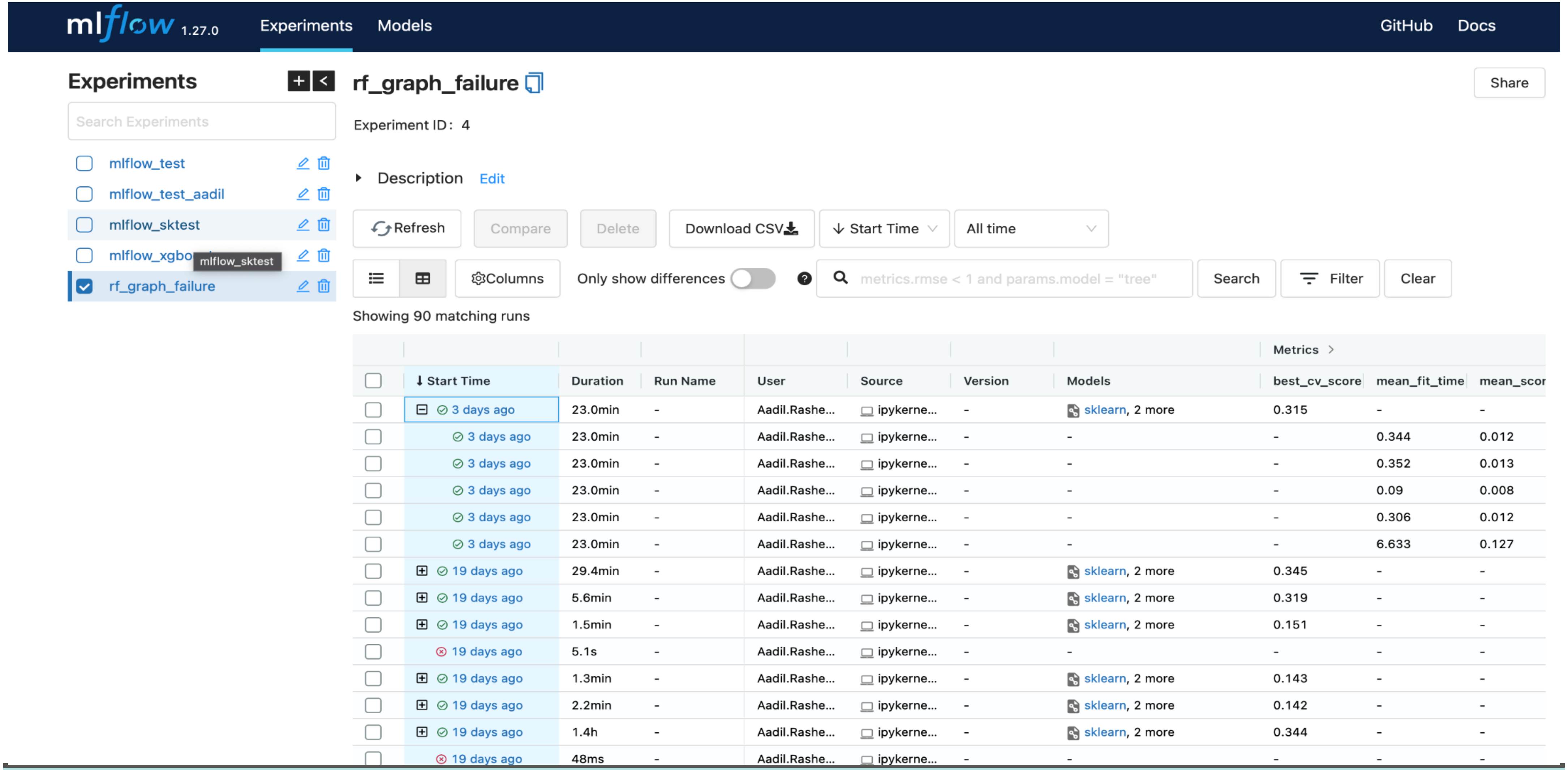
First year data for training

Next four years data for creating prediction output labels

Long term graft failure prediction performance


ROC curve


Normalized confusion matrix



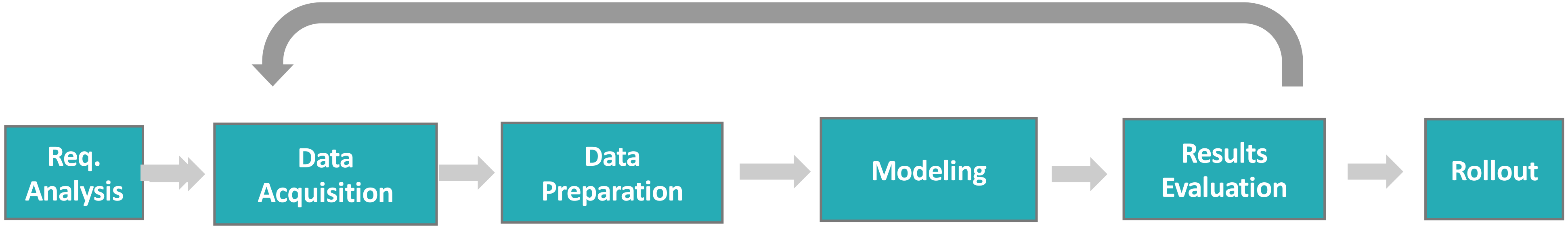
rf_graph_failure

Experiment ID: 4

Showing 90 matching runs

								Metrics		
☐	Start Time	Duration	Run Name	User	Source	Version	Models	best_cv_score	mean_fit_time	mean_scor
☐	⊟ ✓ 3 days ago	23.0min	-	Aadil.Rashe...	ipykerne...	-	sklearn, 2 more	0.315	-	-
☐	✓ 3 days ago	23.0min	-	Aadil.Rashe...	ipykerne...	-	-	-	0.344	0.012
☐	✓ 3 days ago	23.0min	-	Aadil.Rashe...	ipykerne...	-	-	-	0.352	0.013
☐	✓ 3 days ago	23.0min	-	Aadil.Rashe...	ipykerne...	-	-	-	0.09	0.008
☐	✓ 3 days ago	23.0min	-	Aadil.Rashe...	ipykerne...	-	-	-	0.306	0.012
☐	✓ 3 days ago	23.0min	-	Aadil.Rashe...	ipykerne...	-	-	-	6.633	0.127
☐	⊞ ✓ 19 days ago	29.4min	-	Aadil.Rashe...	ipykerne...	-	sklearn, 2 more	0.345	-	-
☐	⊞ ✓ 19 days ago	5.6min	-	Aadil.Rashe...	ipykerne...	-	sklearn, 2 more	0.319	-	-
☐	⊞ ✓ 19 days ago	1.5min	-	Aadil.Rashe...	ipykerne...	-	sklearn, 2 more	0.151	-	-
☐	⊗ 19 days ago	5.1s	-	Aadil.Rashe...	ipykerne...	-	-	-	-	-
☐	⊞ ✓ 19 days ago	1.3min	-	Aadil.Rashe...	ipykerne...	-	sklearn, 2 more	0.143	-	-
☐	⊞ ✓ 19 days ago	2.2min	-	Aadil.Rashe...	ipykerne...	-	sklearn, 2 more	0.142	-	-
☐	⊞ ✓ 19 days ago	1.4h	-	Aadil.Rashe...	ipykerne...	-	sklearn, 2 more	0.344	-	-
☐	⊗ 19 days ago	48ms	-	Aadil.Rashe...	ipykerne...	-	-	-	-	-

Machine learning cycle



Req. Analysis → **Data Acquisition** → **Data Preparation** → **Modeling** → **Results Evaluation** → **Rollout**

Data Acquisition
- Data dictionary
- Anonymization of patient

Data Preparation
- Data cleaning
- Cohort selection
- Handling missing value

Modeling
- Feature engineering
- Continuous and categorical data handeling
- Cross validation

Results Evaluation
- Hyper parameter optimization
- Feature selection
- Model selection

Future work and conclusion

- The data was used for trained was small size, and it was highly imbalance in classes. This aspect should be addressed more advance.
- Data cleaning and feature engineering were the most challenging part.
- We are deploying the CPMs to provide a service for inferencing and fine tuning the model
- Our next step is to train our CPM in federated learning infrastructure on more data from different center.

thank you