

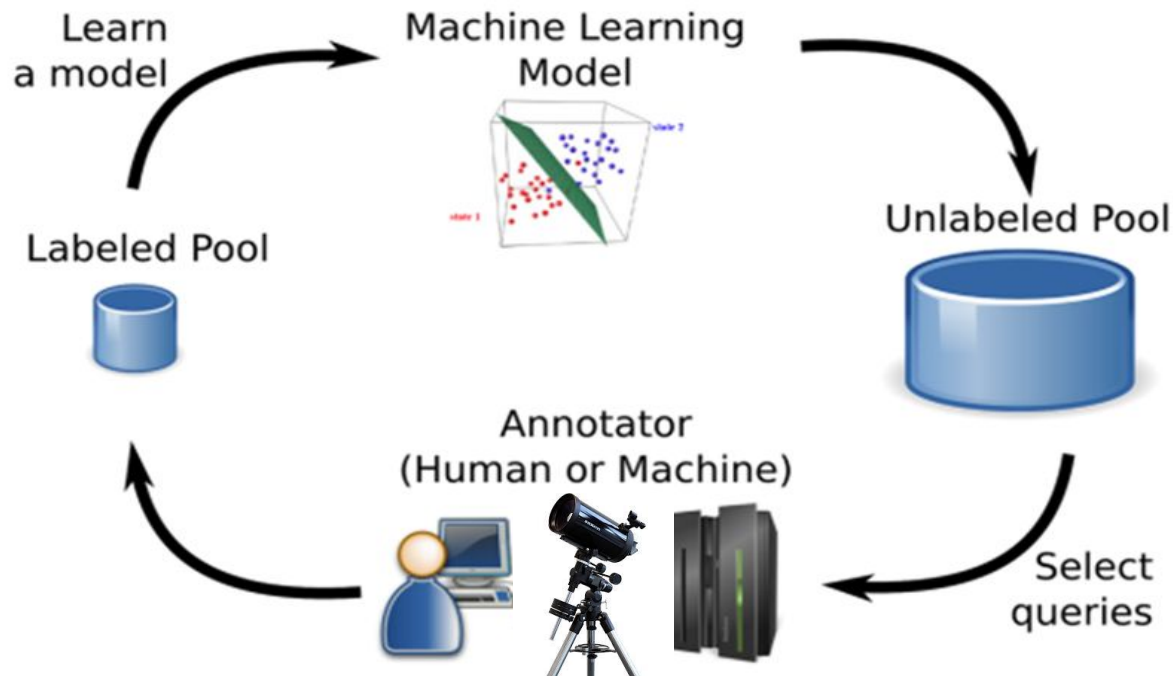
Active Learning in

Maximum classification with minimum training

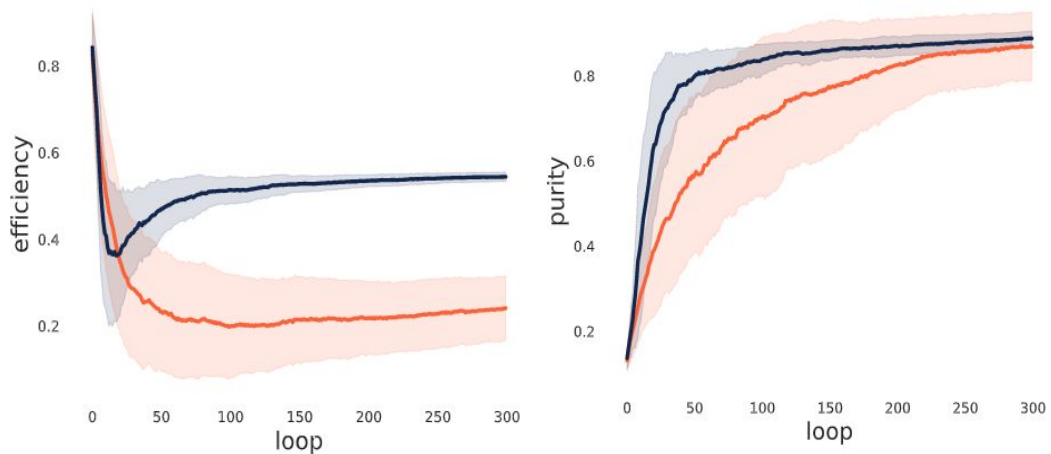
*Enabling Astronomical Transient discoveries in the Rubin era: the Fink-Brazil Workshop
8 May 2024, CBPF - Rio de Janeiro, Brazil*

Active Learning

Optimal classification, minimum training



Case study: Early SN Ia classification



Results after 300 loops:

*Training: 310 alerts
Testing: > 52 000 alerts*

Choose training sample which lead to better results
and train a Random Forest classifier ...

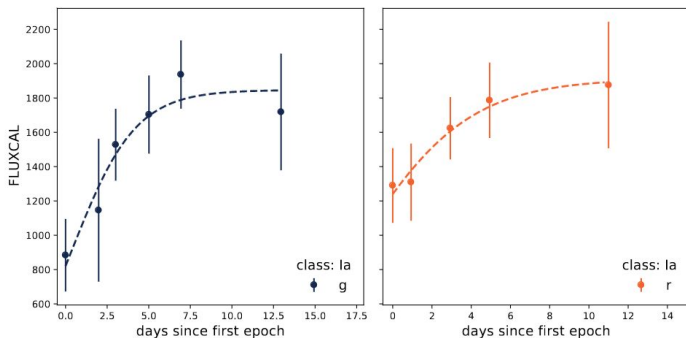


Trained ML
model

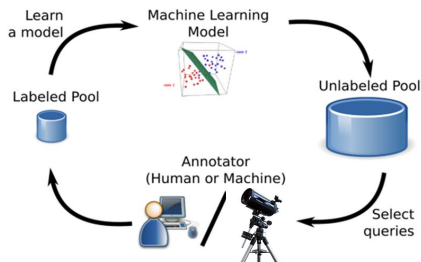


Early SN Ia Fink reported to TNS

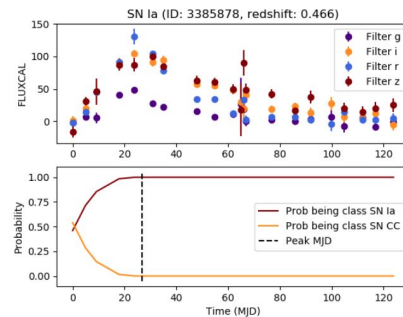
- Sigmoid features from *Leoni et al., 2022*



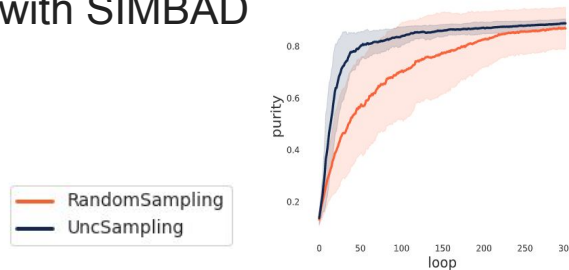
- Off-line train with Active Learning



- Agreement with SuperNNova



- Less than 20 detected points
- At least 3 points per filter
- Xmatch with SIMBAD



Early SN Ia Fink reported to TNS

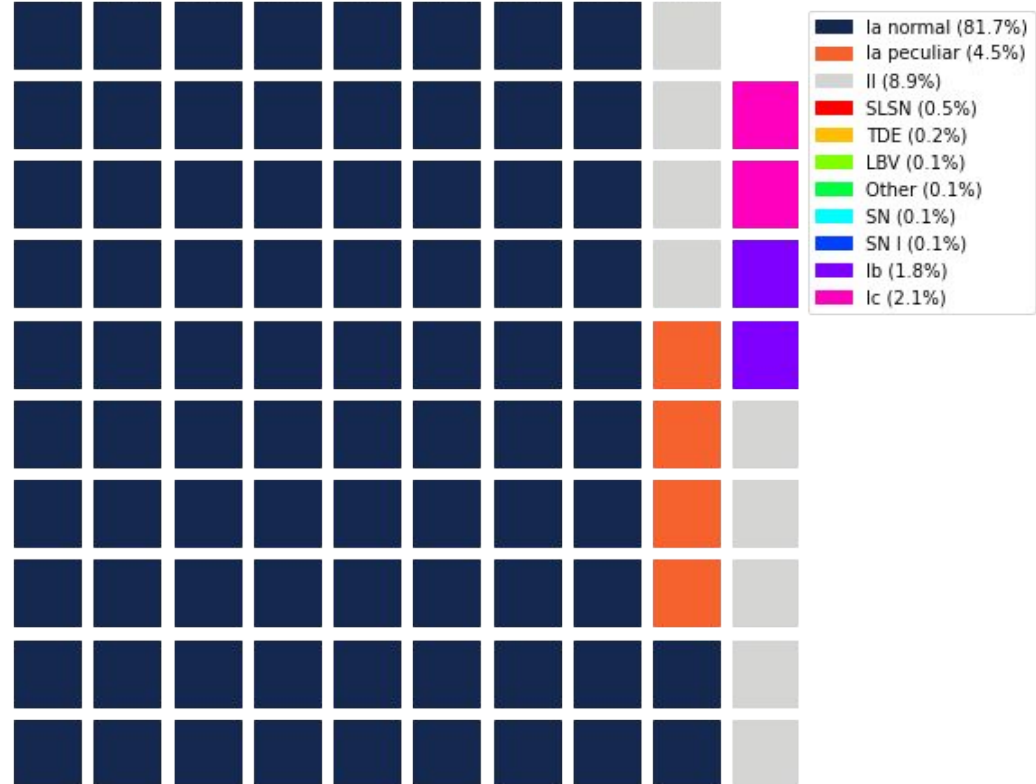
Not a known transient/variable +
Score (early SNIa & SNN SN | SNIa)

= 3017 reported to TNS
= 1632 followed-up

Marco, Emille, Julien, +++

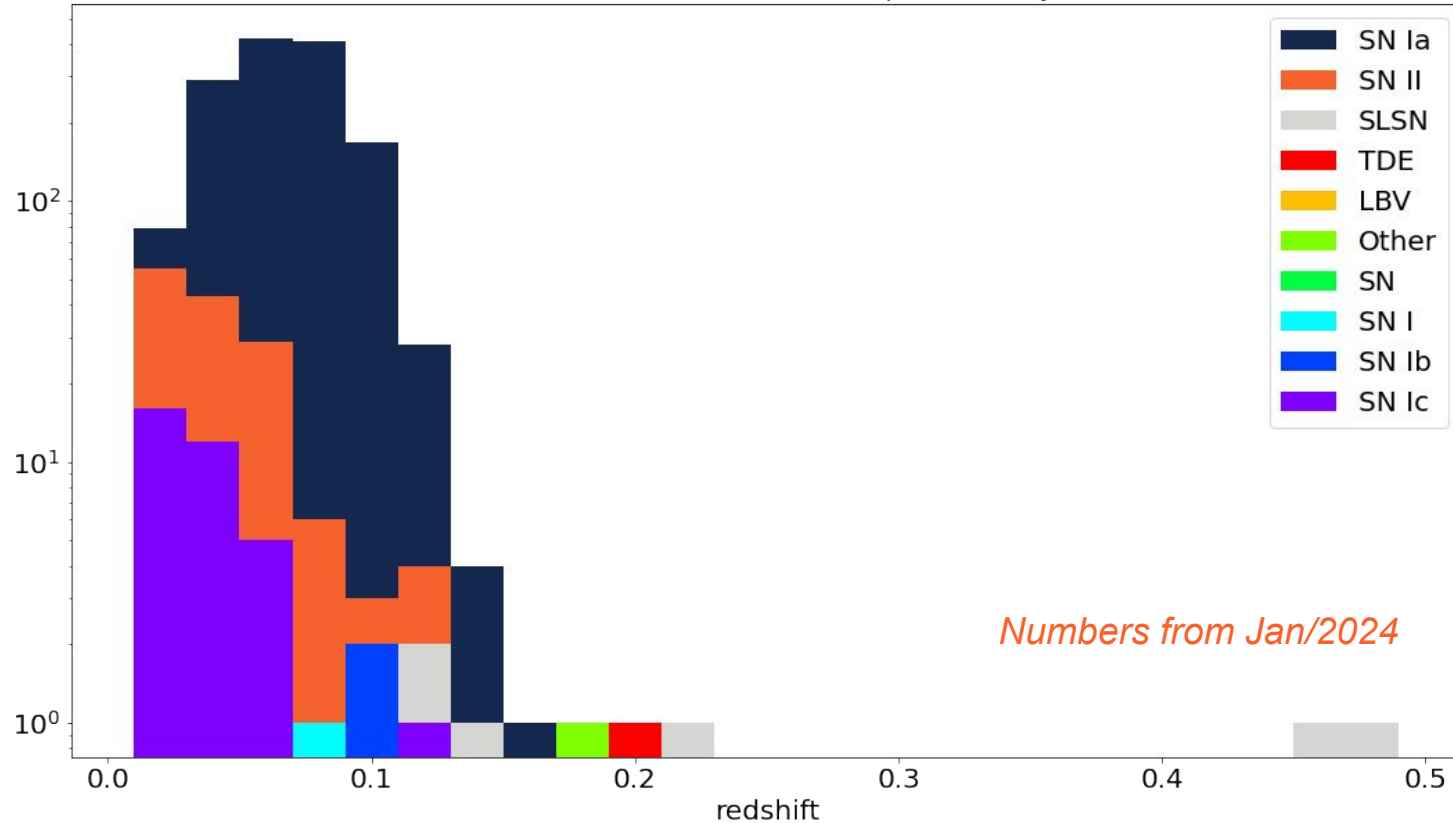
Numbers from Jan/2024

Fink early SNIa TNS reported and followed-up spectroscopically



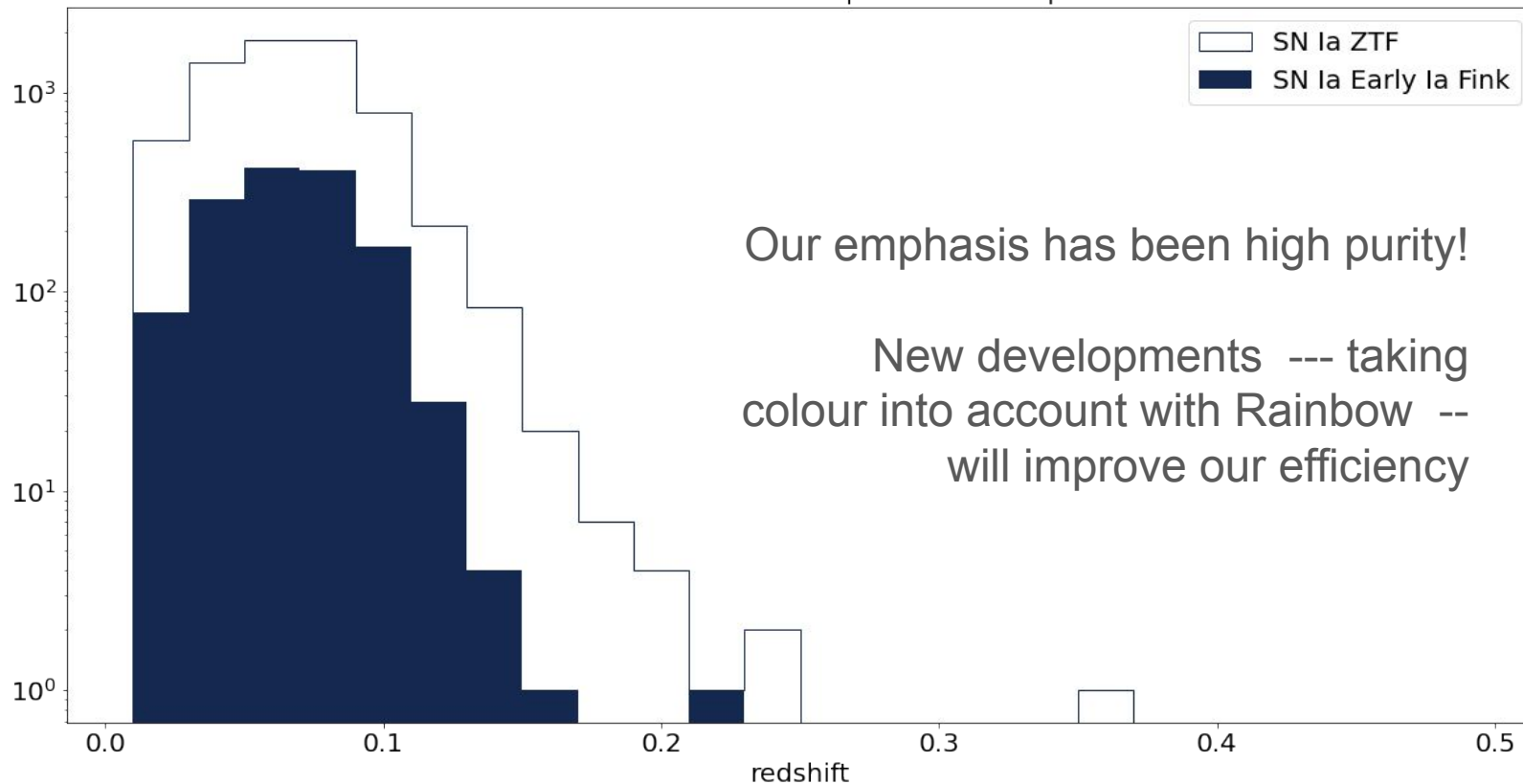
Early SN Ia Fink reported to TNS

Redshift distribution of Fink TNS reported early SNIa



Early SN Ia Fink reported to TNS

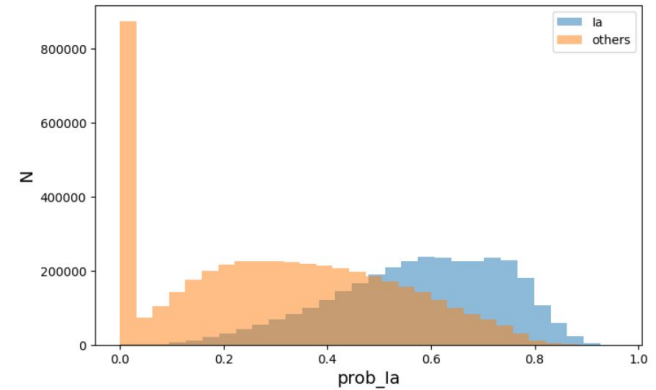
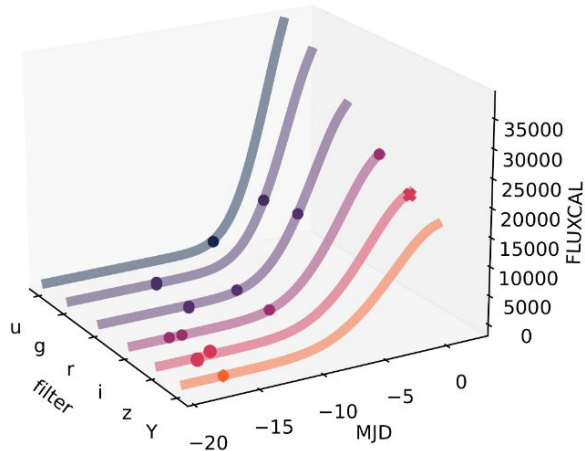
Redshift distribution of ZTF spec and Fink reported



ELAsTiCC challenge

- Cadence is far from ideal
- We needed to be able to extract features with a lower number of points

Example of light curve fitted with sigmoid + rainbow

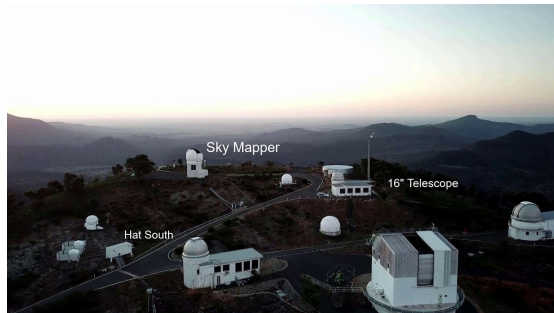


Threshold	Efficiency	Purity
0.5	0.70	0.70
0.6	0.49	0.77
0.7	0.26	0.85



Active Learning for real

Candidates in the decision boundary are sent to observation at the Siding Spring Observatory, Australia



fink-... ▾ ≡ ✎

📁 Try Slack Pro f...

- # announcements
- # anomaly-detection
- # at2023sze_ztf23...
- # bot_al_loop
- # bot_anomaly
- # bot_anomaly_area
- # bot_kilonova-can...
- # bot_known_tde_f...
- # bot_manga
- # bot_tns
- 🔒 brazil_cnrs
- # classifier_referen...
- # conferences_eve...

bot_al_loop ▾

Modified Julian Date

Thursday, May 21

5:35 =====

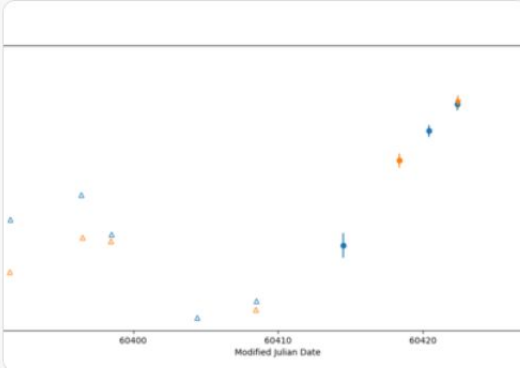

ID: ZTF24aajqwsm

EQU: 257.550506, 9.2523519

Score: 0.553

Classification: Unknown

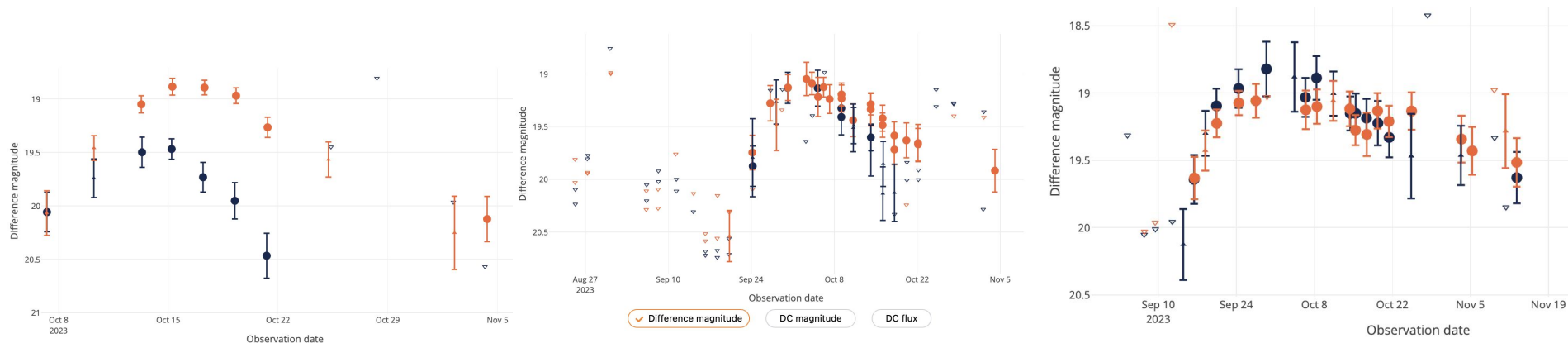
✕ 2 files ▾



AL: improving training sets

Aka: follow-up to identify early non conclusive SNe Ia or non Ia

2 CV, 16 SN Ia, 1 SN Ib, 2 SN Ibn, 1 SN Ic, 5 SN II, and others low SNR



ANU 2.3m IFU for spectroscopic follow-up + extra spectra by DEBASS and ePESSTO+

Working towards a publication later this year



Take home message

- Machine Learning models trained on real data are **complementary** to those trained on simulations
- Real data training will always be within the **small-data** regime
- Active learning is a way to remove the human from the loop, **for classification of well known classes**
- The technique is **class-independent**
- For Rubin all of this should be automated

