Multiwavelength Classification of X-ray Sources with a Supervised Machine-Learning Approach

Steven Chen[1], Hui Yang[1], Oleg Kargaltsev[1], Jeremy Hare[2], Yichao Lin[1]

1. George Washington University
2. NASA Goddard Space Flight Center

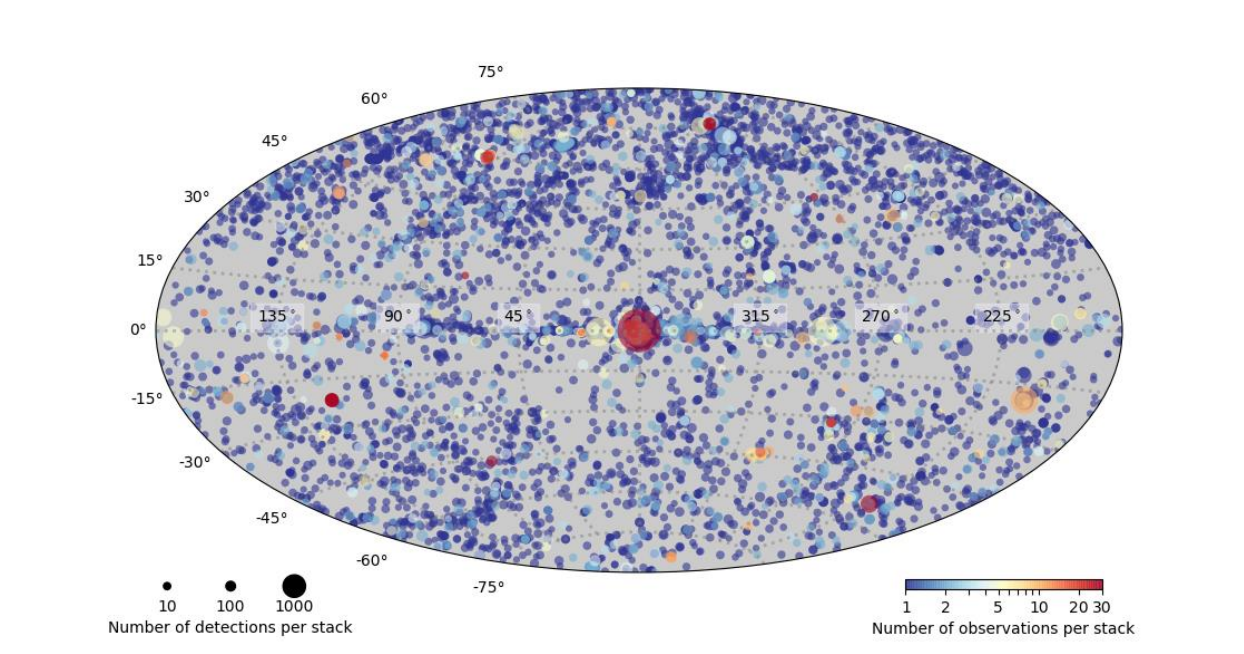
https://www.nasa.gov/mission/chandra-x-ray-observatory/

Compact Object (CO) Source Populations

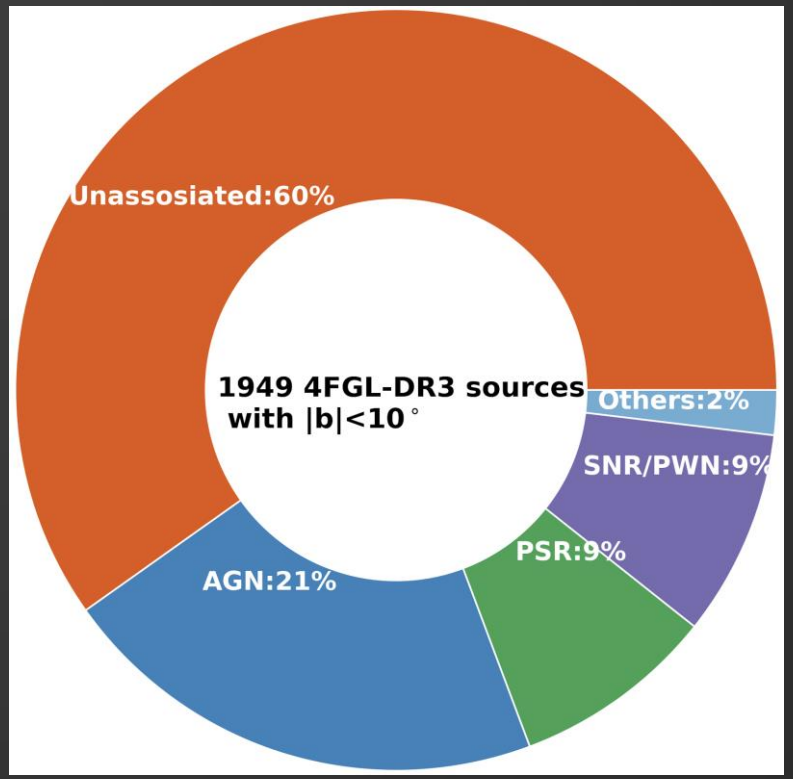
- Galactic CO population is poorly constrained
  - 3,000 NSs, ~30 BHs detected
  - $10^{8\text{-}9}$ NSs, $10^{6\text{-}8}$ BHs expected (Camenzind, 2007)
- CO populations constrain stellar evolution models, including:
  - Massive binary evolution
  - Supernova explosion physics
  - Binary CO merger rates
- Many COs exist in dense stellar environments.
- Growing population of unclassified Gamma-ray sources (GeV, TeV), highly likely to be Blazars or COs.

Modern X-ray observatories find millions of serendipitous sources, most of which remain unclassified

Catalogs	# of unique X-ray sources
Chandra Source Catalog (CSC) 2.1	~500k
4XMM-DR13	~650k
eROSITA All-Sky Survey DR1	>1 million



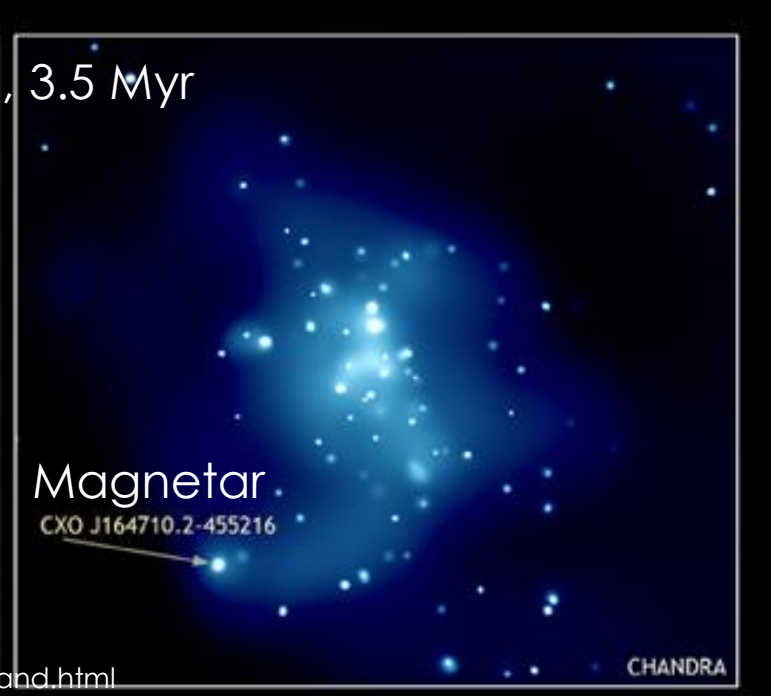
CSC v2.0 Detection Map from CXO

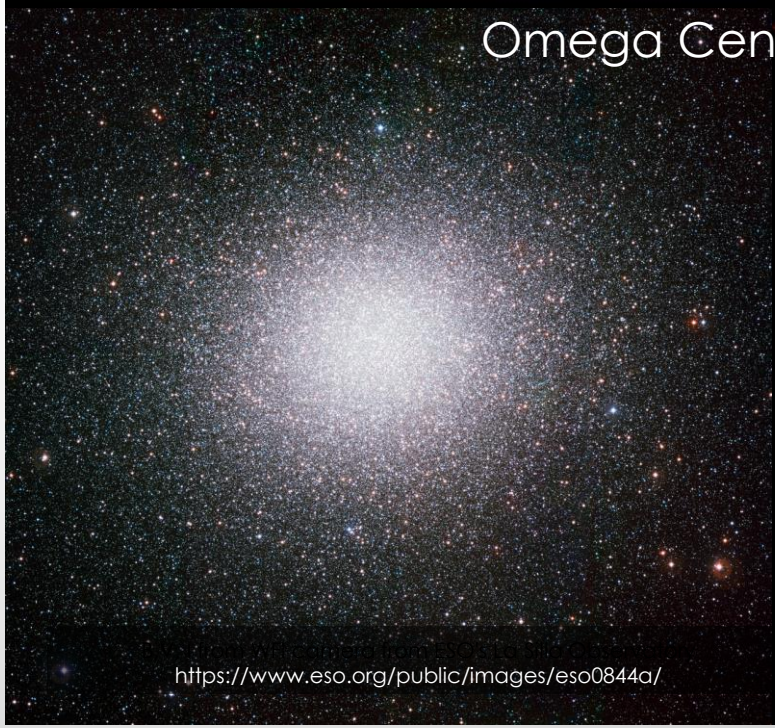


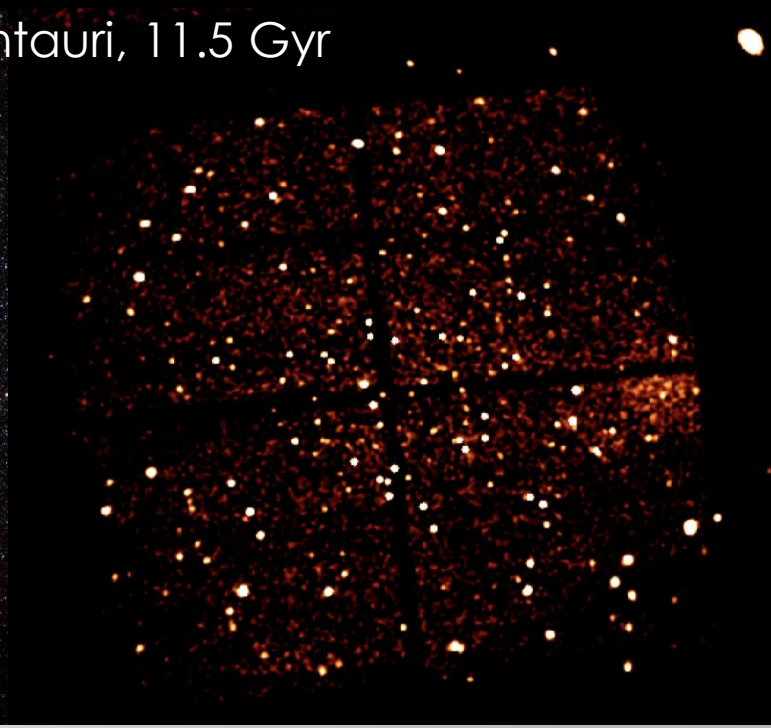
Breakdown of 4FGL classifications for Galactic plane sources within |b|<10°

Star Clusters

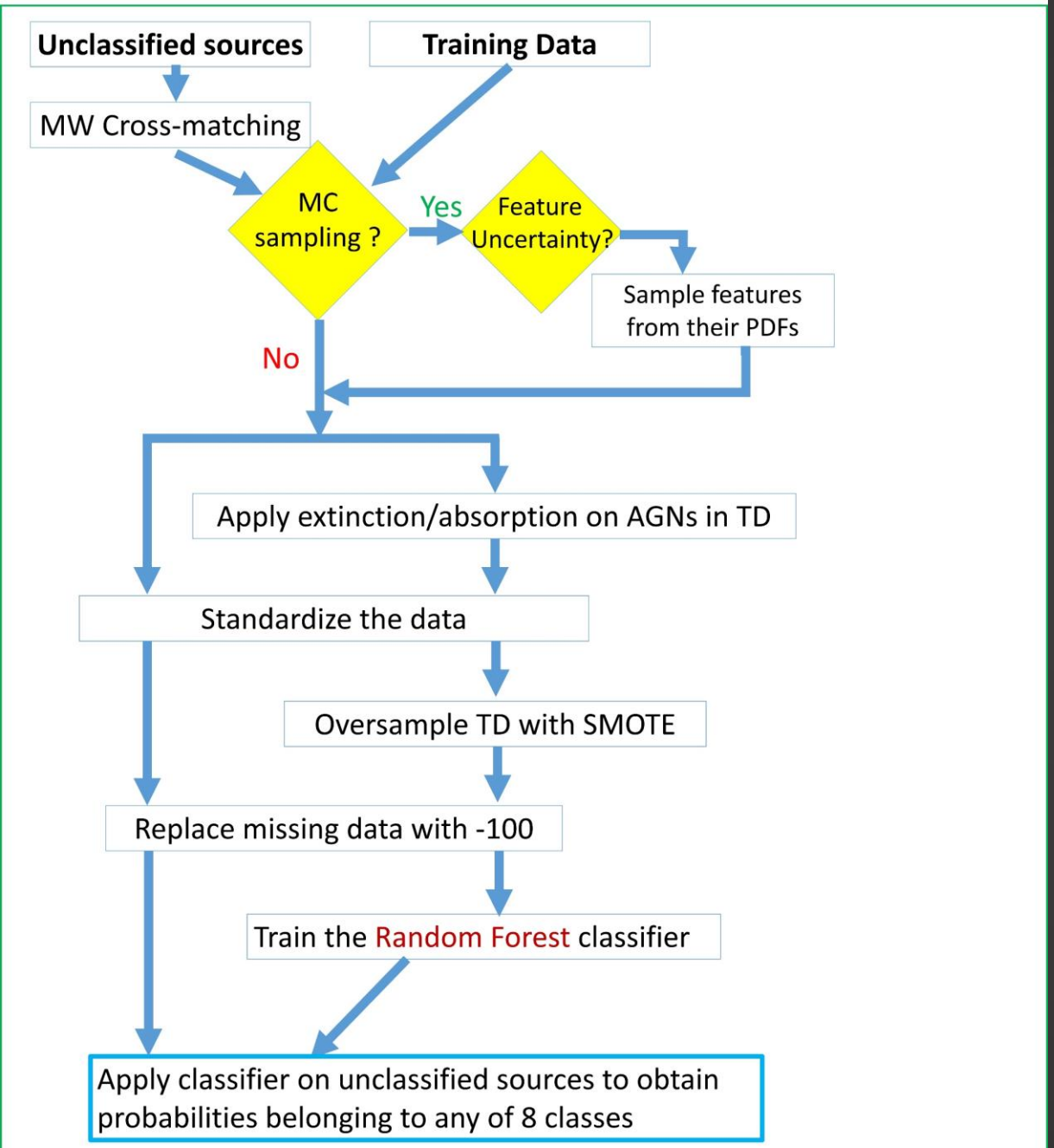
- Gravitationally bound groups of stars
- Open Clusters
  - Young (mostly <1 Gyr)
  - Less massive (<$10^5\ M_\odot$)
  - In galactic disk
  - Younger ones may contain recently formed CO systems (e.g., magnetar in Westerlund 1)
- Globular Clusters
  - Old (~10 Gyr)
  - Massive (>$10^6\ M_\odot$)
  - Off galactic disk
  - Contains many CO systems
- Contains thousands of unclassified Chandra sources.







MUWCLASS pipeline flow chart



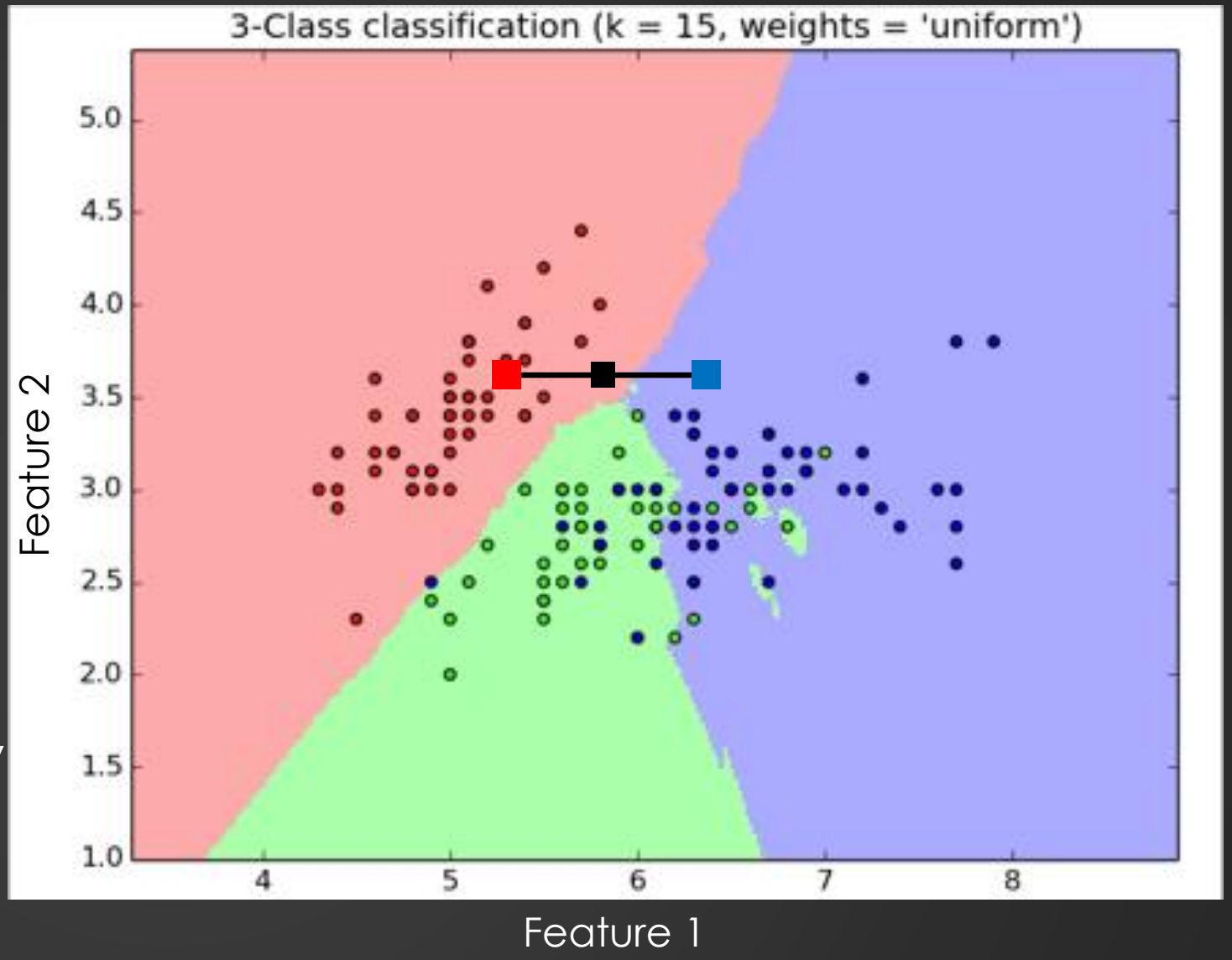
MUWCLASS

- 30+ multiwavelength (MW) features:
  - Gaia, 2MASS, WISE, HST
  - X-ray Fluxes
  - Magnitudes
  - Colors
  - CSC Variability
  - Luminosities
- Multiple random forest classification runs, each time sampling feature uncertainties
- Probabilistic cross-matching to MW counterparts (NWay)
- Correcting for extinction bias on AGNs in TD
- Correcting class imbalance

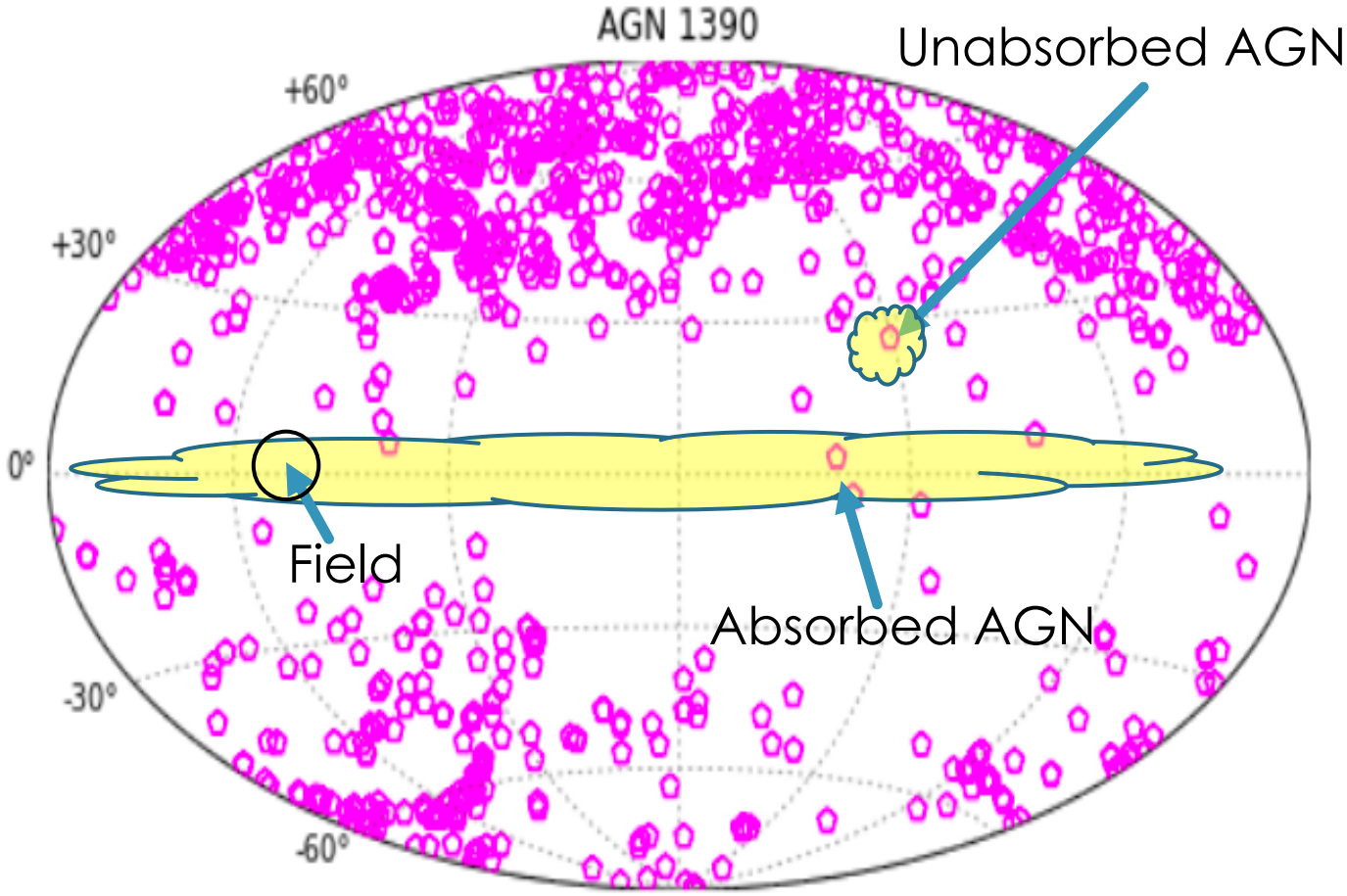
Uncertainty Sampling

- Samples uncertainty in fluxes, magnitudes, other features
- Applies to TD and unclassified sources
- Uncertainties may impact classification
- Multiple random forests, each sampling uncertainties randomly

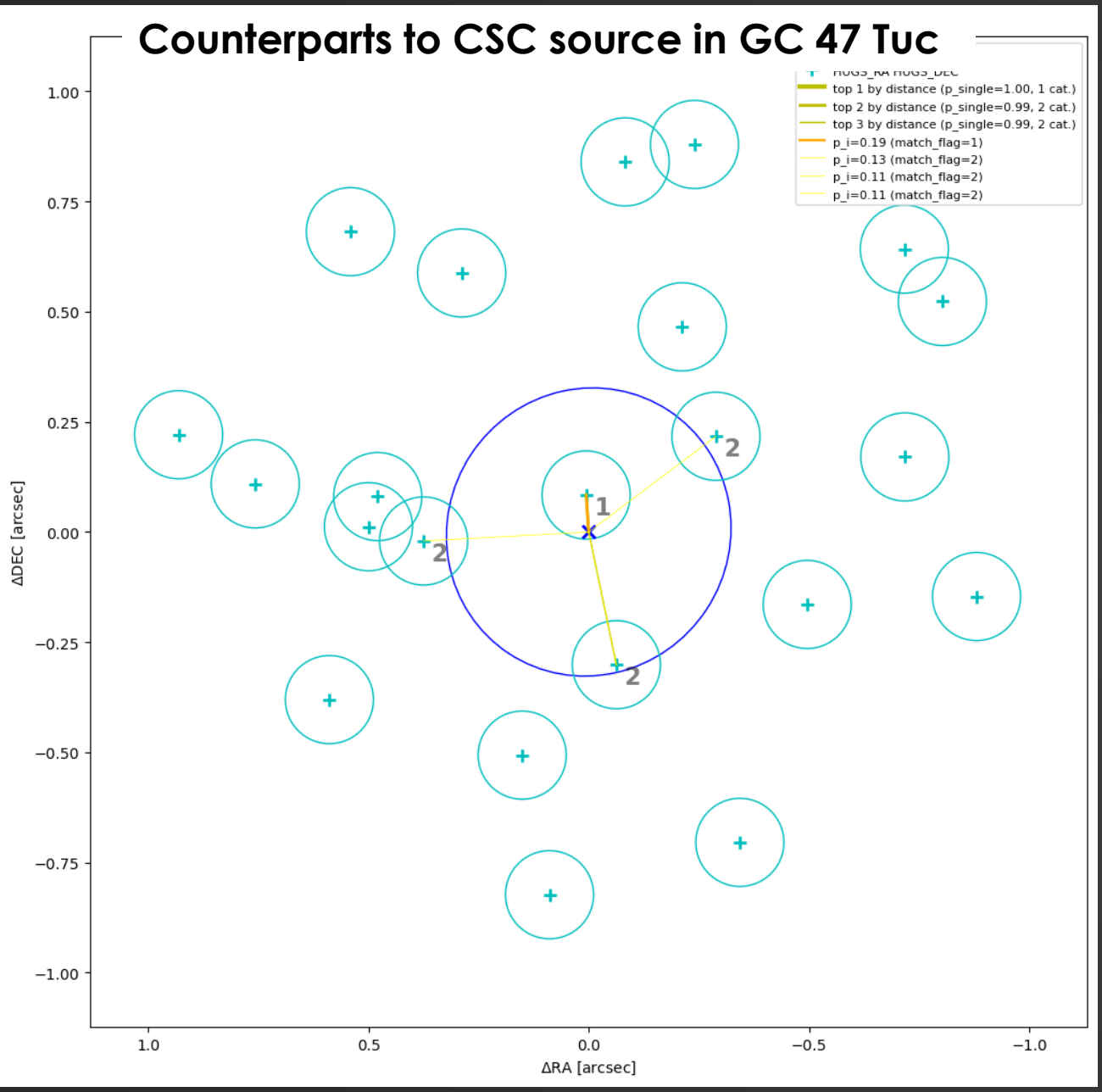
Feature 1 Value	Red	Blue	Green
5.8	0.6	0.2	0.2
5.3	1	0	0
6.3	0.1	0.85	0.05



Correcting Extinction/Absorption bias for AGNs in TD



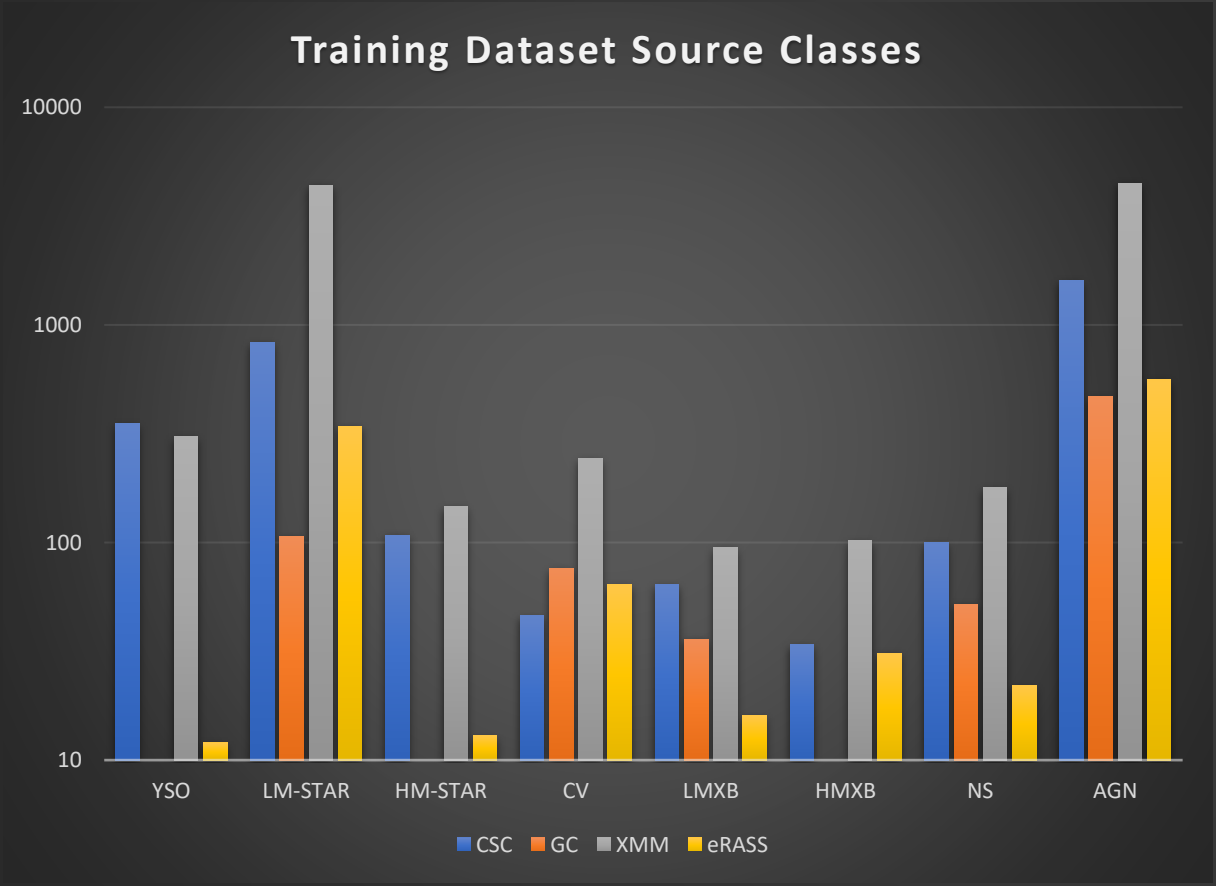
- Most AGN in TD off galactic plane, low absorption
- Real AGN appear very different in plane
- Need to correct for bias:
- Redden all AGN in TD in direction of source to be classified
- Deredden all sources in TD and field to be classified
  - Only possible if extinction up to distance of sources known
  - Practical for globular clusters



Probabilistic Cross-matching

- Multiwavelength properties crucial for classification
- Assigning correct multiwavelength counterpart tricky in dense environments
- How to choose association when multiple counterparts of same X-ray source exist?
  - Use closest counterpart
  - Use most confident classification
  - Consider both factors
  - Combine classification probabilities of different associations
- Automation requires probabilistic approach, possible with NWay (Salvato 2017)

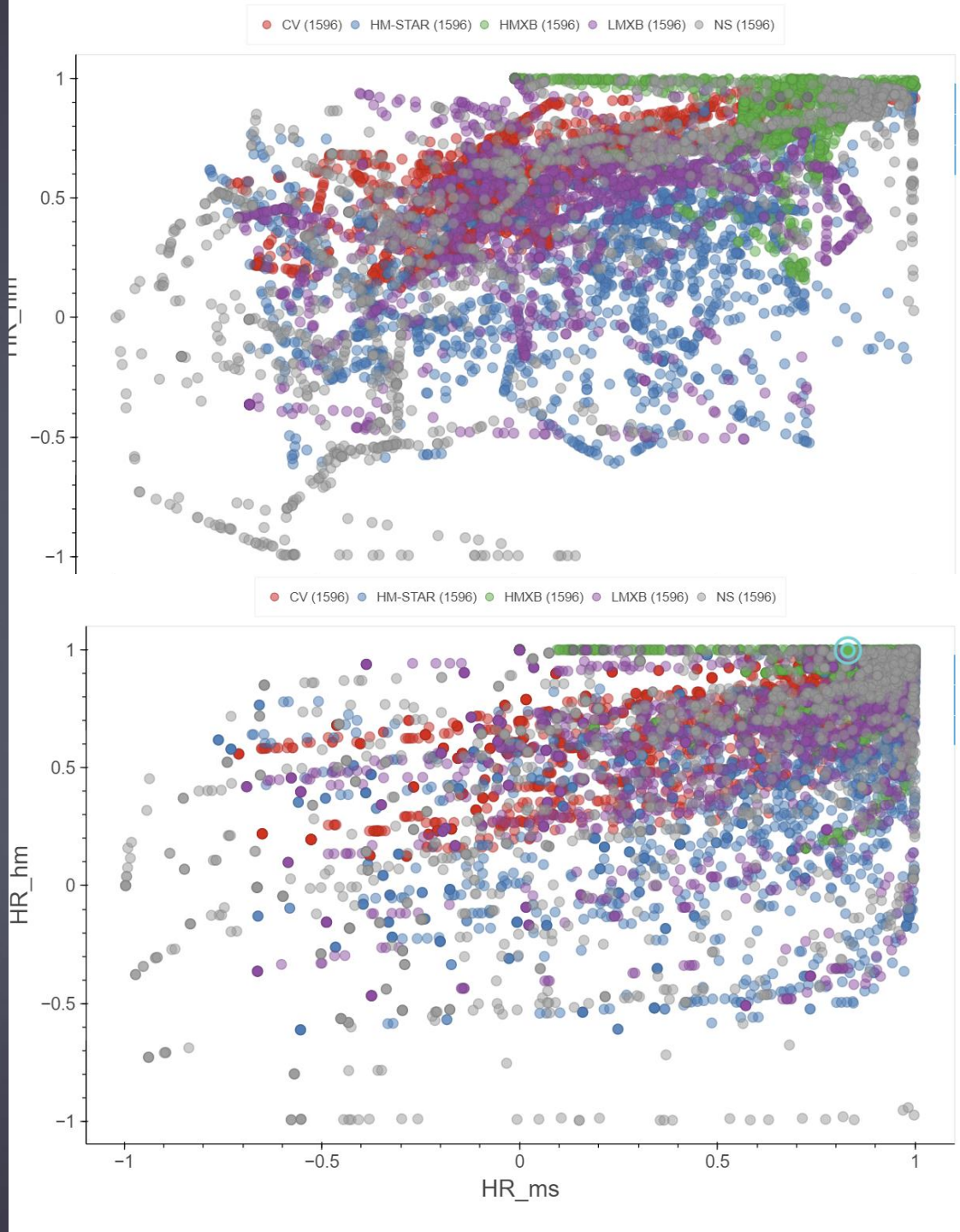
Training Datasets



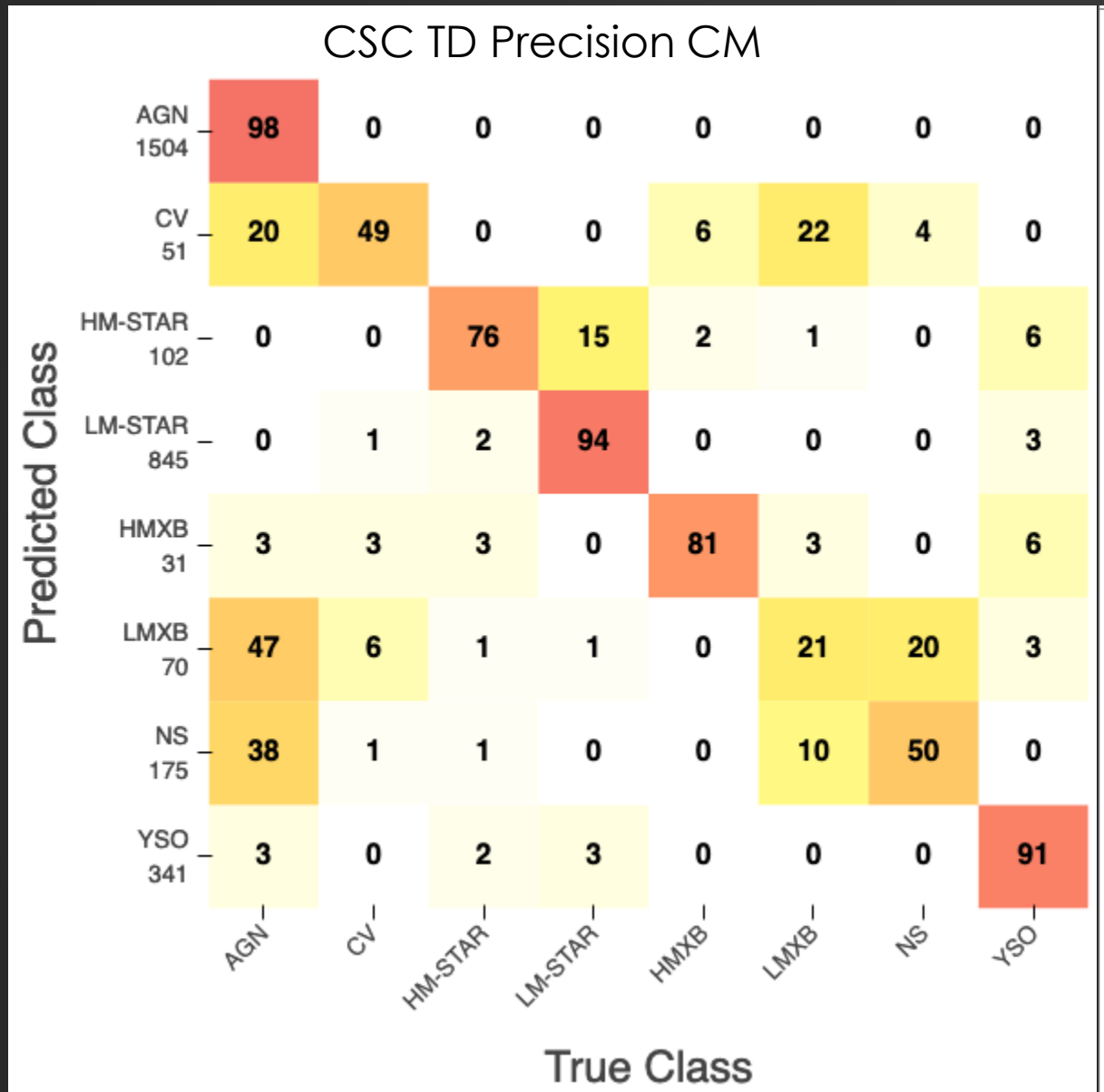
- Based on cross-matching to classified sources in literature
- CSC TD
  - Gaia, 2MASS, WISE
  - All sky, published in Yang et al. 2022
- GC TD
  - Very old, very dense
  - Requires HST counterparts
  - Separate TD, published in Chen et al. 2023
- 4XMM TD
  - Cross-matched to CSC first to get improved X-ray positions, when possible
  - All sky. Submitted as RN, Lin et al. 2024
- eROSITA TD
  - Based on cross-matching to other TDs
  - Very Preliminary!

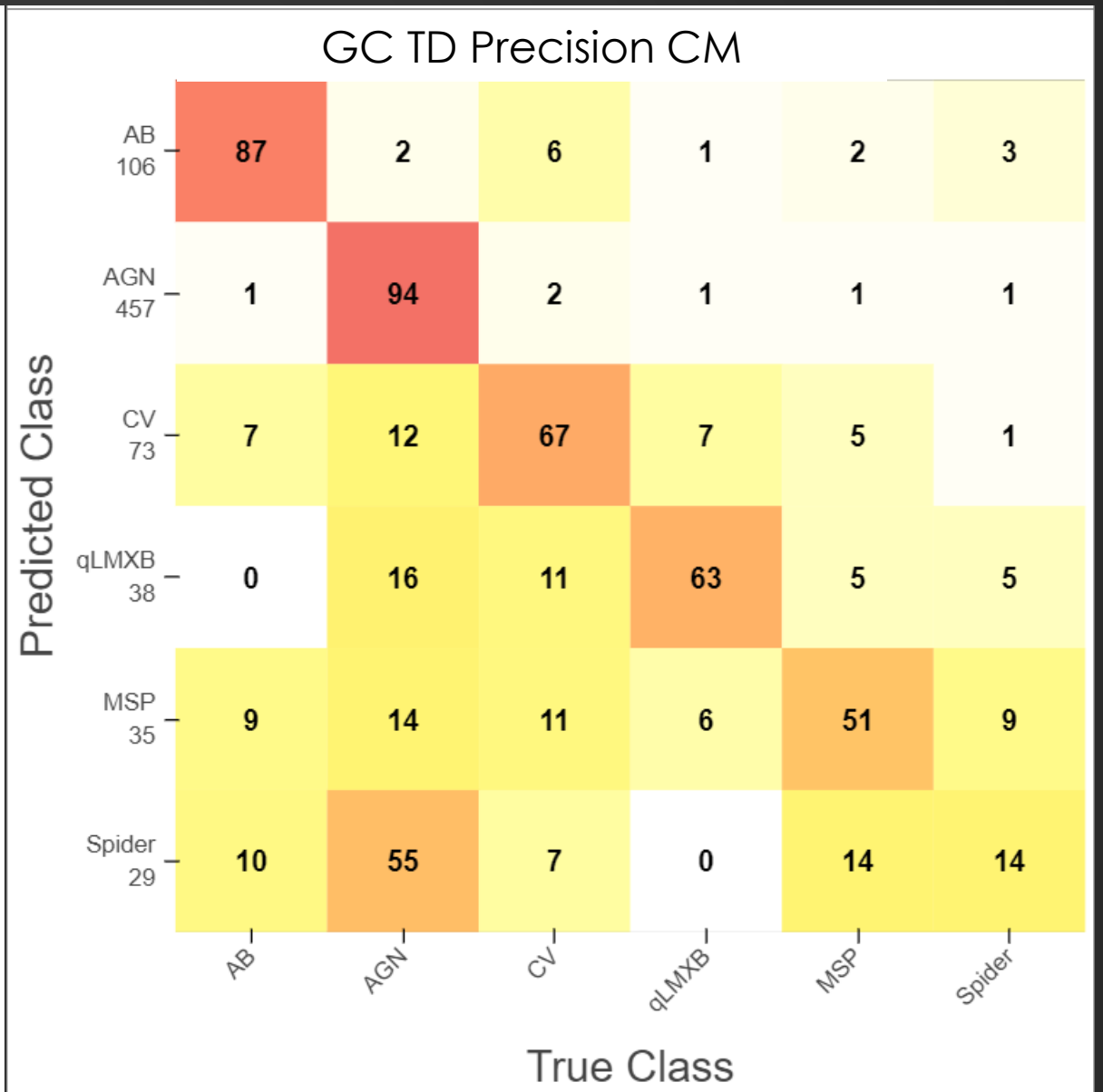
Class Imbalance

- Some classes overrepresented (AGNs), while some classes rare (NSs)
- SMOTE samples points in feature space between objects of one class
  - Very linear, unphysical
- We developed a method of physically motivated oversampling:
  - Sample sources randomly reddened and absorbed in feature space according to per class distribution in TD



LOO Evaluation: Confusion Matrices



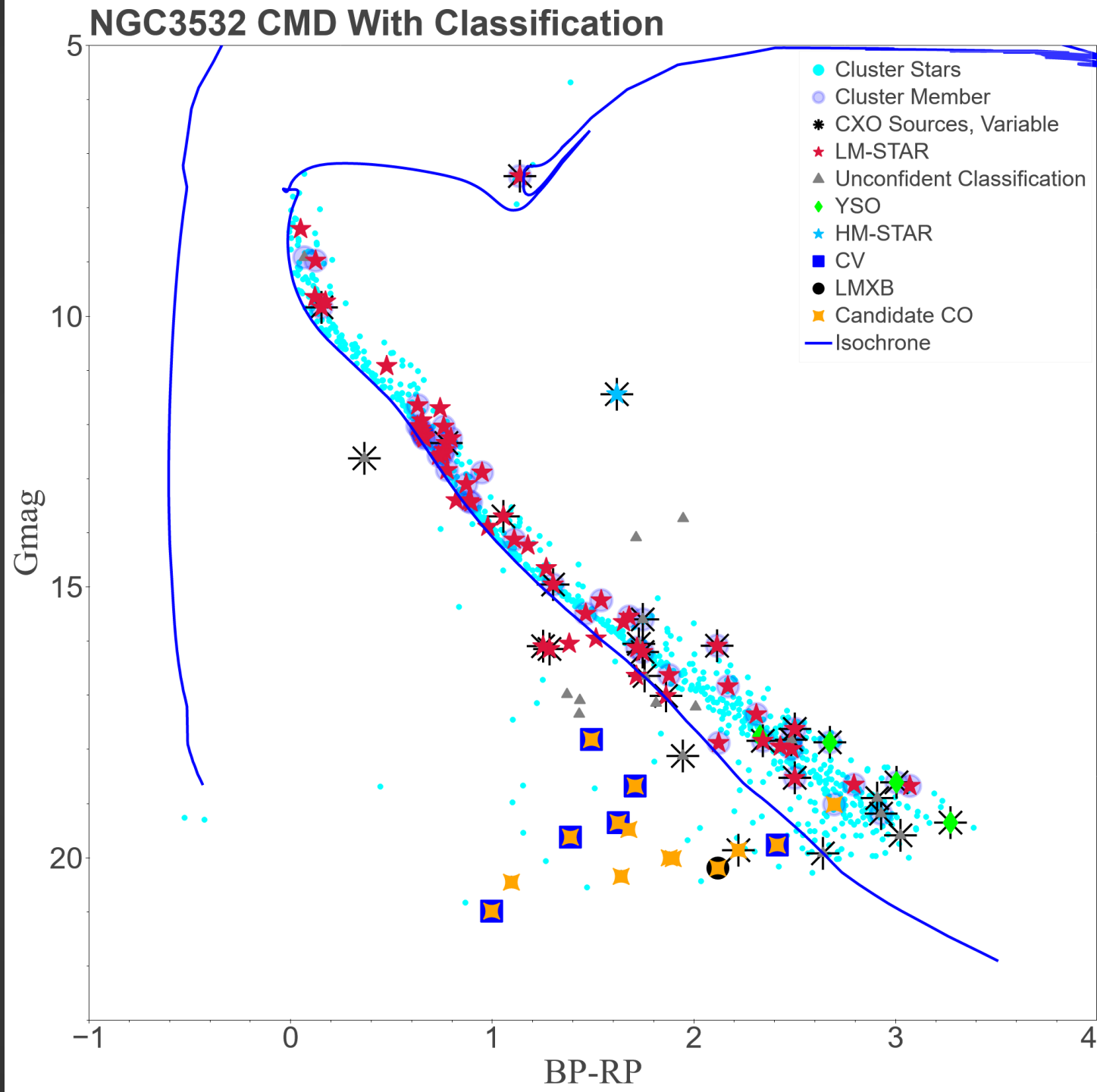


Evaluation

- Well-populated classes have high degree of classification accuracy
- CO classes are underpopulated in TDs, do not perform as well
  - Often confused with AGNs due to similarities in X-ray features
  - Intrinsically diverse
  - Trade-off between more populous but more heterogeneous classes, and less populous but more homogeneous classes
- Biases:
  - Sources present in TD are brighter/closer/less absorbed compared to sources we classify
  - Faint AGN populations in deeper fields may be different from bright AGN in TD
- Missing MW counterparts may be due to lack of depth, or intrinsic faintness, which may have physical significance that MUWCLASS is agnostic to.

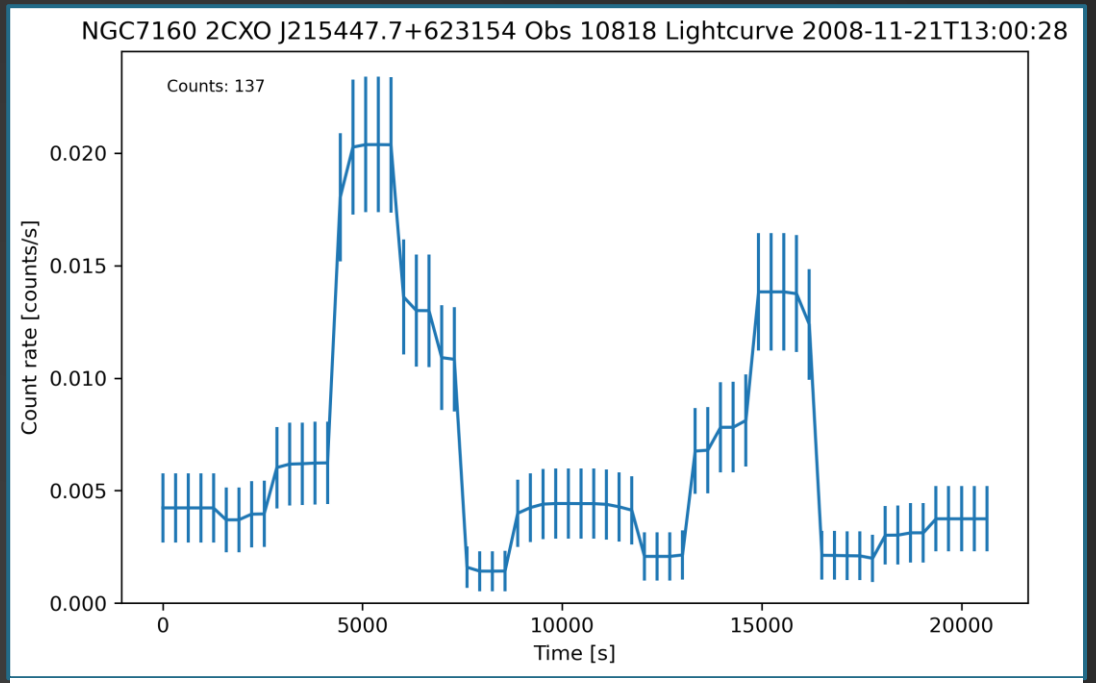
CSC Sources in Open Clusters

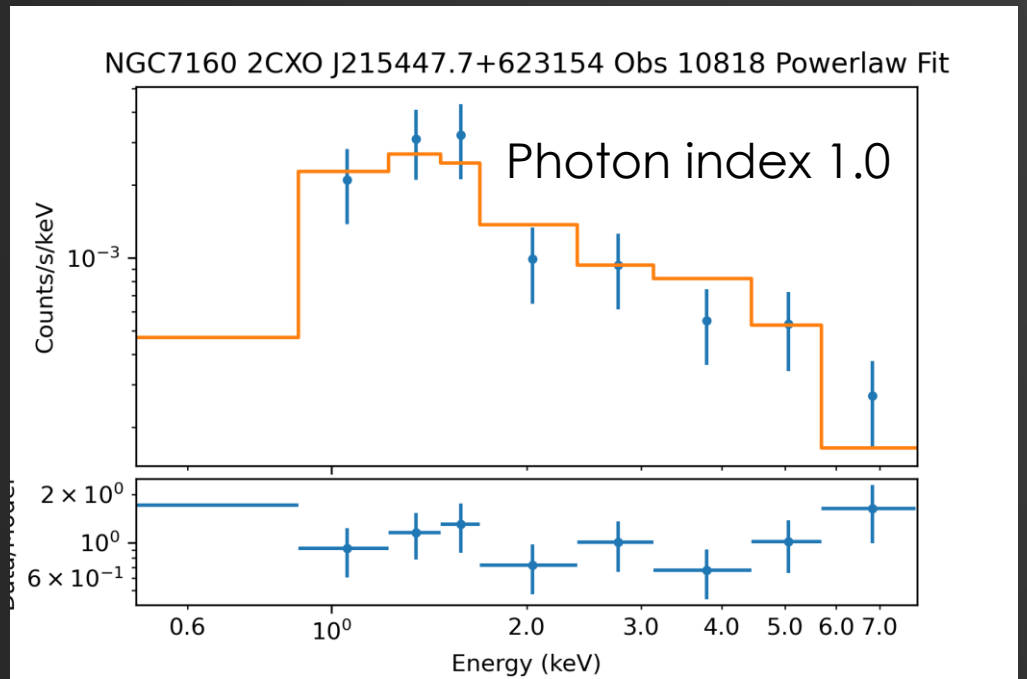
- High confidence of LM-STAR classifications due to presence of main sequence
- Confidently identified coronally active stars allows for studying stellar activity as a function of age, mass, rotation period.
- Cluster member COs expected to be rare but can be interesting (e.g., electron capture SN remnants)
- Discovered several serendipitous background AGN and candidate COs

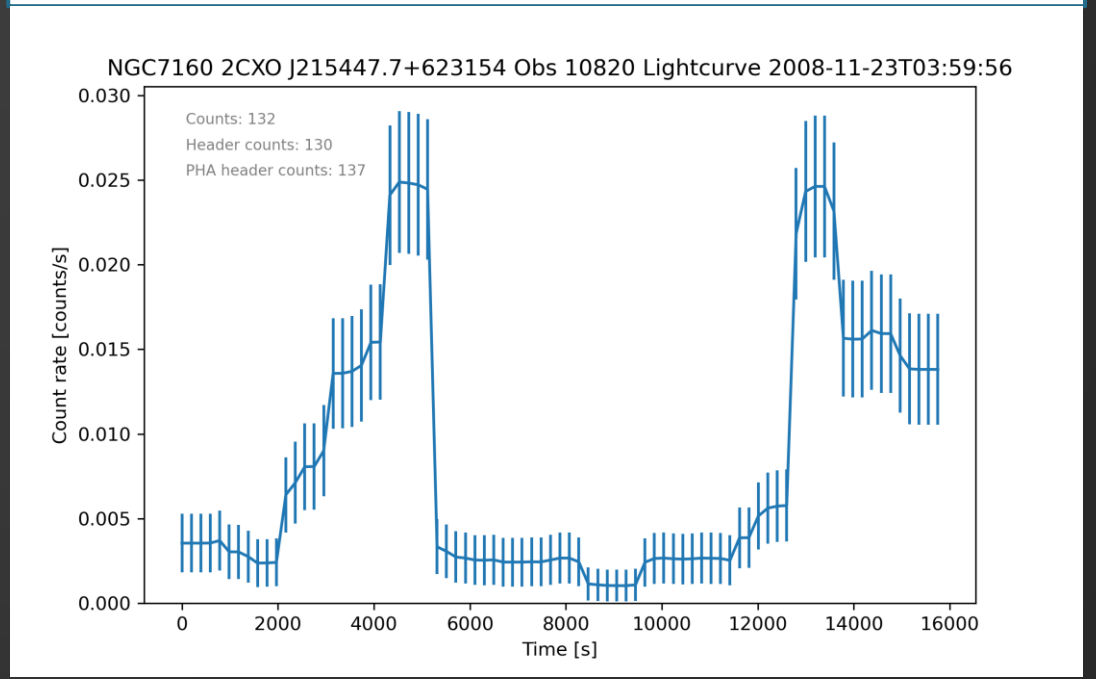


Example of Candidate CO: NGC 7160

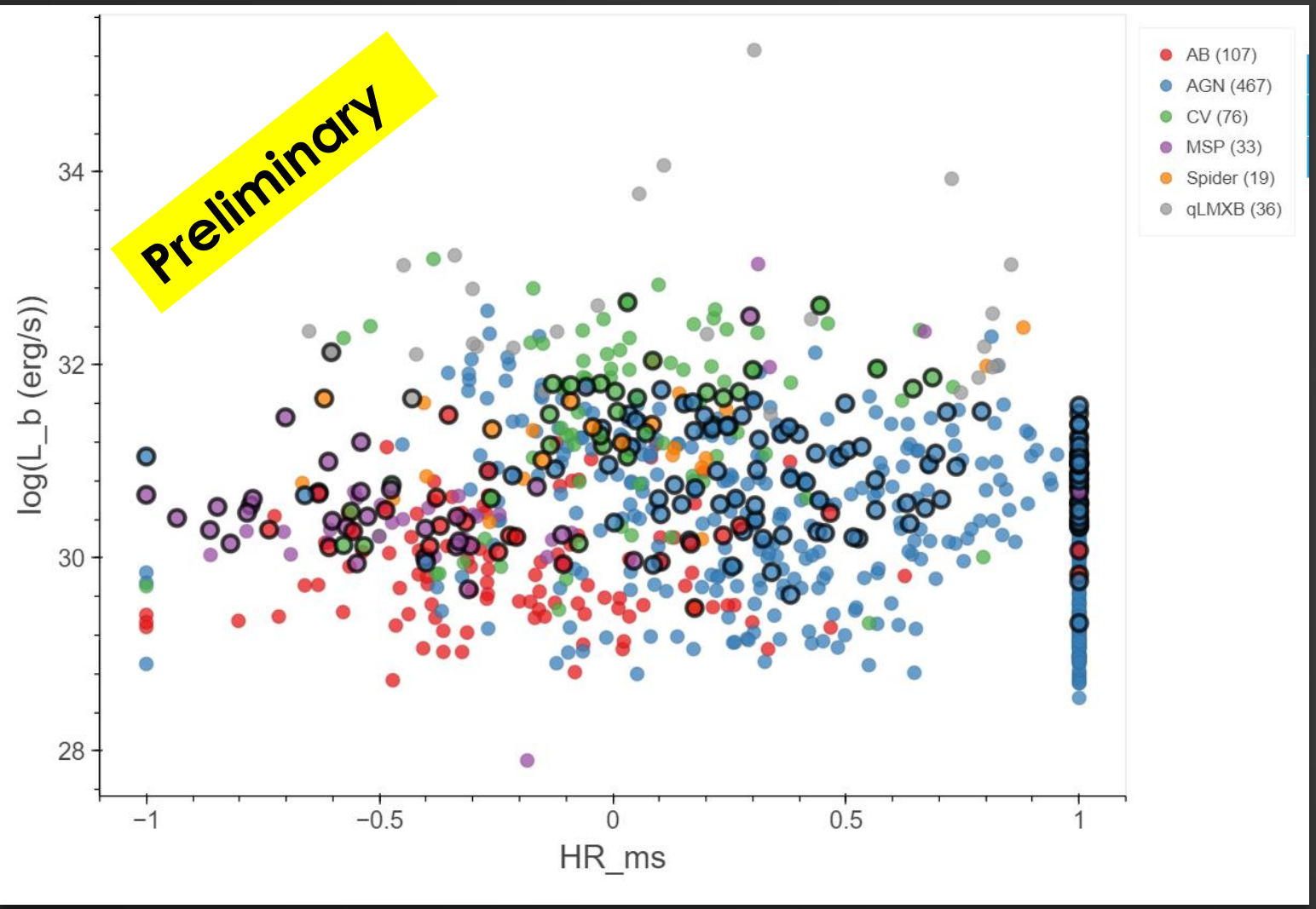
- MUWCLASS classified as CV
- No optical counterpart in Gaia
- Clear ~10 ks periodicity in 4 Chandra observations
- Previously suggested as CV in literature





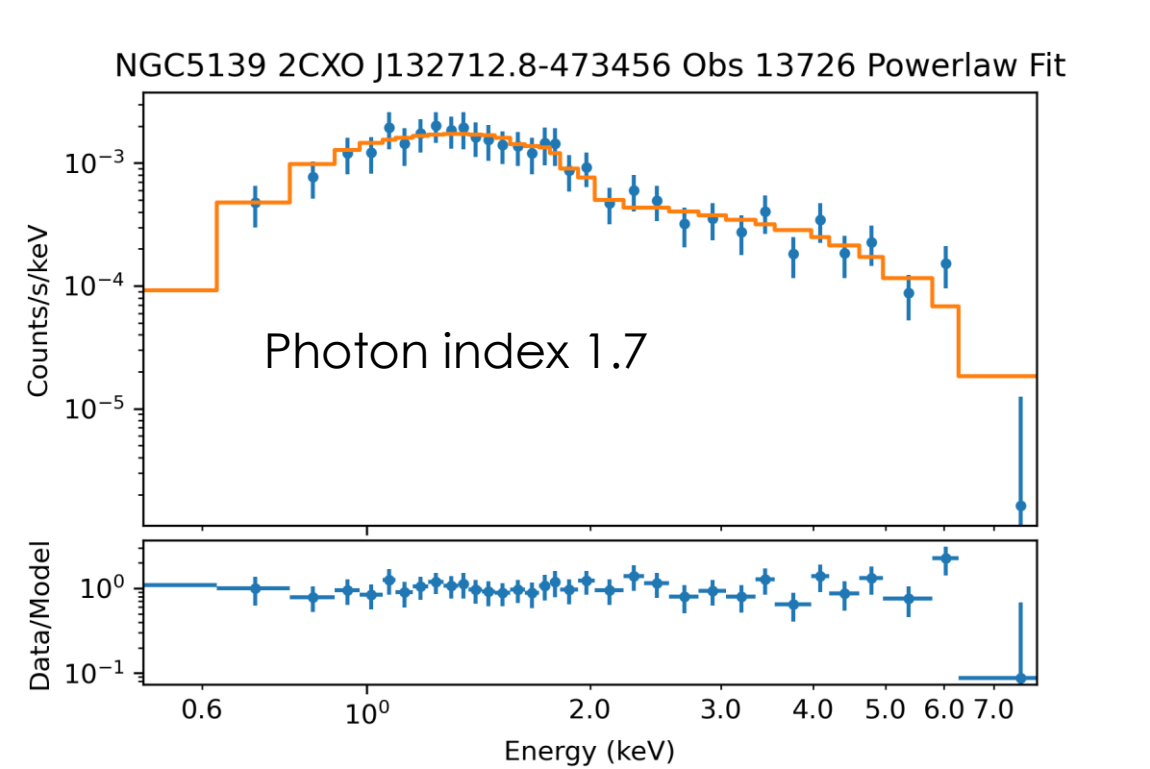


CSC Sources in Omega Centauri

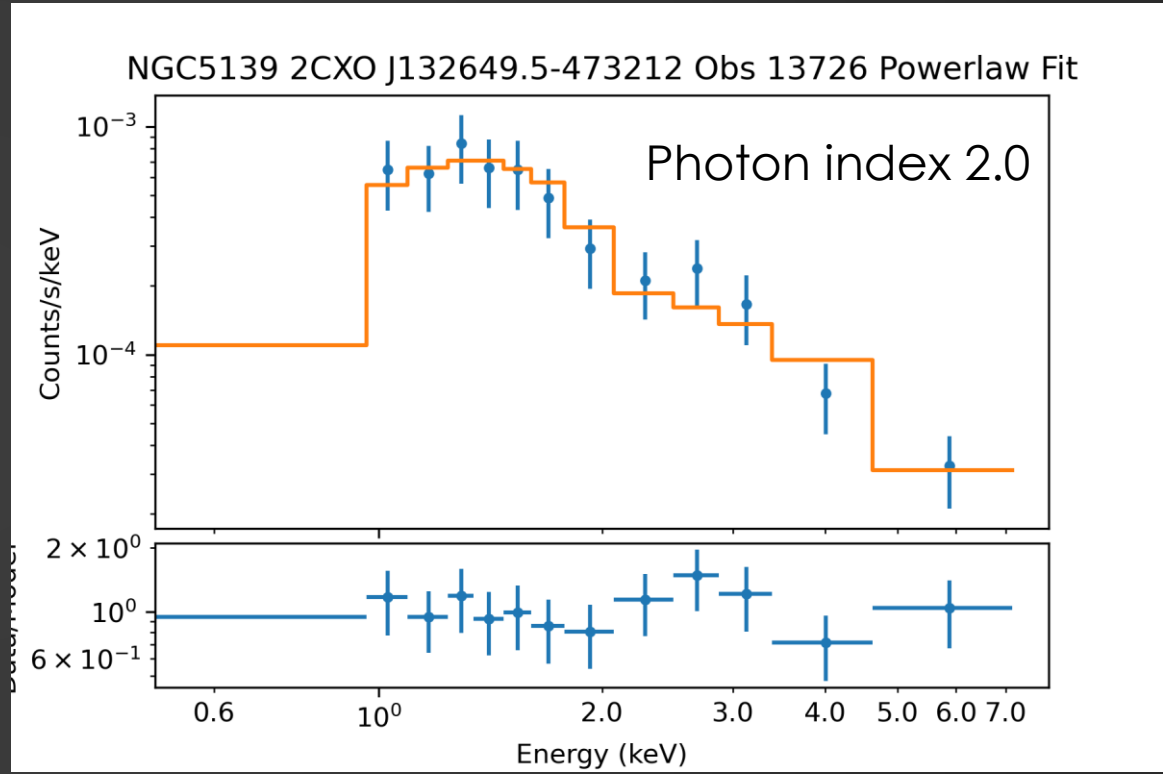


- Classified sources outlined by black circles, TD sources are not
- ~25% of source classifications depend on the choice of HST counterpart
- Classifications mostly correct for 18 sources with known class, even when removed from TD
- Many classified CVs, due to similarity to TD CVs in X-rays and optical features
- Large number of classified MSPs due to similarity to TD MSPs in X-rays, while lacking HST counterparts
- Requires more detailed analysis

Interesting Sources in Omega Cen

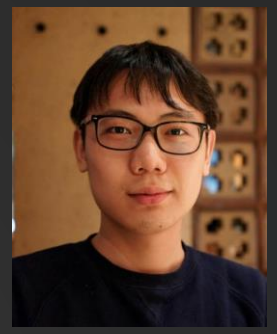


Bright source classified as CV with all possible counterparts



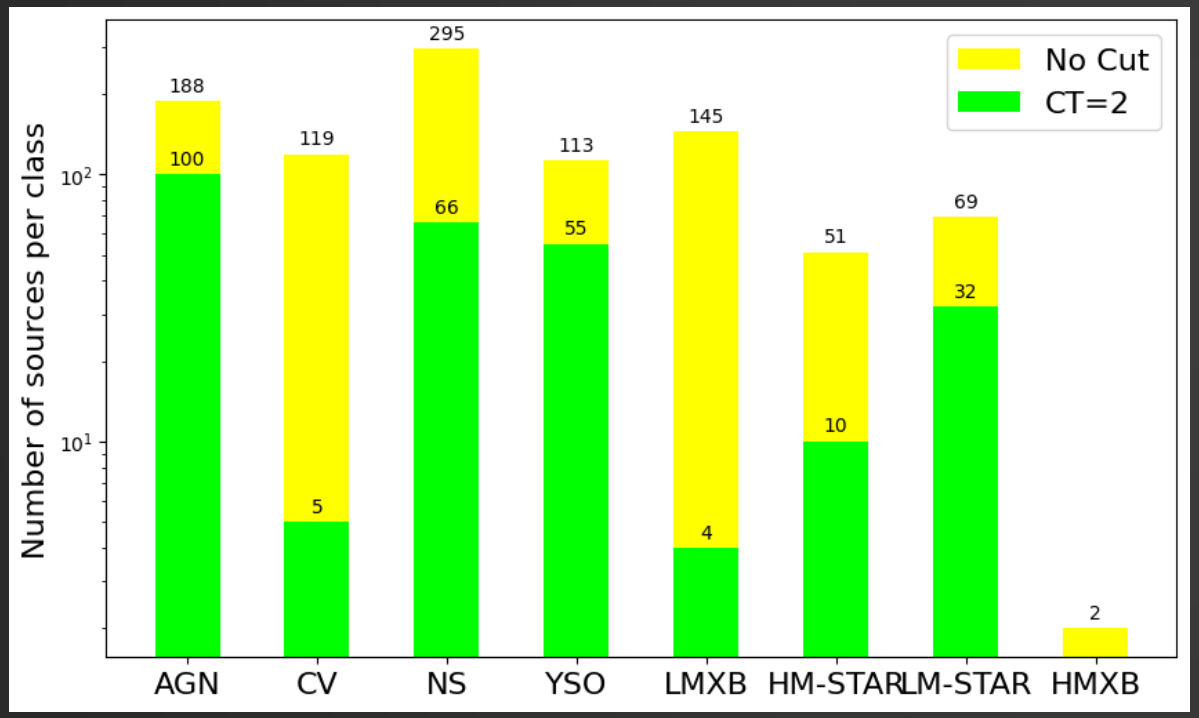
Bright source classified as spider with all possible counterparts

Exploring Unidentified 4FGL-DR4 Sources (unIDs) with MUWCLASS (to be submitted)

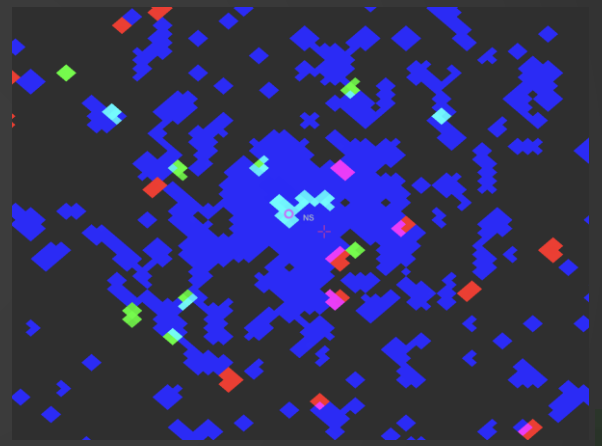


Hui Yang
job hunting in 2024

- Classified 1206 CSCv2 X-ray sources within 73 unIDs
- Identified 107 X-ray sources as potential X-ray counterparts to GeV sources.
- Candidate unID classifications: 3 NS, 2 XRB, 13 AGN, 2 SFR, 21 ambiguous

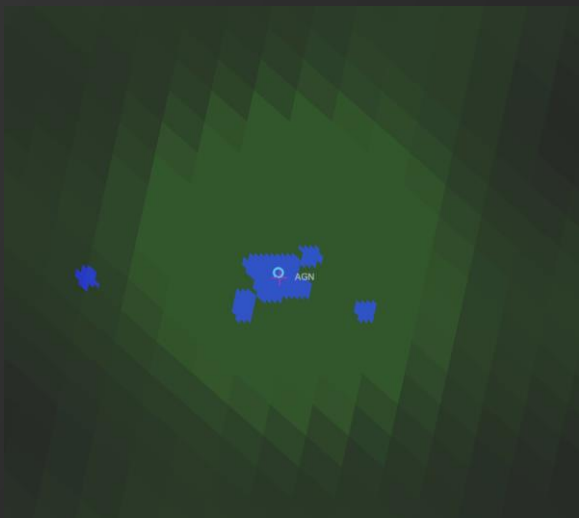


- Classification breakdown of X-ray sources within 95% error ellipses of unIDs
- Accounting for feature uncertainties help selecting confident classifications cutting at classification confidence threshold (CT)



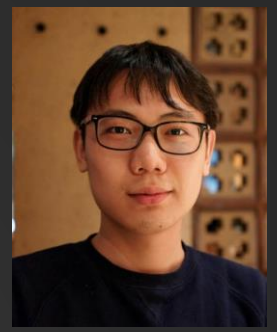
- 2CXO J184443.3–030518 in 4FGL J1844.4–0306
- Classified as NS
- Overlaps with extended emission, possible PWN.

- 2CXO J005806.2-460419 in 4FGL J0058.3-4603.
- Green: radio, Blue: CXO
- Classified AGN/Blazar
- Addition of radio surveys important for future development

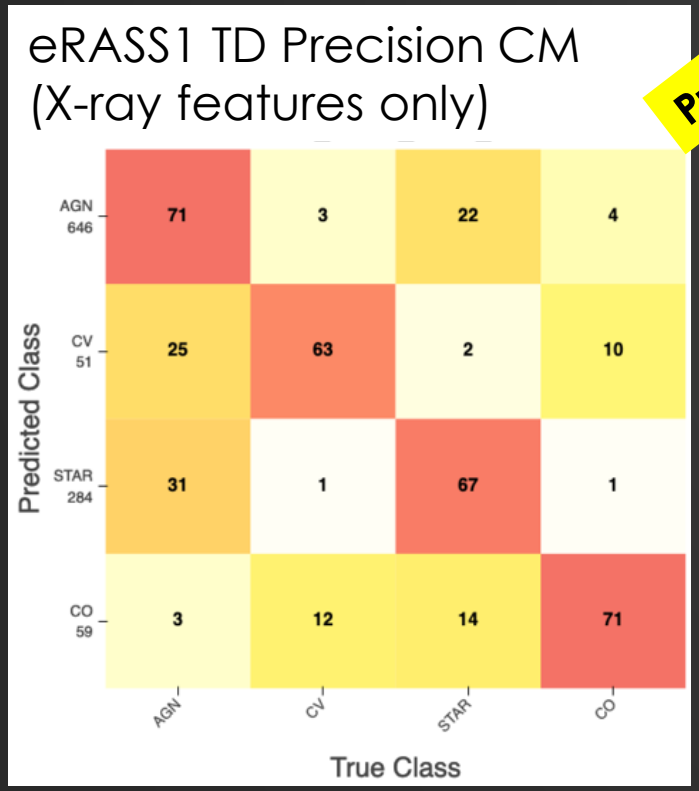


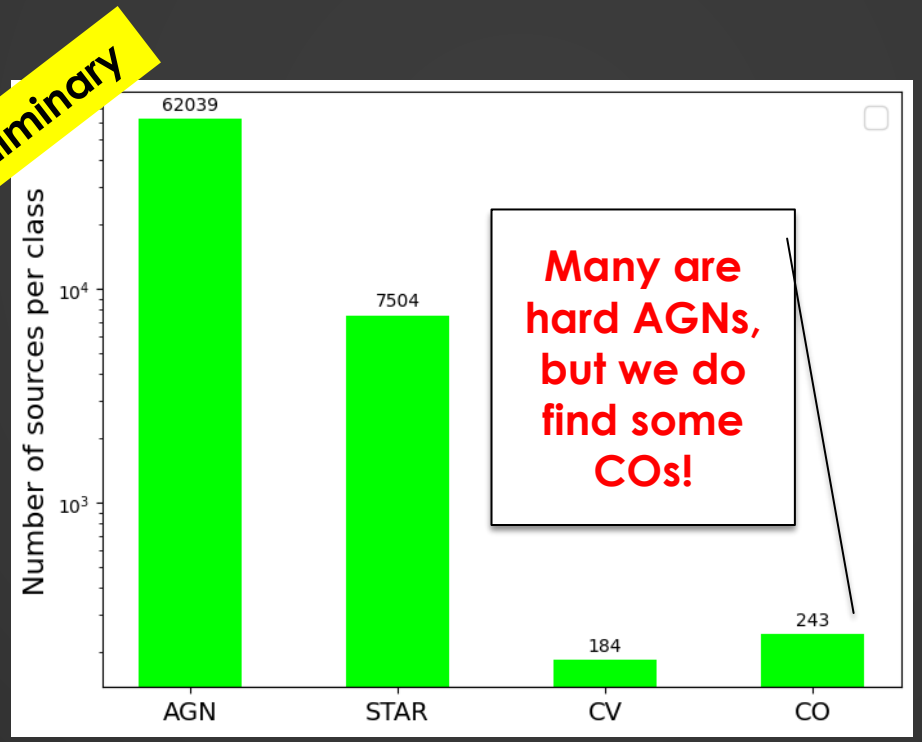
Preliminary classification of eRASS1 sources: Hunting for COs

- Classified ~1M eRASS1 sources using only X-ray features (band fluxes, HRs), and Identified 243 CO candidates.
- Adding MW features narrows down to 29 CO candidates.

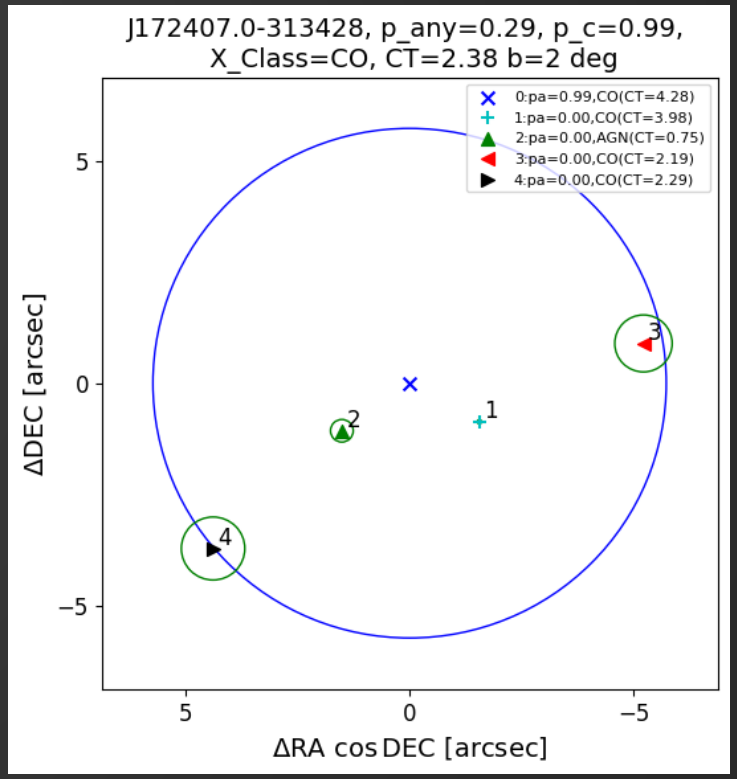


Hui Yang
job hunting in 2024





Classification breakdown of eRASS1 sources using only X-ray features



Example of counterparts in source field

Conclusion

- MUWCLASS is a powerful tool for rapidly classifying many sources in different environments
  - Can substantially increase statistic for population studies of confidently classified source classes, e.g., flaring stars and AGNs.
  - Identify unusual/interesting sources for follow up observations.
  - Classify sources at other wavelengths (e.g., radio, Gamma rays).
- Future improvements:
  - Integration of additional sensitive surveys, including radio.
  - Galactic Plane survey: Only CXO has sufficient angular resolution. AXIS mission concept ideal for future surveys.
  - Expansion of TD. Community-based living database of classified X-ray sources.