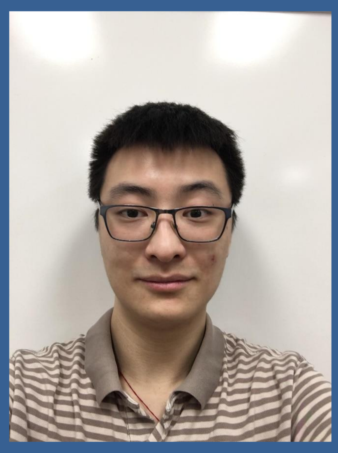
Exploring the X-ray Source Population in Globular Clusters with a Machine Learning Approach

Steven Chen[1], Hui Yang[2], Oleg Kargaltsev[1], Jeremy Hare[3]
[1]George Washington University, Washington, DC; [2]Institut de Recherche en Astrophysique & Planétologie, Toulouse, France; [3]Goddard Space Flight Center, Greenbelt, MD



Globular Clusters

- Around 150 globular clusters (GCs) are found in Milky Way (Kharchenko et al. 2013).
- >10 Gyrs old, masses range from ~10^5 to ~10^7 $M_\odot$
- X-ray sources detected in GCs include active binaries (ABs), Cataclysmic Variables (CVs), isolated or binary millisecond pulsars (MSPs), and low mass X-ray binaries (LMXBs).
- Only Chandra + HST observations have sufficient angular resolution to securely establish counterparts to X-ray sources in GCs, especially fainter ones.
- Great progress made in identifying X-ray sources in past two decades, constraining models of binary evolution, cluster dynamics, and compact object physics (Zhao & Heinke, 2022; Bahramian et al. 2020, etc.).
- However, 80% of ~2000 CXO sources in all Galactic GCs remain unclassified.
- Large number of sources, large number of source properties, and multiple possible counterparts and motivate machine learning (ML) classification approach.
- We present our training dataset of X-ray sources in GCs confidently classified in the literature, and classification results on the GC Omega Centauri (NGC 5139).

Omega Centauri (NGC 5139)

- Largest known galactic GC
- d=5.2 kpc, $M=4\times10^6 M_\odot$, age≈11.5 Gyr (Henleywillis et al. 2018)
- Hosts multiple stellar populations, suspected to have a dwarf galaxy origin
- Extensively observed by Chandra, classified sources include 1 qLMXB, CVs, recently discovered Spider MSPs (Henleywillis et al. 2018, Zhao & Heinke 2023)
- 75% CSC sources not classified!
- Recent HST catalog oMEGACat (Häberle et al. 2024)

Machine Learning Classification Pipeline (MUWCLASS)

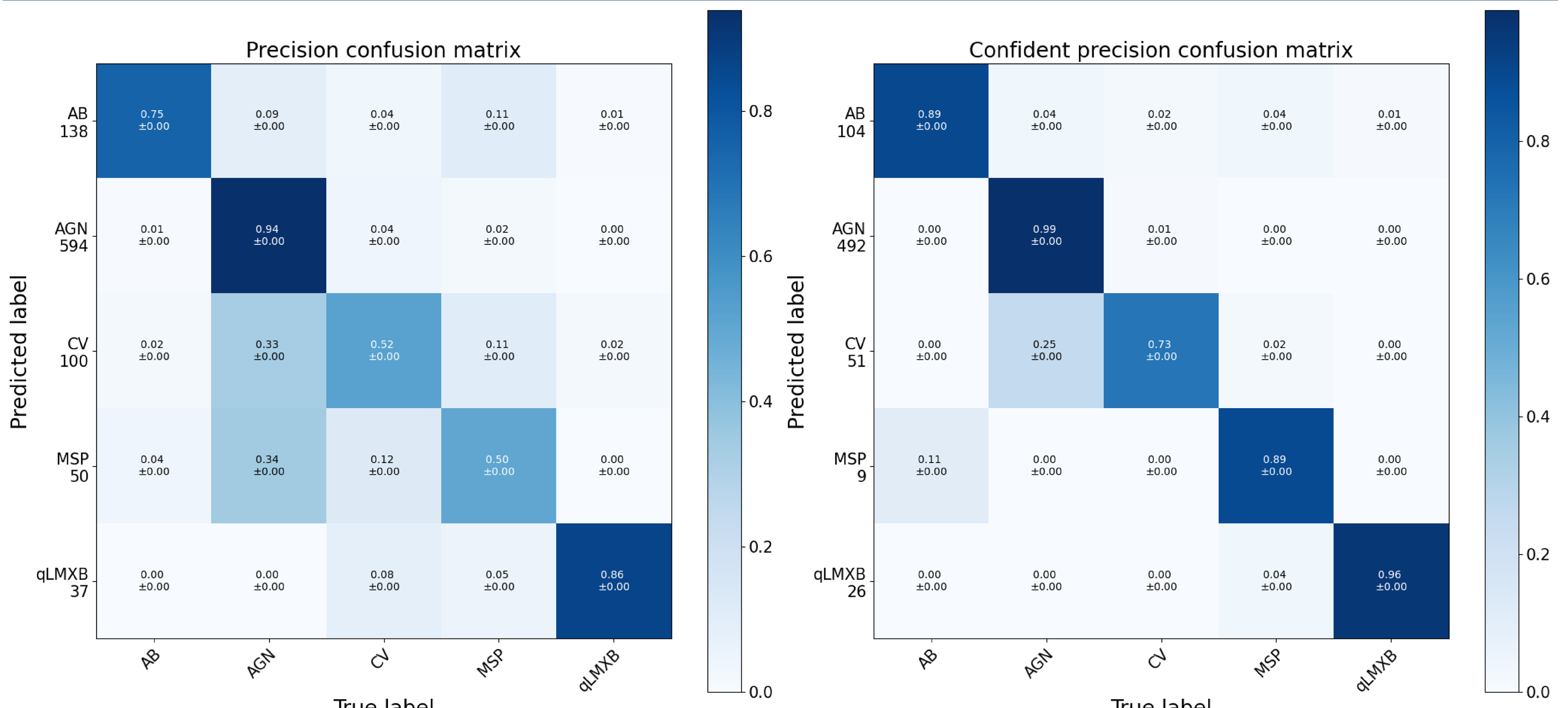
- MUWCLASS originally designed for classifying CXO sources in non-GC environments (Yang et al. 2022, 2024)
  - Random Forest Algorithm (scikit-learn)
  - Crossmatch to Gaia, 2MASS, WISE
  - Samples feature uncertainties
  - Account for various biases
  - Gives vector of probability that source belongs to each source class defined in training dataset (TD)
- Modified to use Chandra Source Catalog (CSC) 2.1, HST for classifying sources in GCs
- Probabilistic crossmatching using NWay (Salvato et al. 2018)
- Features: luminosities, absolute magnitudes, colors, CXO variability
  - GCs have known distances
- Use known extinction to GCs to deredden TD sources, and sources to be classified
- Classes: AB, AGN, CV, MSP, Spider-type MSP, LMXB
- Use upper limits to model cases where true counterpart is too faint to be detected
- To reduce bias, for each source to be classified, we implement following scheme:
  - If the source is missing a feature, remove the feature from pipeline
  - If the source is not missing the feature, remove TD sources that are missing that feature
  - Thus, classification is not influenced by what percent of TD sources of each class is missing feature

Training Dataset

- Compiled X-ray source positions and classifications from 50+ publications on GCs (Chen et al. 2023)
  - ~300 X-ray sources from 25 globular clusters
- Also compiled confident HST counterpart positions, magnitudes, when available in publications
- Crossmatched published X-ray coordinates to CSC
- Crossmatched published HST coordinates to HST UV Globular Cluster Survey (HUGS, Nardiello et al 2018), oMEGACat, and various CDF surveys
- Sources in these catalogs more uniformly processed than in literature, and astrometrically tied to Gaia
- ~150 confident crossmatches to HST
- 606 AGNs taken from Chandra Deep Fields, crossmatched to GOODS, CANDELS, HDUV surveys
- Summary of sources per class in all GCs and in Omega Cen only, with percent of sources with identified counterparts shown below

Class	Total	Counterpart %	Omega Cen	Counterpart %
MSP	62	24	7	0
CV	88	82	6	100
AB	111	100	1	100
qLMXB	37	30	1	100
AGN	621	57	6	100

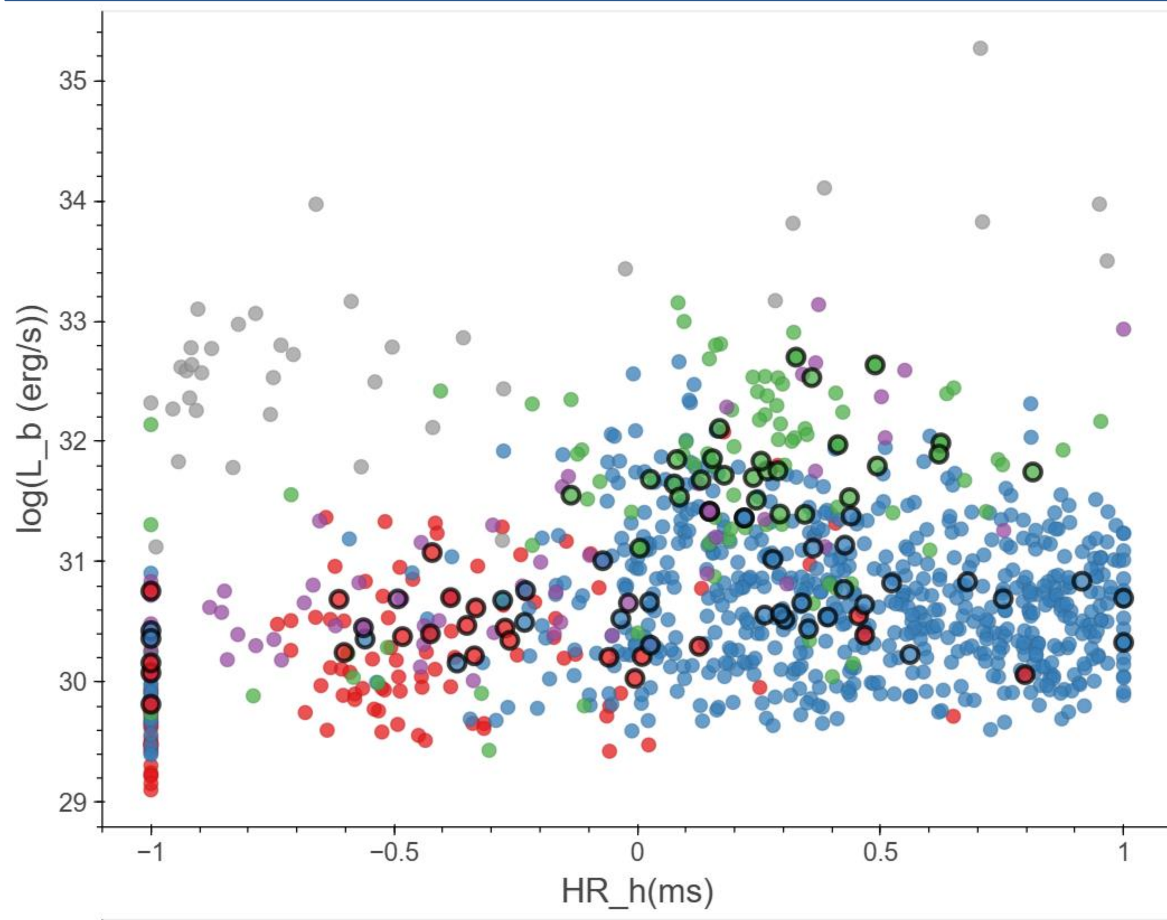
Leave-One-Out Cross-Validation



- Confusion matrices are shown for 10 runs, each run randomly sampling feature uncertainties.
- Most classes perform well, MSPs and CVs perform less well.
  - Matrix for confident classifications (>60% trees in all runs voting for class) is much more diagonal.
- Wrong classifications due to diversity within source class, lack of counterparts, similarity to other classes.

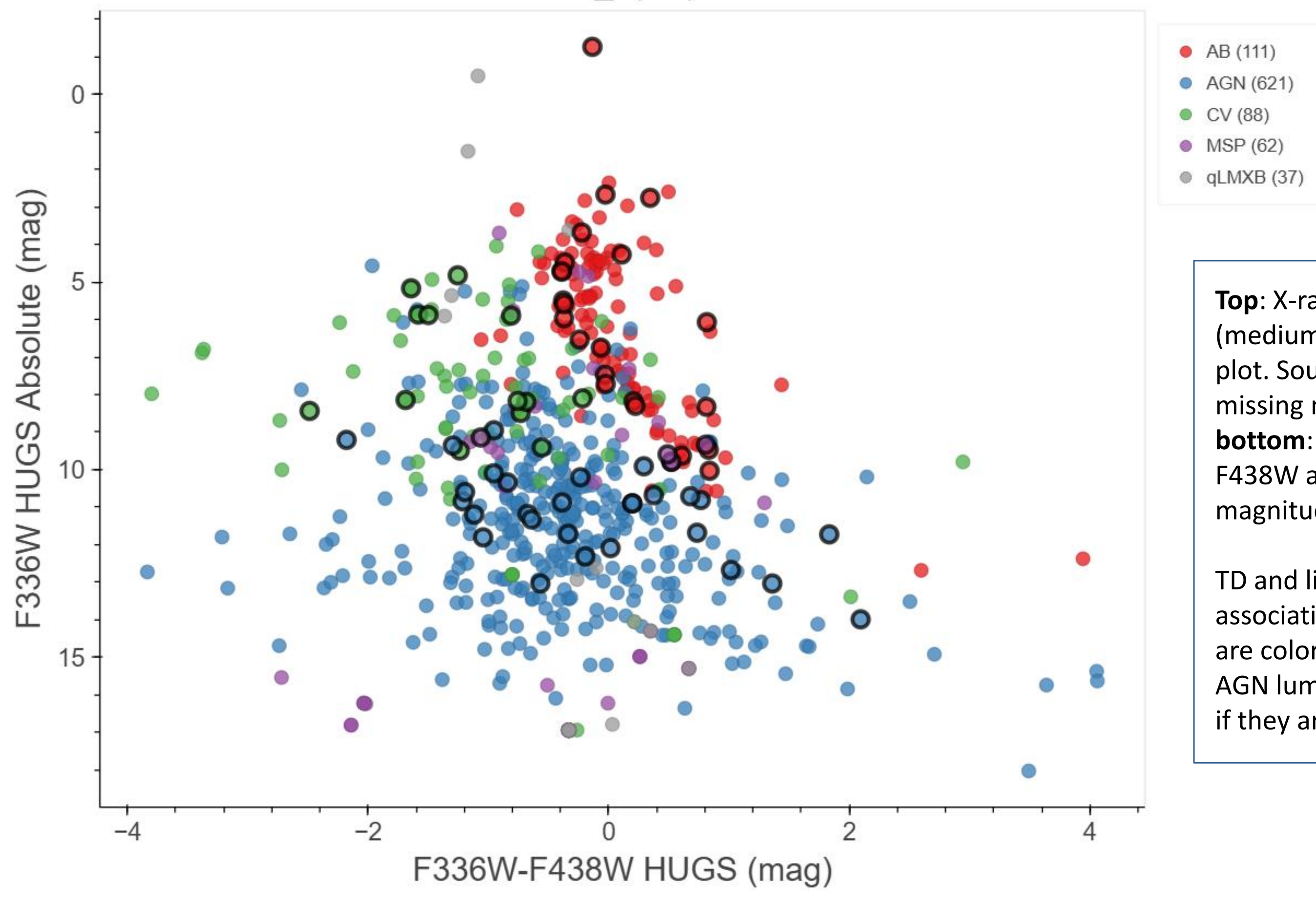
Classified Likely Associations

- Accurate classification heavily depends on finding true counterpart.
- No systematic astrometric offset between CSC and HST sources is found.
- CVs can give us a clue, as they are tightly grouped and distinct in X-ray and optical feature space:
  - More luminous and harder in X-rays compared to ABs and AGNs
  - Counterpart in white dwarf sequence, which is less common than main sequence
- We found 15 sources fulfilling these criteria:
  - Similar to TD CVs in X-rays, with significant probability to be classified as CV ($P_{CV} > 0.2$)
  - Has possible counterpart in white dwarf sequence, classified as CV
  - This counterpart is closest counterpart, often by wide margin ($<0.2''$ separation)
  - 6 are known CVs in Omega Cen, which are each removed from TD before being classified
- These are highly likely to be true CV associations, and we apply these criteria to find likely associations of other classes
- MSPs mostly lack counterpart in TD, classified MSPs largely based on X-ray information



Class	Closest CP	Likely
AGN	120	32
MSP	23	5
CV	29	26
AB	48	22
qLMXB	1	1

Classification summary of sources in ω Cen. Likely classifications found using procedure described above, and includes 6 CVs, 4 ABs, 4 AGNs, 1 MSP, and 1 qLMXB previously known.

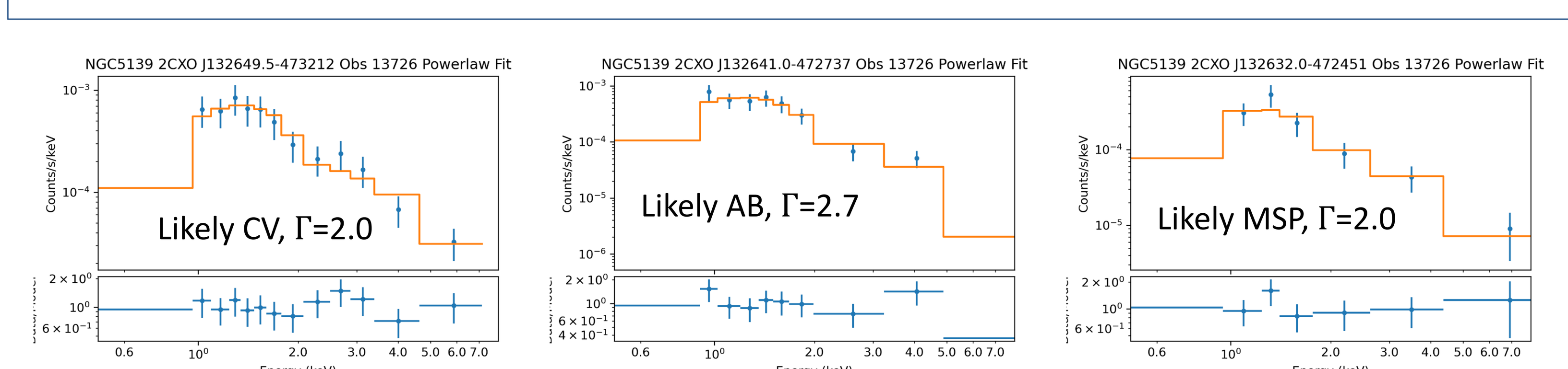


Top: X-ray Luminosity vs hard-(medium+soft) hardness ratio plot. Sources at HR_h(ms) = -1 missing medium or soft band. **bottom**: HST F336W vs F336W-F438W absolute color-magnitude diagram.

TD and likely classified ω Cen associations (outlined in black) are colored by class. AGN luminosities calculated as if they are at cluster distance.

Discussion

- Clear separation between the classes in the TD can be seen, while likely classifications have nearest counterparts that are also close in expected regions.
- 21 sources with known class in ω Cen can be used to check accuracy, when removed from TD
  - 14 sources are classified correctly using the correct association from publication
  - Most incorrect classifications are MSPs or spider MSPs
- Few MSPs pass criteria, while more are classified based on only X-ray information
  - Some with sufficient counts for spectral analysis, see plots below
  - Spectra consistent with MSPs, but other classes also possible
  - Large photon index ($\Gamma \gtrsim 3$) may imply substantial thermal emission contribution



Summary and Future Work

- MUWCLASS is a powerful tool for rapidly classifying many sources in different environments
  - Rapid classification of all associations enables finding metrics to identify true associations
  - Can substantially increase statistic for population studies of confidently classified source classes, e.g., flaring stars, AGNs, CVs
  - Identify unusual/interesting sources for more detailed study, e.g. MSP candidates
- Future work:
  - Classify X-ray sources in all applicable GCs
  - Integration of additional sensitive surveys, including radio
  - Expansion of TD: living database of classified X-ray sources

References

Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R.-D. 2013, Astronomy and Astrophysics, 558, A53
Zhao, J., & Heinke, C. O. 2022, Monthly Notices of the Royal Astronomical Society, 511, 5964
Bahramian, A., Strader, J., Miller-Jones, J. C. A., et al. 2020, ApJ, 901 (The American Astronomical Society), 57
Yang, H., Hare, J., Kargaltsev, O., et al. 2022, The Astrophysical Journal, 941, 104
Yang, H., Hare, J., & Kargaltsev, O. 2024, The Astrophysical Journal, 971, 180Y
Salvato, M., Buchner, J., Budavári, T., et al. 2018, Monthly Notices of the Royal Astronomical Society, 473, 4937
Chen, S., Kargaltsev, O., Yang, H., Hare, J., & Pavlov, G. 2023, Research Notes of the American Astronomical Society, 7, 215
Henleywillis, S., Cool, A. M., Haggard, D., et al. 2018, Monthly Notices of the Royal Astronomical Society, 479, 2834
Nardiello, D., Libralato, M., Piotto, G., et al. 2018, Monthly Notices of the Royal Astronomical Society, 481, 3382
Zhao, J., & Heinke, C. O. 2023, Monthly Notices of the Royal Astronomical Society, 526, 2736
Häberle, M., Neumayer, N., Bellini, A., et al. 2024, The Astrophysical Journal, 970 (IOP), 192

Acknowledgements: This research is supported by Space Telescope Science Institute award HST-AR-16620.