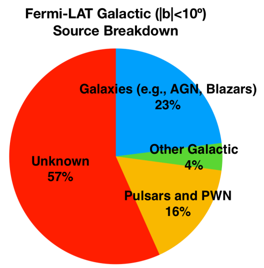
Automated classification of X-rays sources within the extent of Fermi-LAT sources

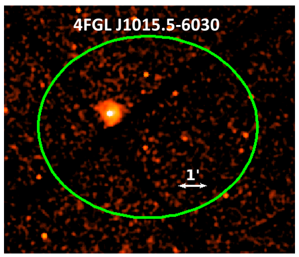
Jeremy Hare[1,*], Hui Yang[2], Oleg Kargaltsev[2], Igor Volkov[2]

1.- NASA/GSFC/USRA 2. - The George Washington University, *NASA Postdoctoral Program Fellow

Introduction

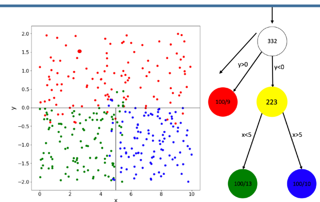
- Fermi-LAT has now been operating for >12 years and in this time is has uncovered many new Galactic GeV sources (e.g., pulsars, PWN, HMGBs)
- Many newly discovered GeV sources have uncertain classifications or remain entirely unidentified; ~1,000 sources discovered by the Fermi-LAT remain unidentified according to the 4FGL DR2 catalog (1).
- Many of these sources have existing X-ray (Chandra or XMM-Newton) coverage and virtually all fields will be covered by the ongoing eROSITA survey
- Additionally, new optical surveys such as ZTF and TESS are providing information about the optical variability of sources
- MW classification can provide an efficient way to learn about the nature of gamma-ray and X-ray sources and can also be applied to radio and X-ray SNR observations.





Machine Learning classification

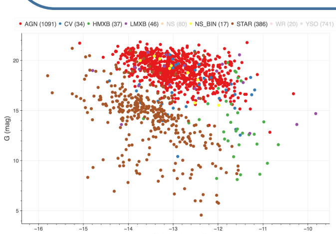
- We focus on supervised learning decision tree algorithms
- We use a Random Forest (RF) algorithm, which builds an ensemble of decision trees, each one different from the rest (2), and making it much less prone to overfitting.
- Our pipeline is currently written in python using several libraries including Scikit-learn, imbalance-learn, and astroquery (3,4,5,6)

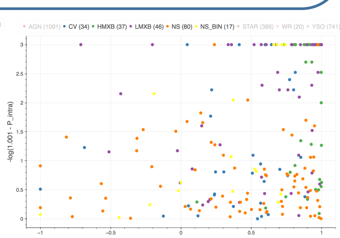


Chandra Training dataset

- The training data set is built and used to evaluate the parameters of objects from known classes and build a decision tree.
- The release of the Chandra source Catalog version 2 (CSCv2; 7,8) has enabled us to construct a new Chandra based training dataset consisting of roughly 2500 sources.
- Chandra's superior angular resolution allows for more accurate cross-matching to MW catalogs, and reduces the chance coincidence probability of spurious matches.
- In total the CSCv2 training dataset consists of 24 MW features, including X-ray time domain features (e.g. per observation and between observation variability).
- We have also updated the MW catalogs used to include Gaia DR3.
- Beta version of online plotting tool (used for figures below) can be found at https://home.gwu.edu/~kargaltsev/XCLASS/ (we would greatly appreciate any feedback/suggestions which can be sent to harej10@gmail.com)

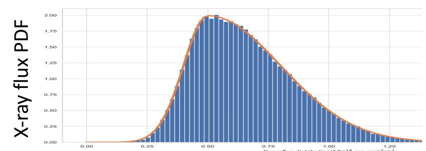
	AGN	PSR	PSR_BIN	CV	HMXB	LMXB	STAR	WR	YSO
Number	1091	80	17	34	37	46	386	20	741





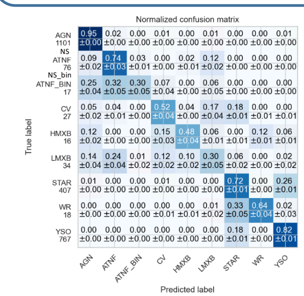
Accounting for uncertainties

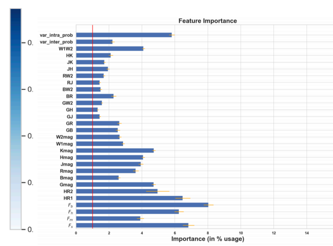
- We now account for uncertainties in the X-ray and MW features for each source
- This is accomplished by sampling each source's features within their uncertainty probability distribution and classifying all samples (i.e., a Monte-Carlo approach)
- Accounting for the uncertainties typically lowers the overall classification confidences but provides more realistic and reliable classification probabilities



Cross Validation

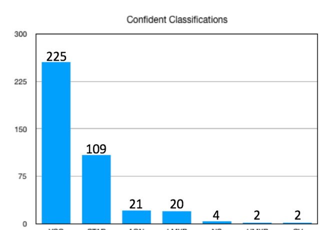
- We use leave-one-out cross validation to test our pipelines performance.
- Cross-validation also accounts for the uncertainties in source features.
- Overall pipeline accuracy is about 84%, up to 94% for ``confident'' classifications
- Performance dominated by most populated sources in training dataset (i.e., AGN, STARs, YSO), pipeline performs worse for underpopulated classes
- Can be used to narrow down the interesting sources for further follow-up.

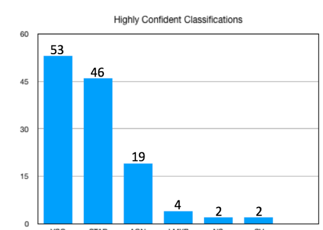




Chandra observations of Unidentified Fermi-LAT sources

- About 50 fields of unidentified Galactic (|b|<10°) Fermi-LAT sources have been covered by Chandra over its lifetime, of which about 30 are currently contained in the CSCv2.
- We selected all X-ray sources that lie within the Fermi-LAT positional uncertainties to run through our classification pipeline.
- We then cross-matched these sources with MW catalogs to extract all 24 features used by our pipeline.
- Classification without accounting for uncertainties takes only a few minutes on a modern laptop. However, using the Monte-Carlo approach to account for uncertainties is very computationally expensive, requiring about 30 s per simulation per core or about 11 hours on a 40 core machine.
- Overall, there were about 1900 X-ray sources located within the 95% positional uncertainties of unidentified Galactic Fermi-LAT sources, of which about 400 (or 20%) were confidently classified and about ~10% were highly confidently classified.



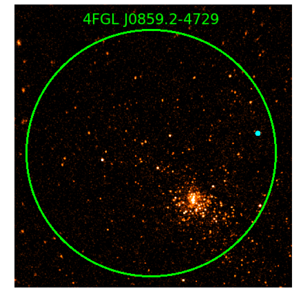


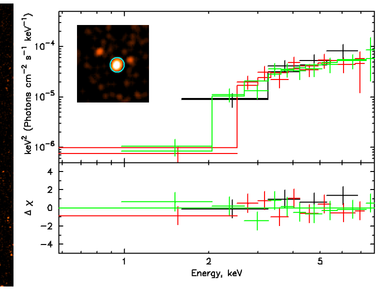
Acknowledgements

JH acknowledges support from an appointment to the NASA Postdoctoral Program at the Goddard Space Flight Center, administered by the Universities Space Research Association under contract with NASA. Support for this work was provided by the NASA through Chandra Awards AR3-14017X and AR0-21007X and by NASA Astrophysics Data Analysis Program award 80NSSC19K0576.

4FGL J0859.2-4729

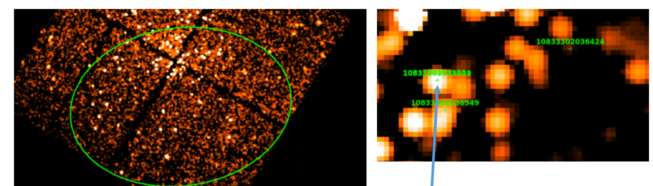
- Unidentified Fermi-LAT source that overlaps with the young star cluster RCW 38.
- May be multiple sources contributing to GeV emission
- High classification confidence NS candidate classified in this field (shown as teal source in Figure below).
- Source has no optical/IR/NIR counterparts but field is heavily absorbed.
- Source's X-ray spectrum appears to be non-thermal and is well described by an absorbed power-law model with a photon-index of 1.6.
- Large absorption may be reason for lack of detection of MW counterpart, deeper limits needed to search for counterpart.

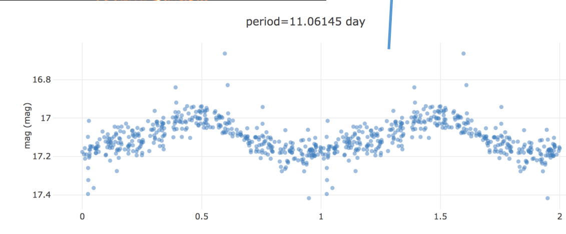




Ongoing and Future Work

- Once we have finished the analysis of fields in the CSCv2, we will begin analyzing and running our pipeline on fields that have not yet been added to the CSC. This will add an additional ~20 source fields.
- We are anticipating adding data from additional frequency bands, such as radio fluxes from the VLASS (or MeerKAT) surveys upon release.
- We are also working on adding additional time domain features from the many all sky optical surveys currently taking place (e.g., ZTF, ASAS-SN, TESS). This can help to provide additional variability information (e.g., periodicity, flaring) for sources too faint to measure their variability in X-rays.
- Below is an example of an 11 day modulation in the ZTF light curve of an optical counterpart to an X-ray source in the field of an unidentified Fermi-LAT source.





https://irsa.ipac.caltech.edu/cgi-bin/g2p/nph-g2p

References

1) Abdollahi, S., Acero, F., Ackermann, M., et al. 2020, ApJS, 247, 33
2) Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.
3) Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011
4) Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, AAP, 558, A33
5) Astropy Collaboration, Price-Whelan, A. M., SipHocz, B. M., et al. 2018, AJ, 156, 123
6) Lemaitre, G., Nogueira, F., & Aridas, C. K. 2016, arXiv:1609.06570
7) Evans, I.~N., Allen, C., Anderson, C.~S., et al. 2019, AAS/High Energy Astrophysics Division
8) Evans, I.~N., Primini, F.~A., Miller, J. B., et al. 2020, American Astronomical Society Meeting Abstracts #235