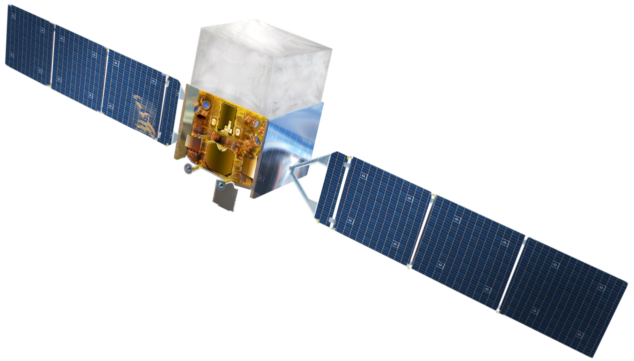


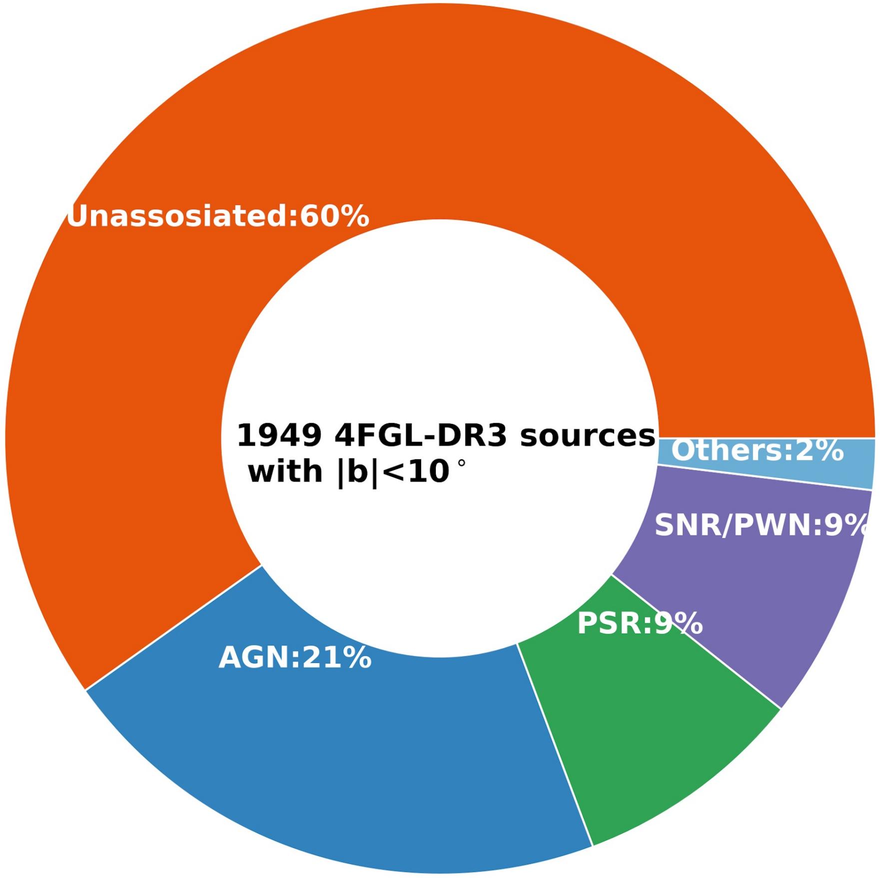
Classification of 4FGL sources with CSCv2 and multi-wavelength surveys

Hui Yang[1], Jeremy Hare[2], Oleg Kargaltsev[1], Steven Chen[1], Igor Volkov[1], Blagoy Rangelov[3]

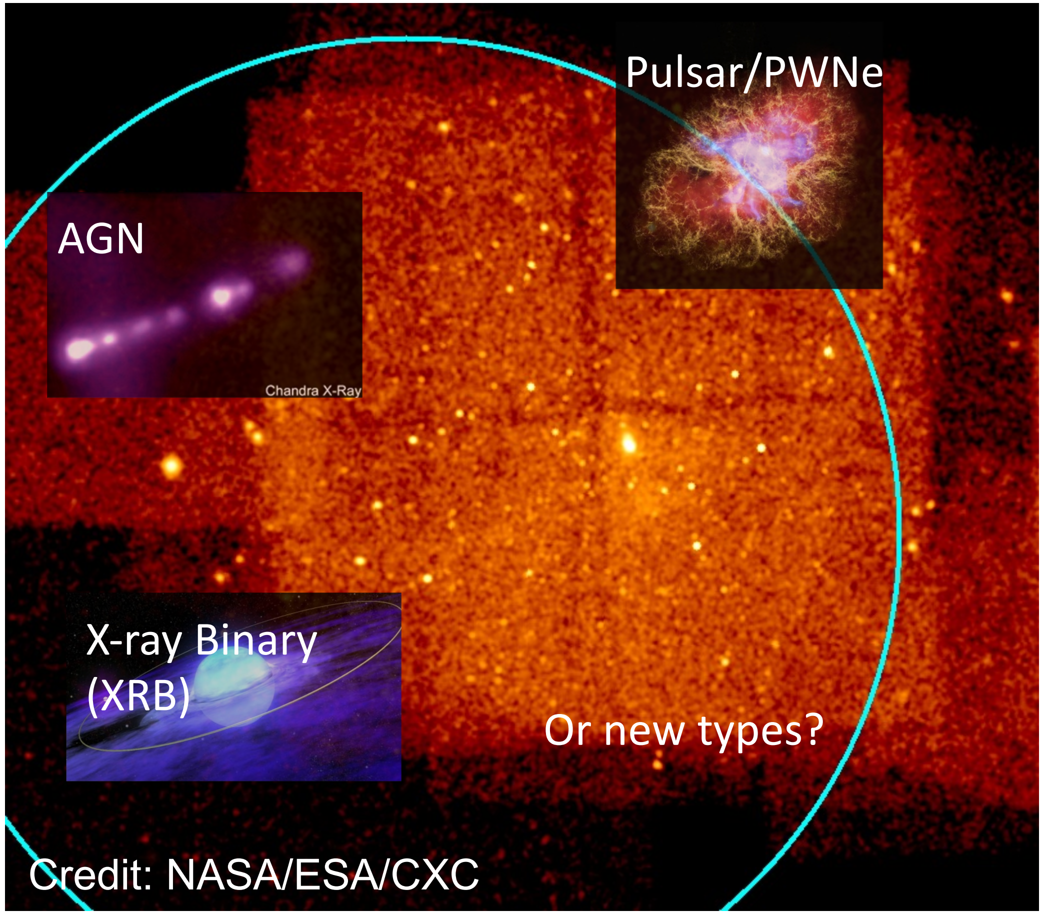
[1] The George Washington University ,[2] NASA GSFC,[3] Texas State University

Fermi Summer School, Lewes Delaware, 2023

Unassociated Galactic 4FGL-DR3 sources - identifying X-ray sources of γ-ray emitters



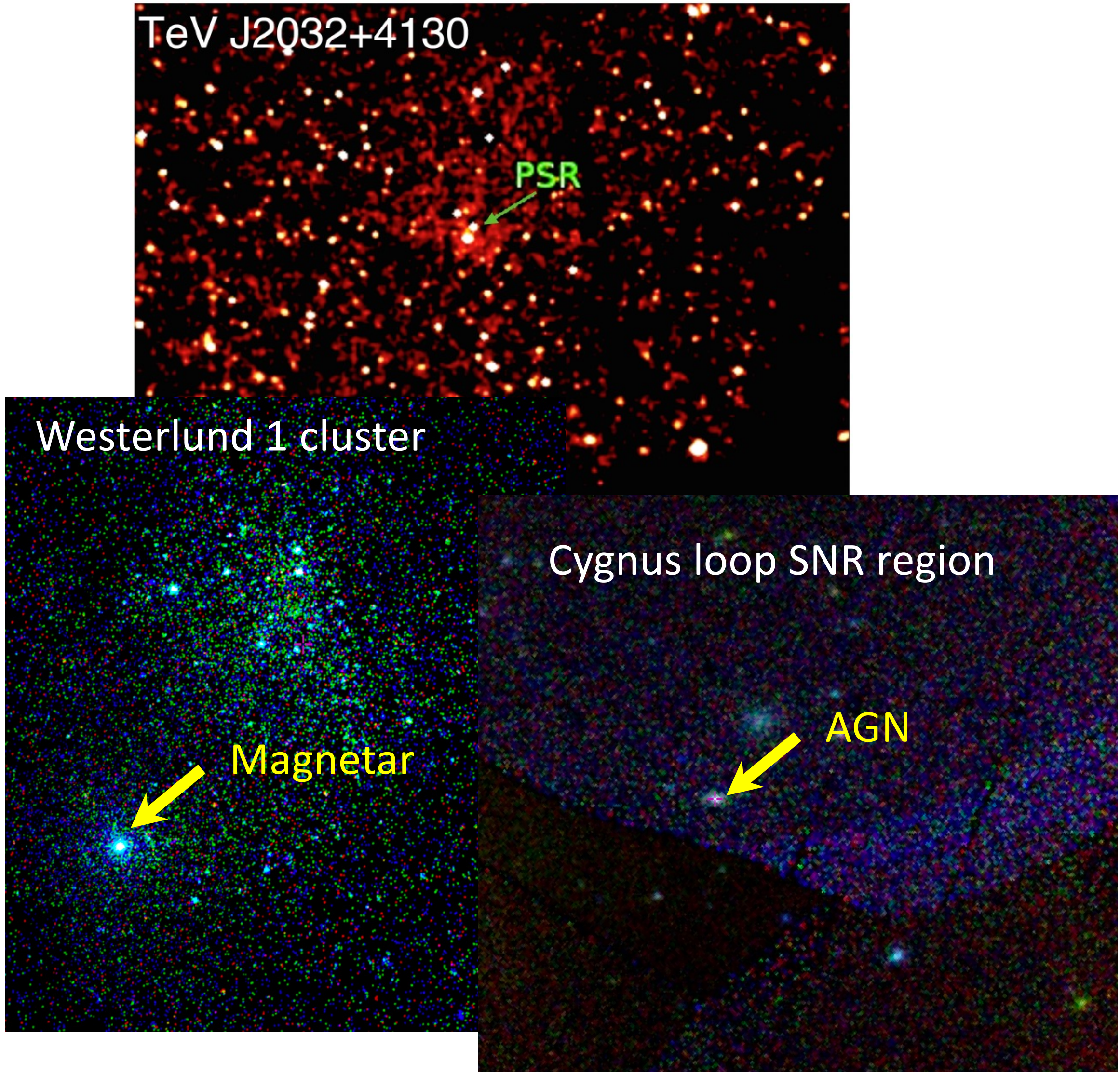
Right: Breakdown of 4FGL-DR3 classifications for Galactic plane sources within |b|<10



Chandra image of γ-ray sources HESS J1809-1917 overlaid with different types of γ-ray emitters

Many PSR, AGN, XRBs may look similar

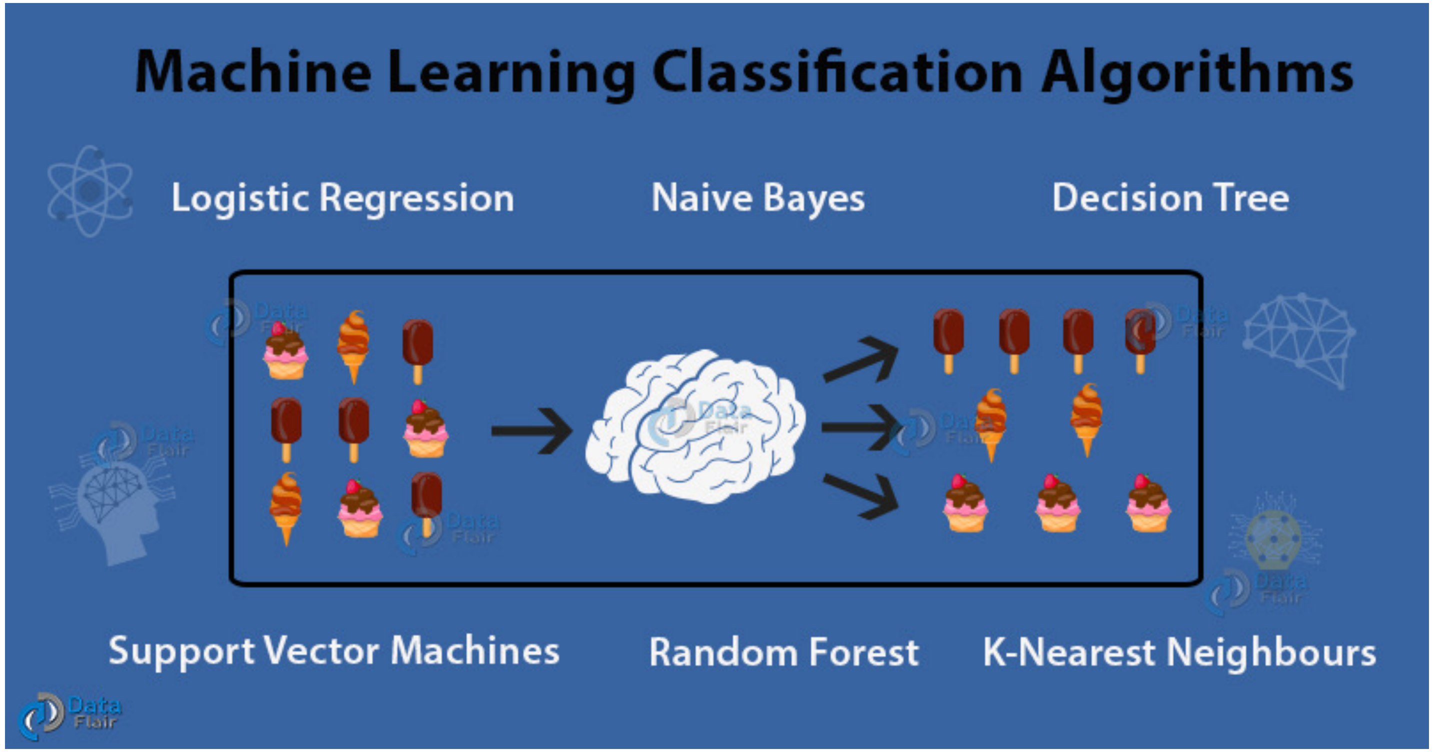
Majority of X-ray sources are faint, and difficult to distinguish from each other.



Credit: O. Kargaltsev

Solution: Machine Learning

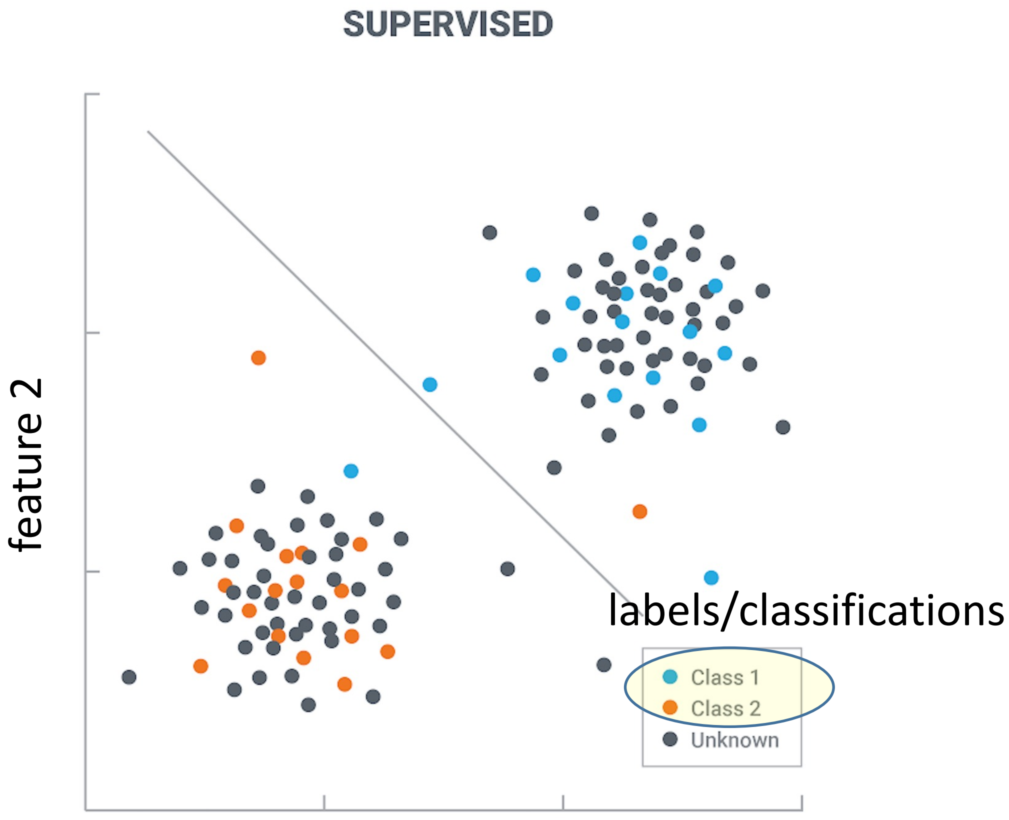
ML is needed to enable efficient classification and to reveal dependencies in large datasets **with high dimensionality.**

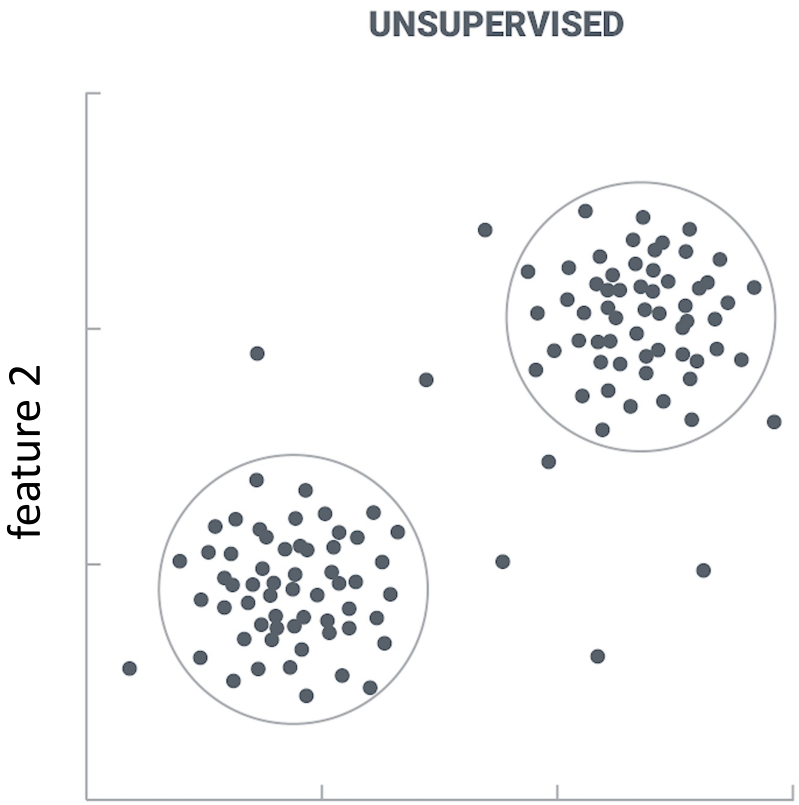


https://data-flair.training/blogs/machine-learning-classification-algorithms/

Machine Learning Basics

- Supervised Learning vs. Unsupervised Learning (Clustering)





feature 1 ⟺ parameters, measurements

feature 1

Credit: LAWTOMATED

The Astrophysical Journal, 941:104 (35pp), 2022 December 20 https://doi.org/10.3847/1538-4357/ac952b

Classifying Unidentified X-Ray Sources in the Chandra Source Catalog Using a Multiwavelength Machine-learning Approach

Hui Yang[1], Jeremy Hare[2,3], Oleg Kargaltsev[1], Igor Volkov[1], Steven Chen[1], and Blagoy Rangelov[4]
[1] Department of Physics, The George Washington University, 725 21st Street NW, Washington, DC 20052, USA; huiyang@gwmail.gwu.edu
[2] NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[3] NASA Postdoctoral Program Fellow
[4] Department of Physics, Texas State University, 601 University Drive, San Marcos, TX 78666, USA
Received 2022 June 27; revised 2022 September 12; accepted 2022 September 19; published 2022 December 15

Abstract

The rapid increase in serendipitous X-ray source detections requires the development of novel approaches to efficiently explore the nature of X-ray sources. If even a fraction of these sources could be reliably classified, it would enable population studies for various astrophysical source types on a much larger scale than currently possible. Classification of large numbers of sources from multiple classes characterized by multiple properties (features) must be done automatically and supervised machine learning (ML) seems to provide the only feasible approach. We perform classification of Chandra Source Catalog version 2.0 (CSCv2) sources to explore the potential of the ML approach and identify various biases, limitations, and bottlenecks that present themselves in these kinds of studies. We establish the framework and present a flexible and expandable Python pipeline, which can be used and improved by others. We also release the training data set of 2941 X-ray sources with confidently established classes. In addition to providing probabilistic classifications of 66,369 CSCv2 sources (21% of the entire CSCv2 catalog), we perform several narrower-focused case studies (high-mass X-ray binary candidates and X-ray sources within the extent of the H.E.S.S. TeV sources) to demonstrate some possible applications of our ML approach. We also discuss future possible modifications of the presented pipeline, which are expected to lead to substantial improvements in classification confidences.

Unified Astronomy Thesaurus concepts: Catalogs (205); X-ray sources (1822); Classification (1907); Random Forests (1935); X-ray binary stars (1811); Active galactic nuclei (16); X-ray stars (1823); Young stellar objects (1834); Cataclysmic variable stars (203); Astrostatistics tools (1887); X-ray surveys (1824); Compact objects (288)

Supporting material: machine-readable tables

Yang, H., Hare, J., Kargaltsev, O., et al. 2022, ApJ, 941, 104

The MUltiWavelength CLASSification Pipeline (MUWCLASS): Training Dataset (TD)

Supervised Machine Learning (ML) approach requires training dataset (TD)

Source Type	Number of CSCv2 sources
Active galactic nucleus (AGN)	1390
Cataclysmic variable (CV)	44
High-mass star (HM-STAR)	118
High-mass X-ray binary (HMXB)	26
Low-mass star (LM-STAR)	207
Low-mass X-ray binary (LMXB)*	65
Pulsar and isolated neutron star (NS)	87
Young stellar object (YSO)	1004
Total 8 source classes	2941

*LMXB also includes non-accreting X-ray binaries (e.g., red-back and black widow systems)

MUWCLASS: (29) Features/Properties

Various Colors

Mid-Infrared from WISE

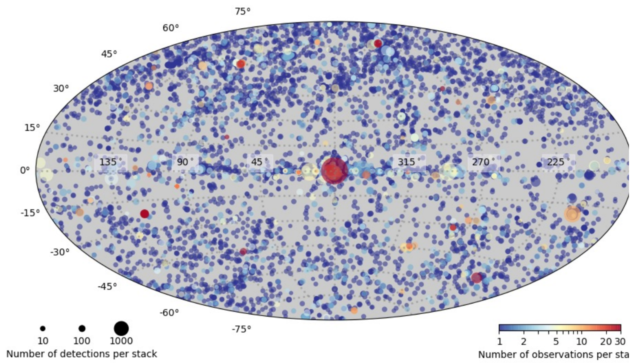
Near-Infrared from 2MASS

Optical from Gaia

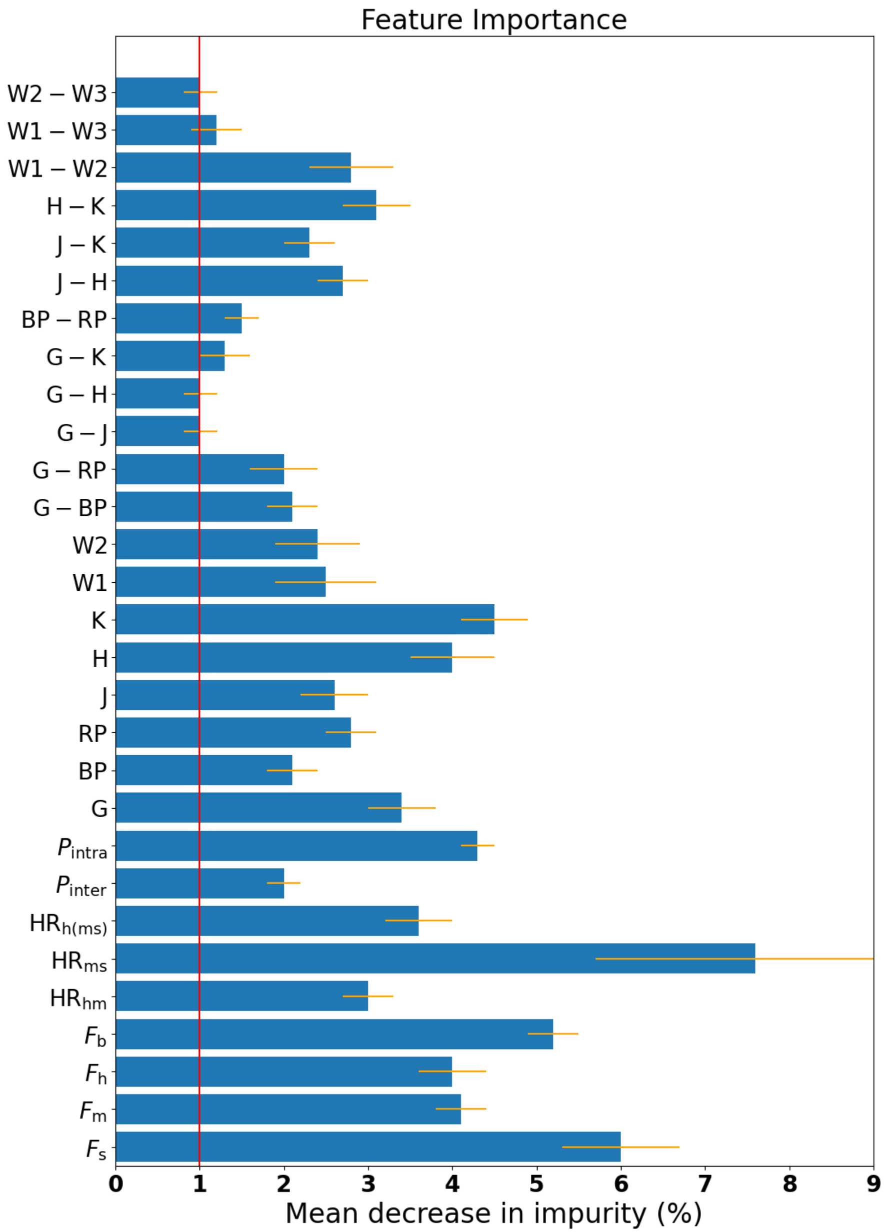
X-ray variability

Hardness ratios

Chandra X-ray fluxes



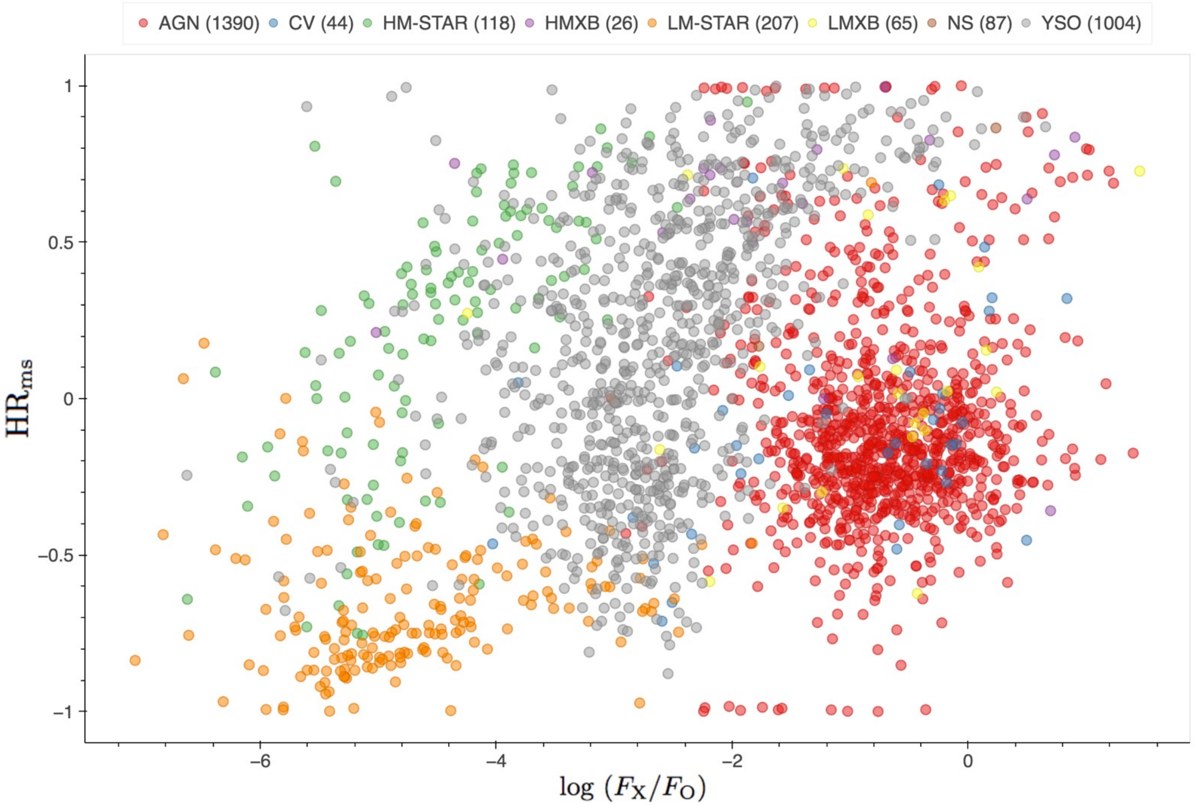
CSC v2.0 Detection Map from CXO

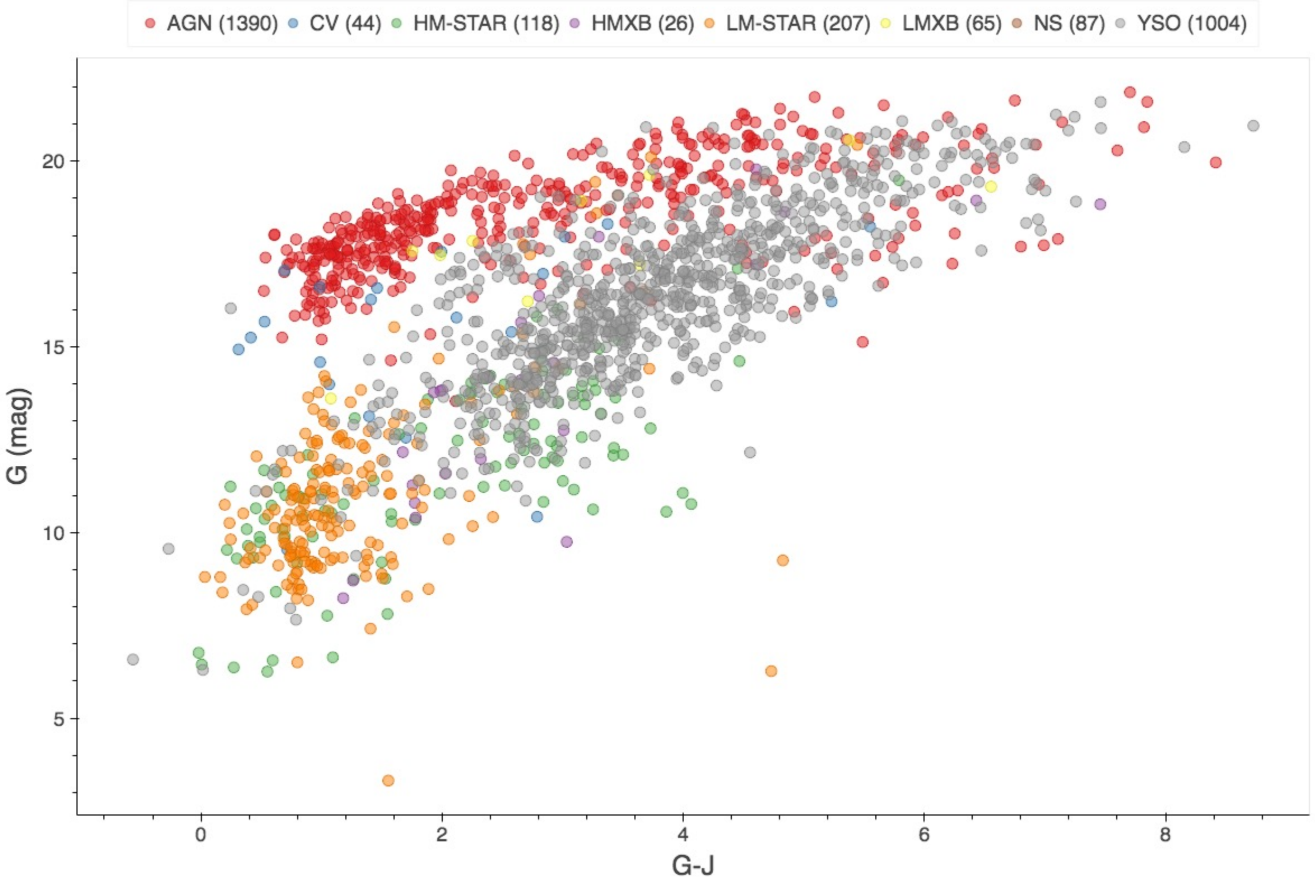


2D representations of our TD

SCAN ME!

Typically traditional classification consists
of using multi-wavelength parameter plots to separate
source classes

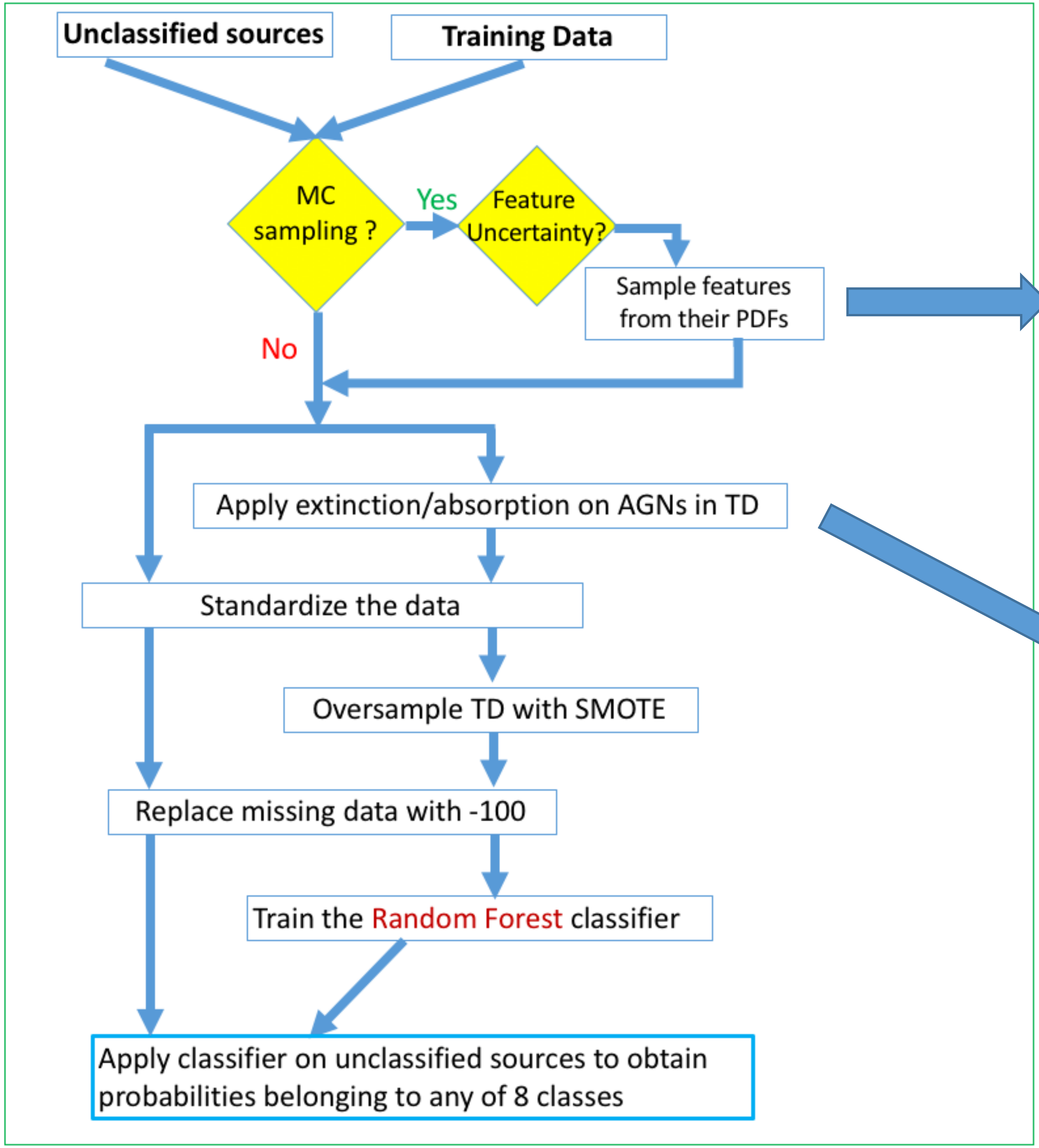


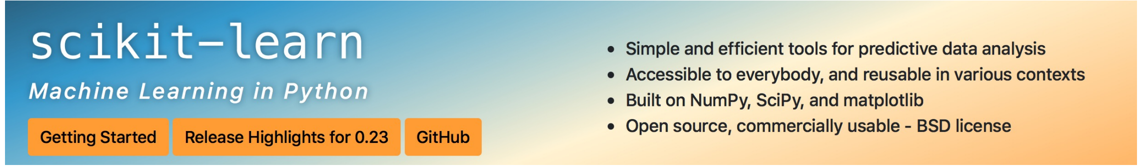


Hard to comprehend more than 2 or 3 dimensions for human

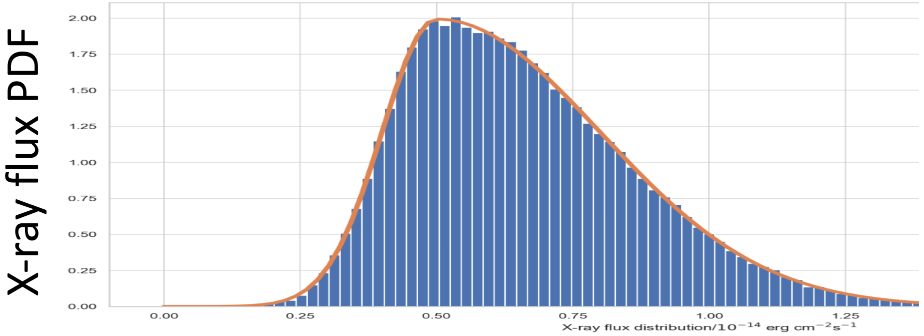
Explore the TD yourself using the visualization GUI at
https://home.gwu.edu/~kargaltsev/XCLASS/ (Yang et al. 2021)

MUWCLASS pipeline flow chart

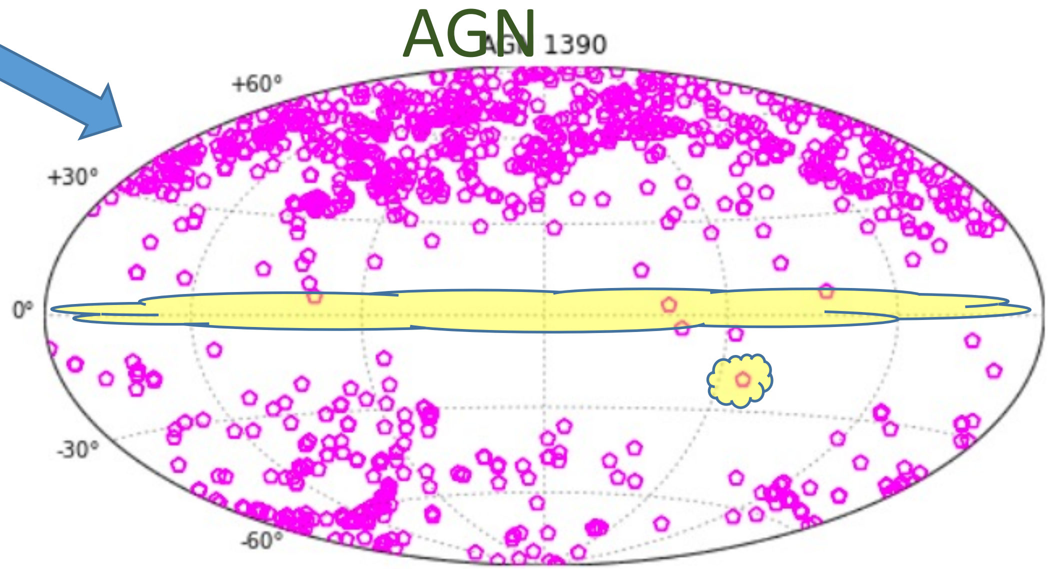




https://scikit-learn.org/stable/



Feature measurement uncertainties are taken into account by sampling from feature PDFs via Monte-Carlo.



Confusion Matrix (true vs. predicted class)

recall CM

True Class \ Predicted Class	AGN	CV	HM-STAR	LM-STAR	HMXB	LMXB	NS	YSO
AGN 1390	93	0	0	0	0	3	2	2
CV 44	9	55	2	2	5	16	5	7
HM-STAR 118	0	0	72	11	0	0	3	14
LM-STAR 207	0	0	6	86	0	0	0	7
HMXB 26	0	8	12	0	69	0	0	12
LMXB 65	11	20	2	2	2	26	34	5
NS 87	1	2	0	0	0	26	68	2
YSO 1004	1	0	2	2	0	0	0	94

Average Accuracy: 88.6%

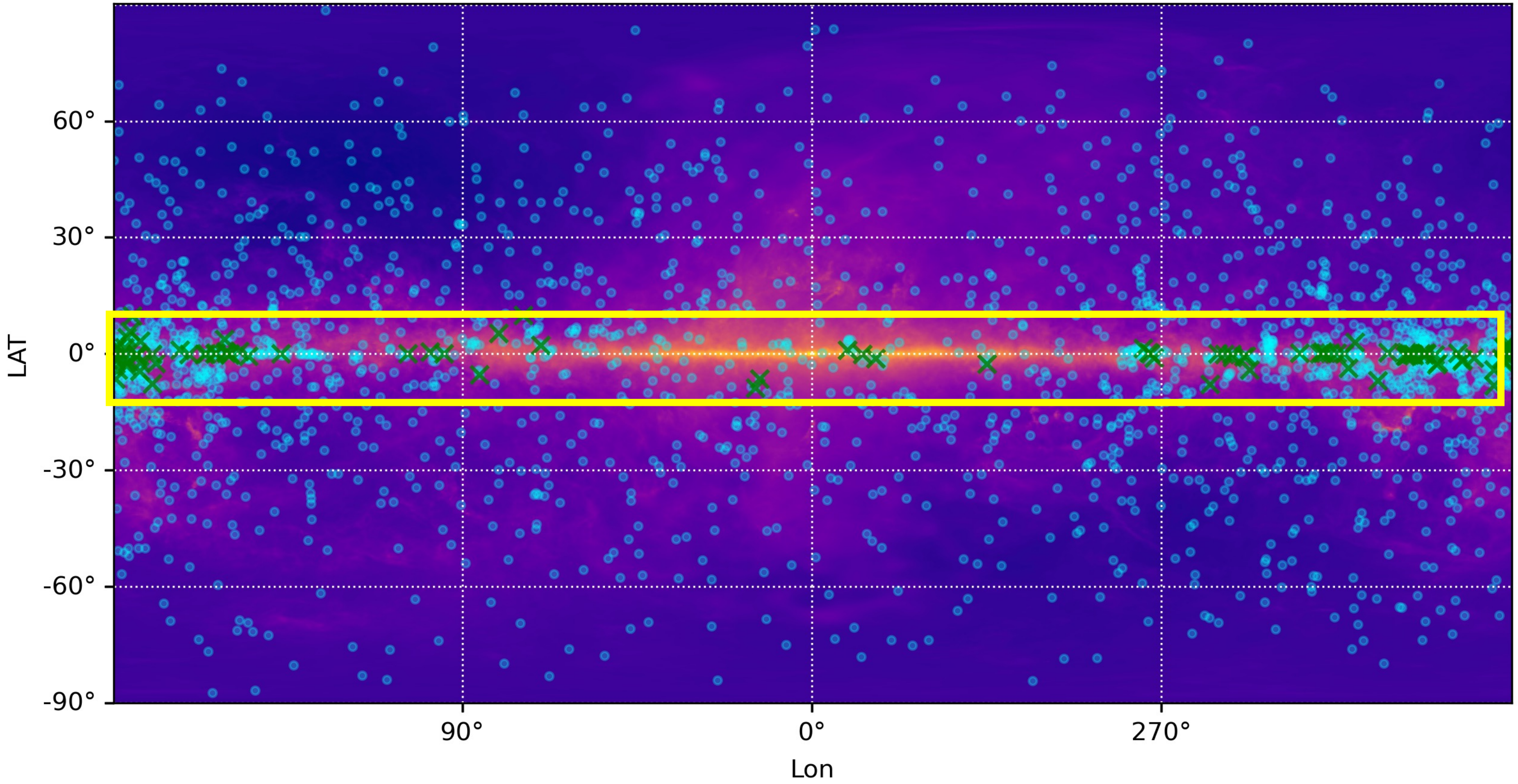
Confusion Matrix of **Confident** Classifications

confident recall CM

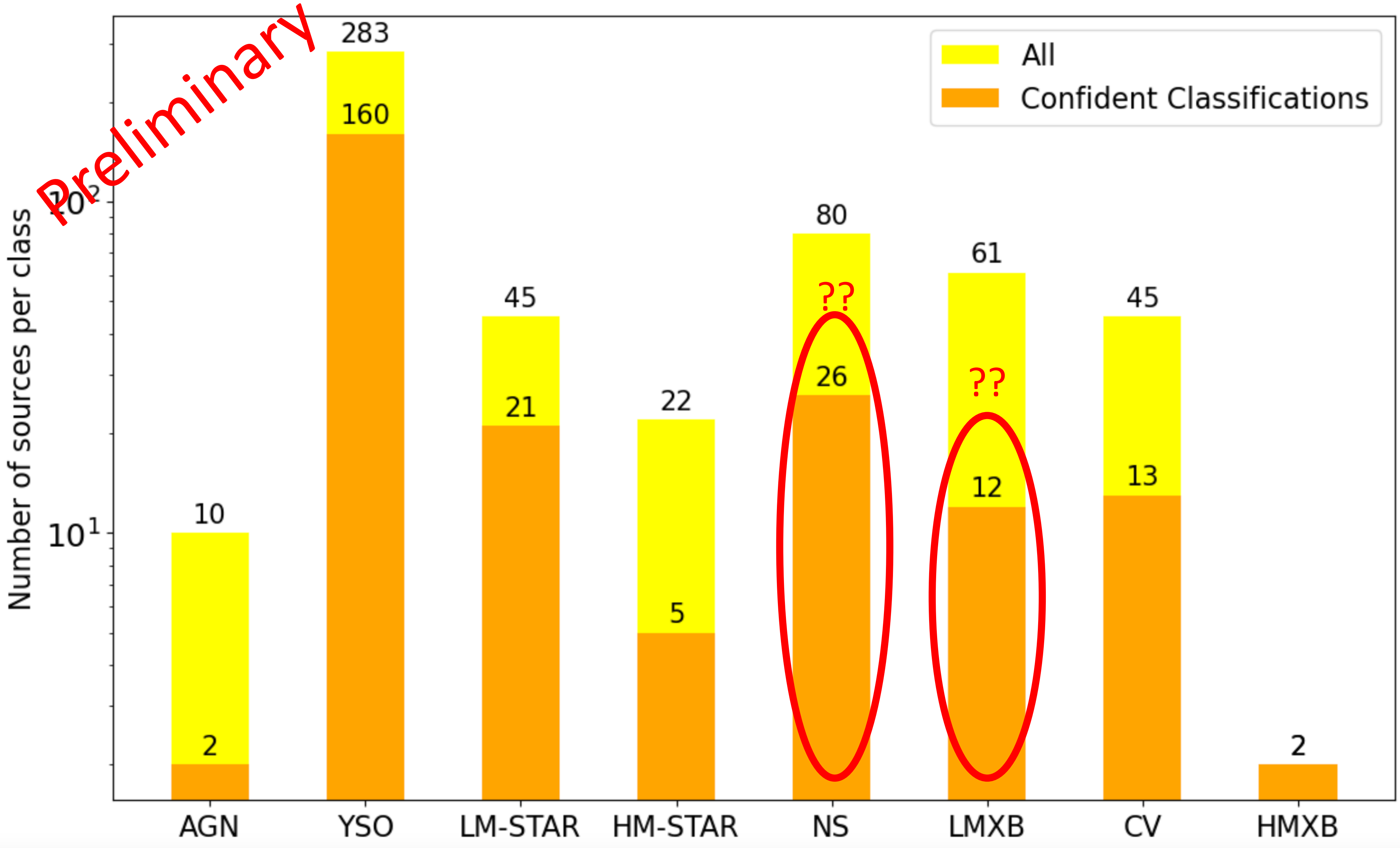
True Class \ Predicted Class	AGN	CV	HM-STAR	LM-STAR	HMXB	LMXB	NS	YSO
AGN 0.87	99	0	0	0	0	0	0	0
CV 0.34	0	73	0	0	7	7	0	13
HM-STAR 0.53	0	0	79	8	0	0	0	13
LM-STAR 0.70	1	0	1	96	0	1	0	2
HMXB 0.62	0	6	12	0	75	0	0	6
LMXB 0.18	25	17	0	0	8	25	17	8
NS 0.43	0	0	0	0	0	11	86	3
YSO 0.88	0	0	1	1	0	0	0	98

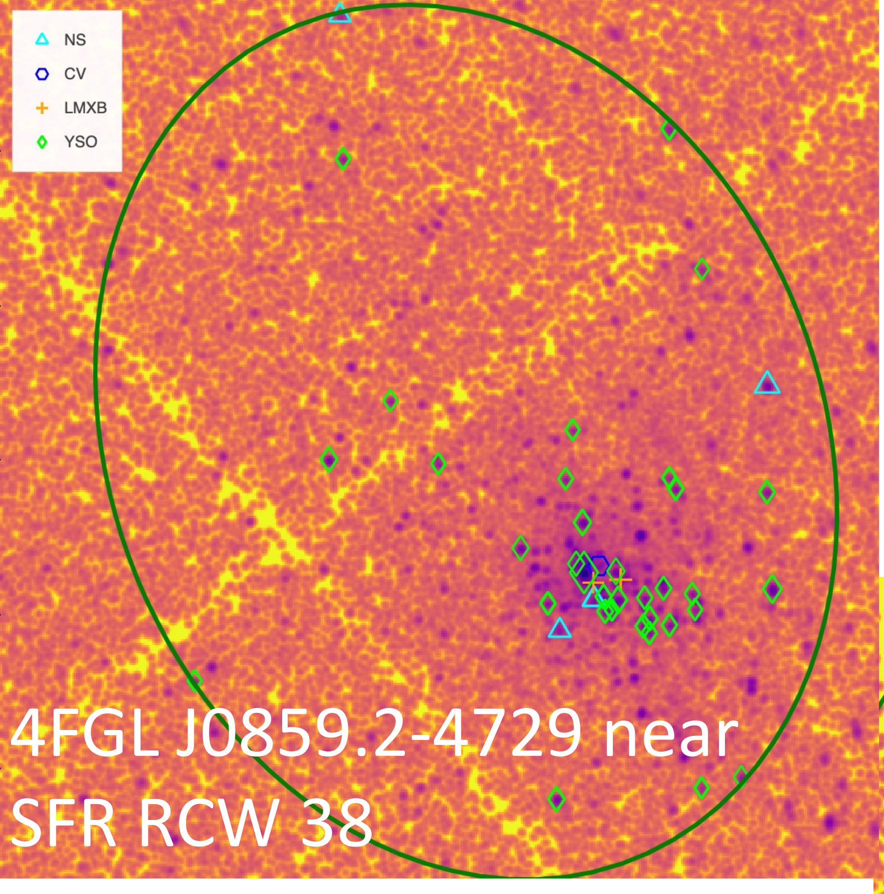
Average Accuracy: 97%

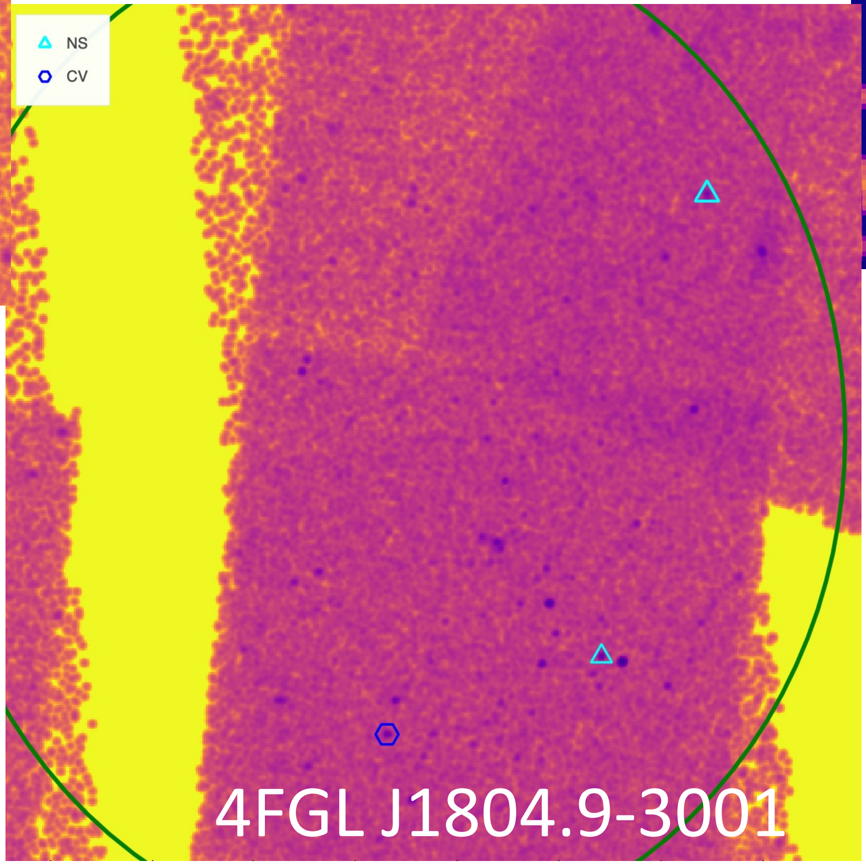
Classification of CSCv2 Sources within unassociated 4FGL-DR3 fields

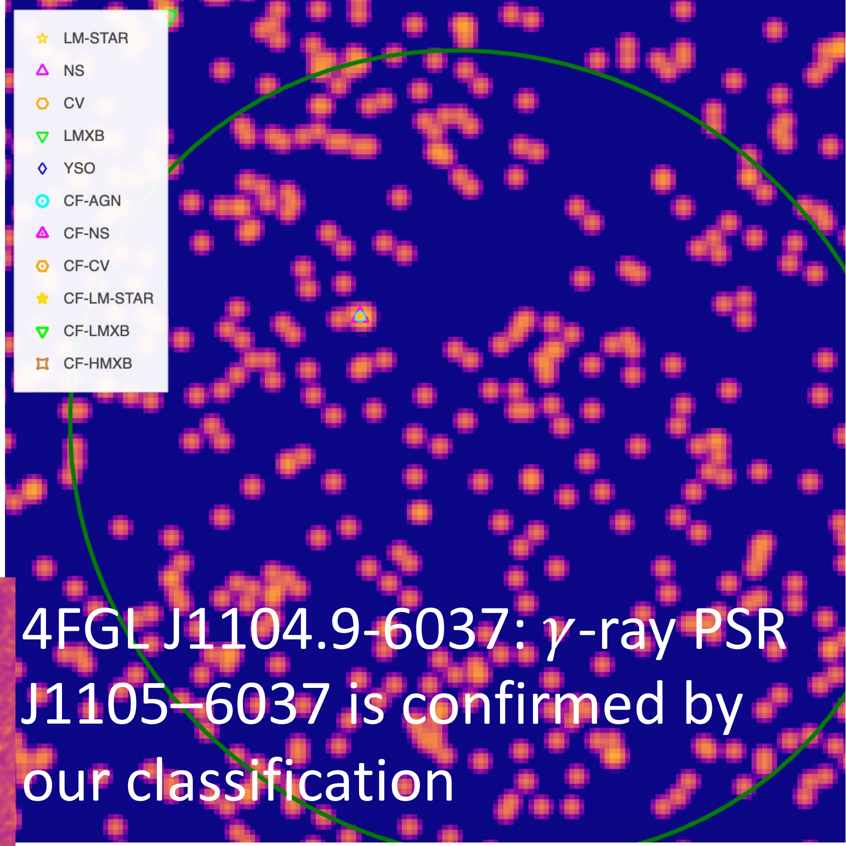


- 37 unassociated 4FGL-DR3 sources within |b|<10° with ≥5 CSCv2 sources within their error ellipses;
- 548 significant (X-ray S/N>5) CSCv2 sources within these 37 FGL sources.


Preliminary
All
Confident Classifications
Number of sources per class
10^2
10^1
283
160
10
2
45
21
22
5
80
26
??
61
12
??
45
13
2
AGN
YSO
LM-STAR
HM-STAR
NS
LMXB
CV
HMXB


NS
CV
LMXB
YSO
4FGL J0859.2-4729 near
SFR RCW 38


NS
CV
4FGL J1804.9-3001

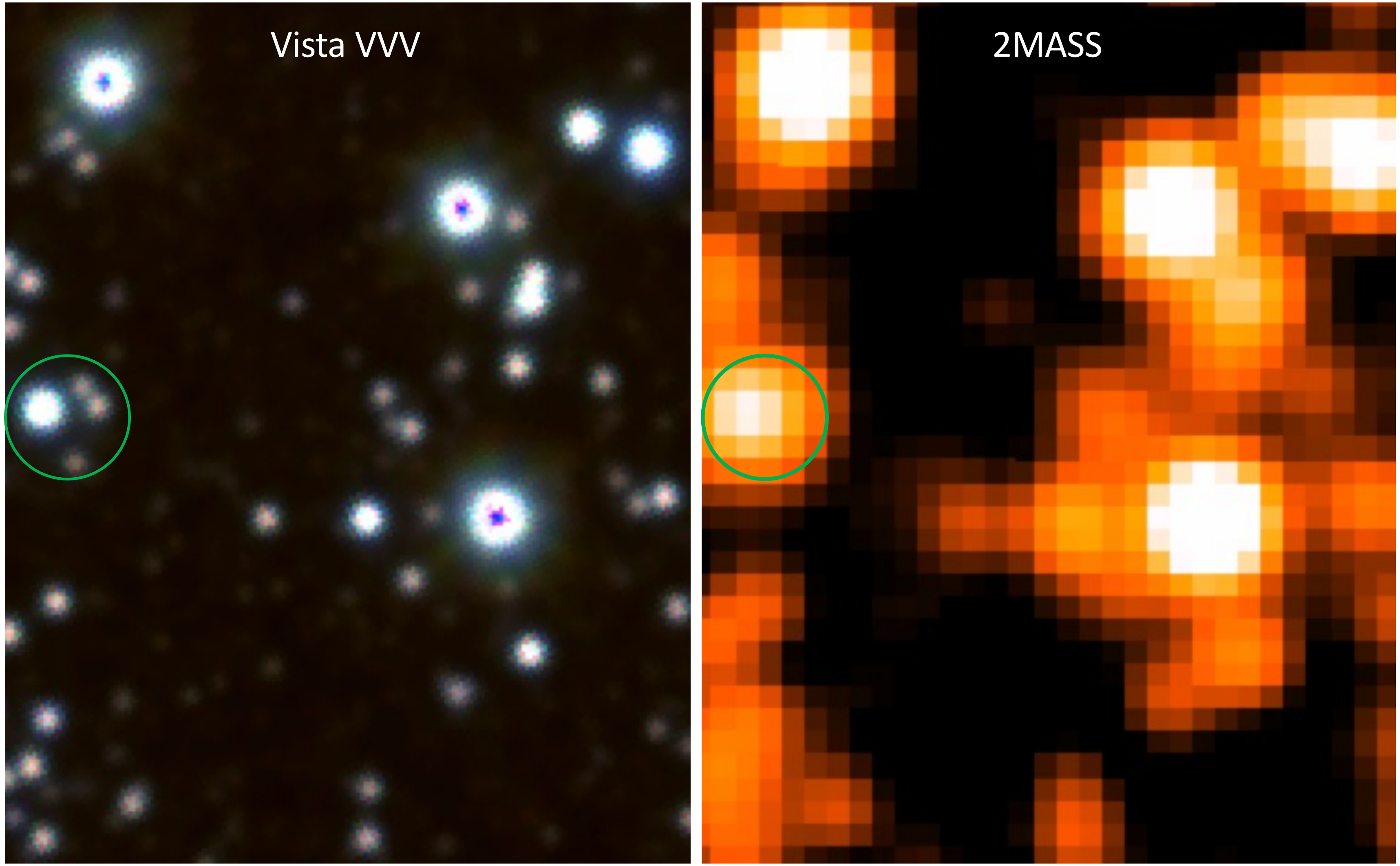

LM-STAR
NS
CV
LMXB
YSO
CF-AGN
CF-NS
CF-CV
CF-LM-STAR
CF-LMXB
CF-HMXB
4FGL J1104.9-6037: γ-ray PSR
J1105–6037 is confirmed by
our classification

Issues and Biases

- NS virtually all too faint to be detected by multi-wavelength surveys used in training dataset
- Many LMXBs also have counterparts too faint to be detected by multi-wavelength surveys used in training dataset, hence they become confused with the NS class
- Sources too faint to be detected by these MW surveys (e.g., M-dwarfs, absorbed AGN) will be preferentially classified as NS/LMXBs

Future Improvements: deeper surveys

- Update to more sensitive surveys (e.g., Pan-STARRs, DECaps, Vista VVV)

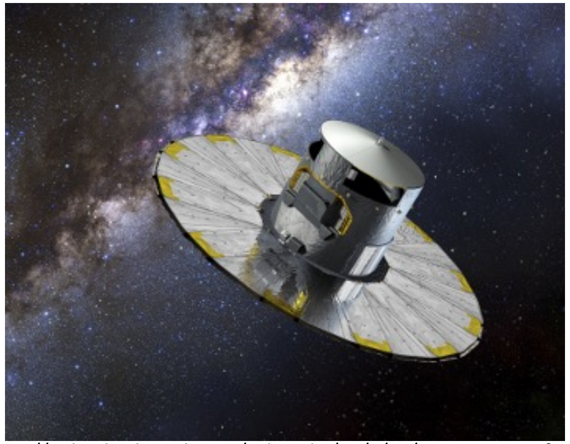


Future Improvements: New Multi-wavelength Features

ASKAP

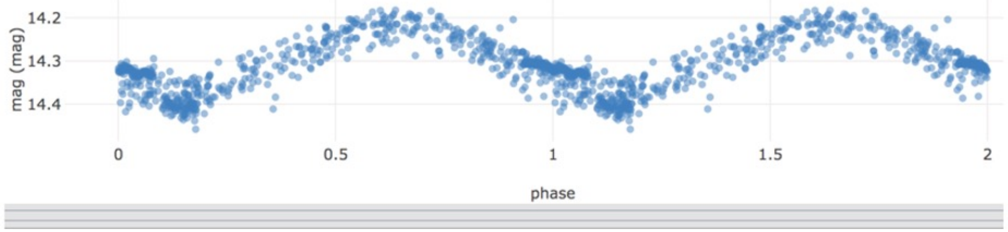
- Inclusion of new radio surveys:
- Australian SKA Pathfinder Telescope (ASKAP)
- VLA All-sky Survey (VLASS)
- MeerKAT source catalog

Gaia



https://upload.wikimedia.org/wikipedia/en/0/01/Gaia_spacecraft.jpg

- Distances and proper motions from Gaia DR3



ZTF

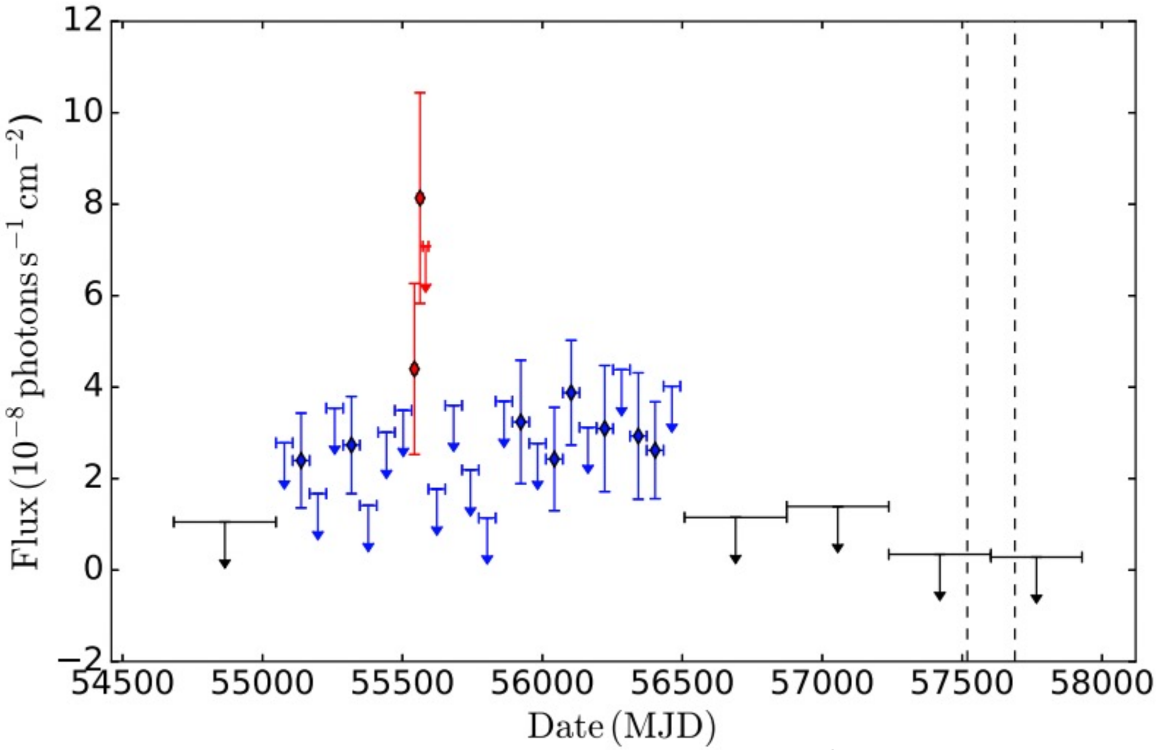
https://www.ipac.caltech.edu/project/ztf

- Large field of view optical time domain surveys:
- ZTF
- TESS
- VCRO

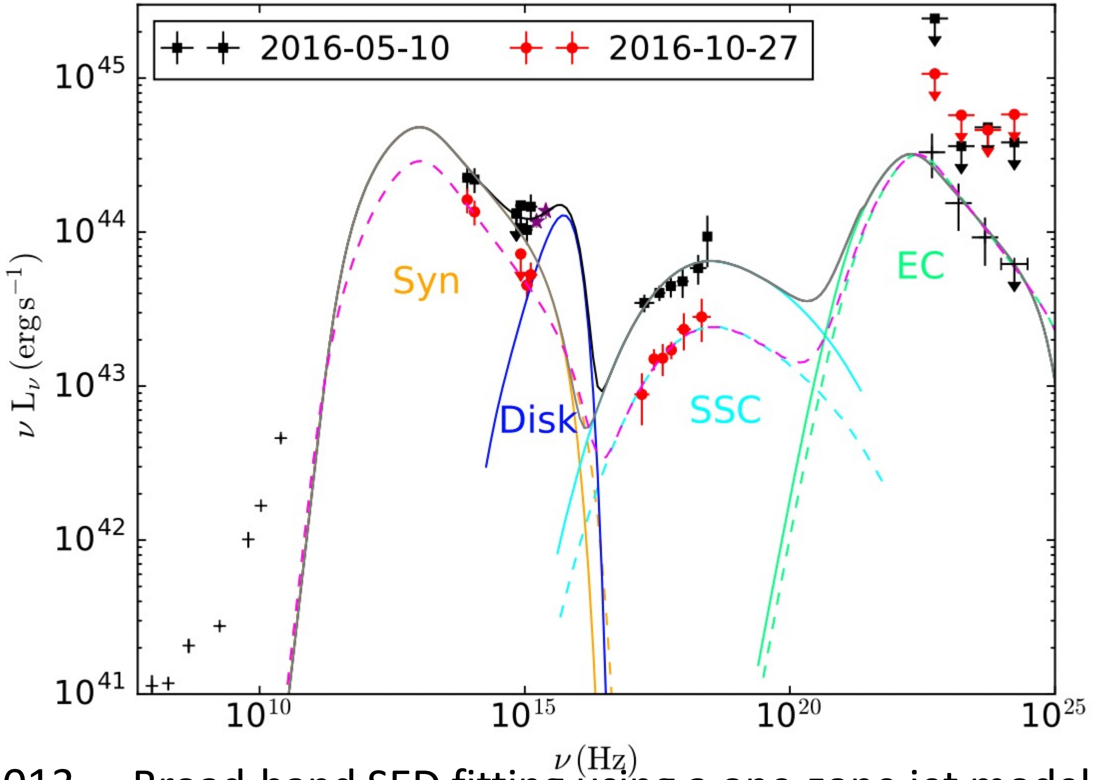
MNRAS 477, 5127–5138 (2018)
Advance Access publication 2018 April 16
doi:10.1093/mnras/sty904

SDSS J211852.96−073227.5: a new γ-ray flaring narrow-line Seyfert 1 galaxy

Hui Yang,[1,2★] Weimin Yuan,[1,2★] Su Yao,[3,4] Ye Li,[3,4] Jin Zhang,[1] Hongyan Zhou,[5,6] S. Komossa,[7] He-Yang Liu[1,2] and Chichuan Jin[1]



Fermi-LAT light curve shows monthly γ-ray flaring activities during 2009-2013.



Broad-band SED fitting using a one-zone jet model.