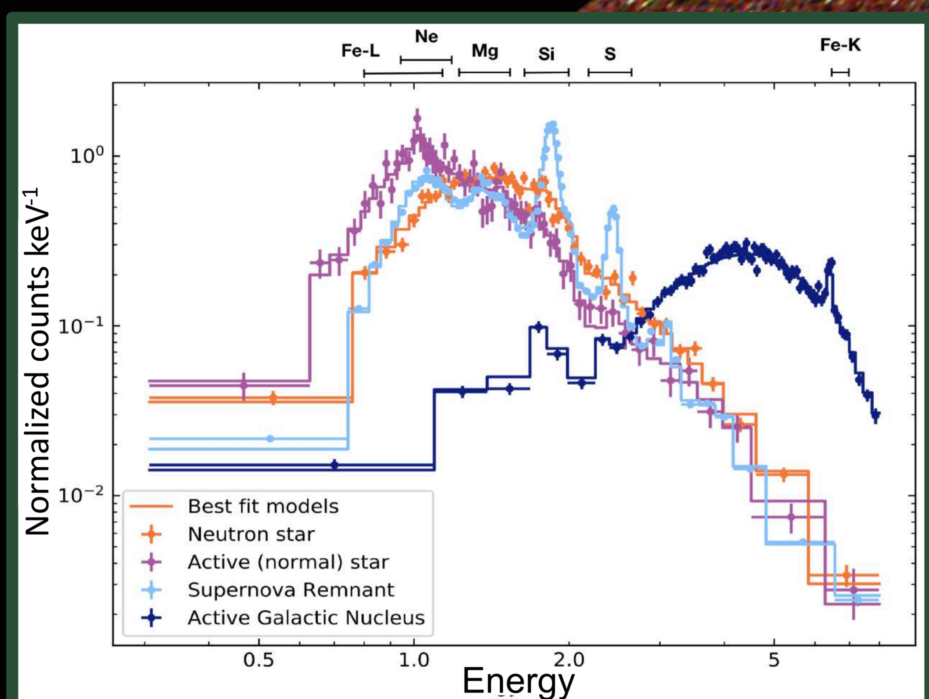
Data Driven Methods to Classify X-ray Sources

Pavan R Hebbar, Craig O. Heinke
University of Alberta

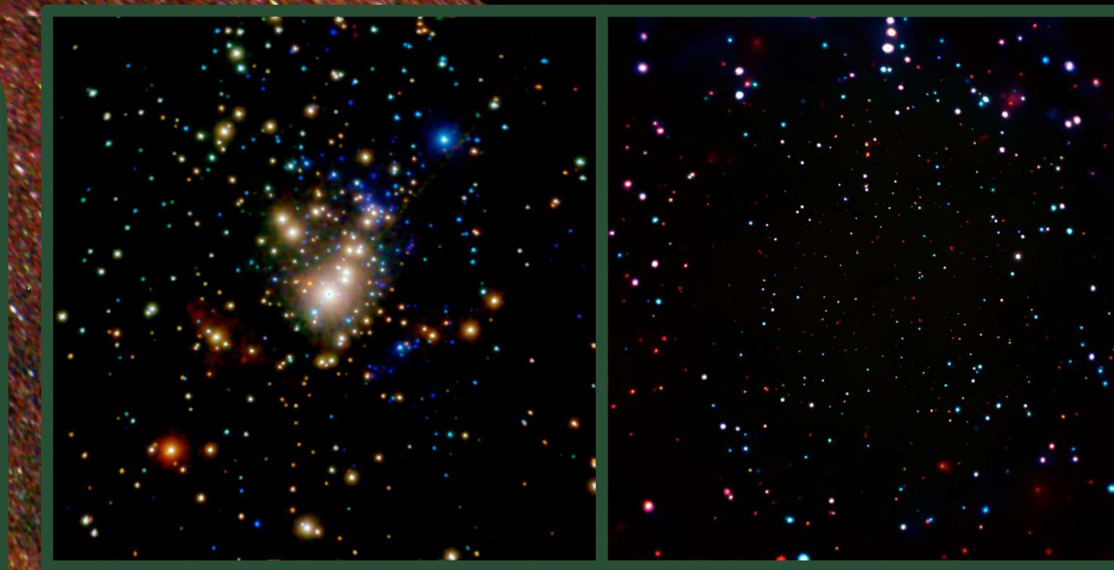


Sample spectra of neutron star, active star, SNR, and AGN.

The X-ray sky

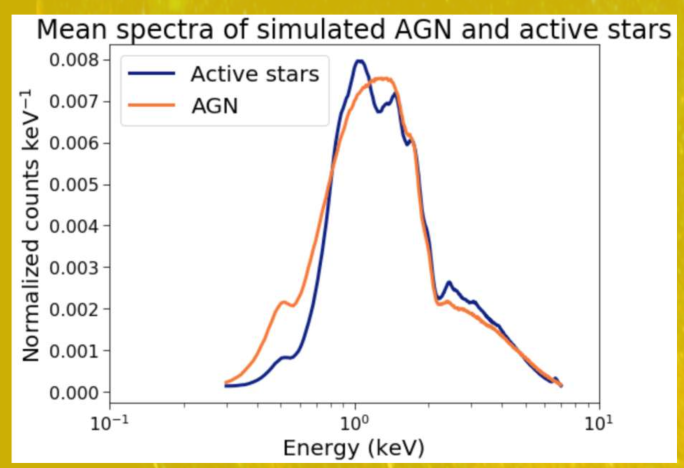
Modern X-ray telescopes have detected hundreds of thousands of X-ray sources (e.g. Evans et al. 2010). These include active galactic nuclei (AGN), supernova remnants (SNRs), active stars, etc. Common methods to distinguish these types of sources include modelling their X-ray spectra or using hardness ratios (HRs) to estimate source properties.

Detailed X-ray spectroscopy of thousands of sources is tedious, and HRs could fail to accurately predict the properties of faint objects. Standard HRs fail to distinguish line-dominated and continuum sources in CCD spectra. Machine-learning (ML) tools that can differentiate such spectra would be accurate and efficient in classifying X-ray sources. Here, we demonstrate the use of neural networks (NNs, an ML method) to differentiate between active stars and AGN.



X-ray image of (*Left*) Stars in Chandra Orion Ultradeep Project (COUP, Credits: CXC/ Fiegelson et al.) (*Right*) AGN in Chandra Deep Field South (CDFS, Credits: CXC/Luo et al.)

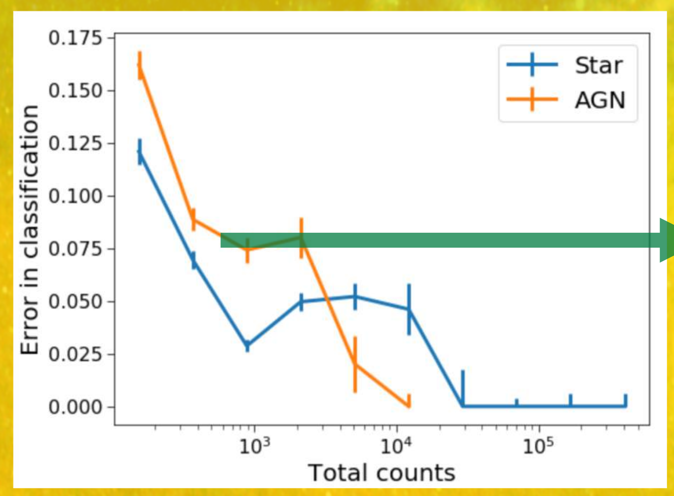
AGN vs. Active star



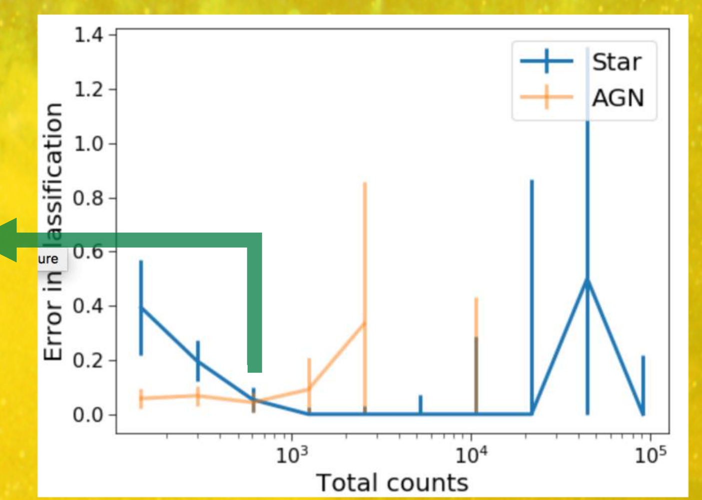
Generated fake spectra of AGN and active stars (100,000 each) from CDFS and COUP. Only spectra with > 100 counts used for analysis.

Active stars have prominent emission lines from Mg, Si, S & Fe. AGN have mainly continuum spectra.

Accuracy increases with more counts



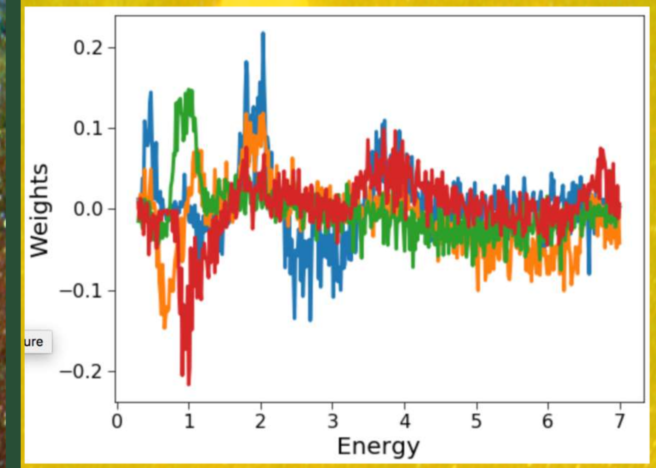
Error < 10% (within error limits) for counts > 400



Net classification accuracy of 91% on the simulated test set (~40,000 spectra not used for training the model).

Net classification accuracy of 90% for the observed spectra of COUP and CDFS sources (~400 spectra in total).

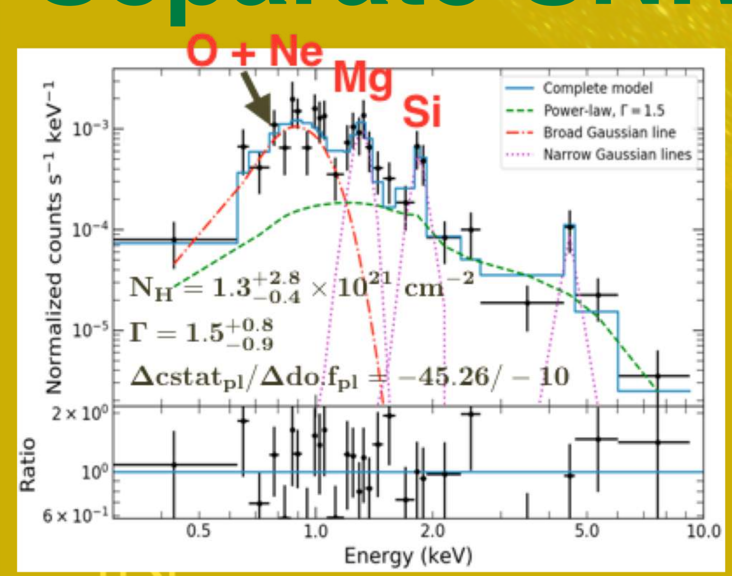
The important energies



The ML model assigns weights to the energy bins. Large weights (magnitude only) imply that the corresponding energies are crucial for classification.

From figure, energies ~0.7–1.5 keV (Fe-L, Mg-K lines), ~1.8–2.5 keV (Si-K, S-K lines), and ~3–4 keV (unknown) are important.

Separate SNR and AGN



Identifying emission lines in dwarf galaxy Henize 2–10 allowed Hebbar et al. (2019) to show that the candidate AGN is most likely a young SNR.

Conclusions & Future

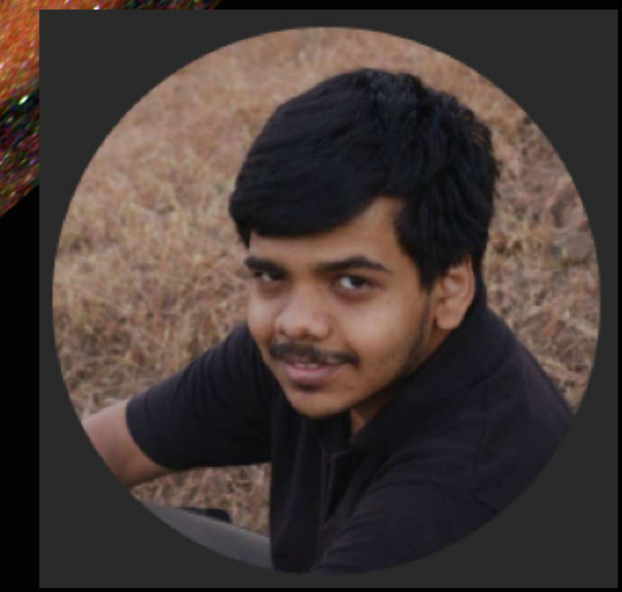
- Tested the use of NN as an accurate and efficient tool to differentiate AGN and active star spectra.
  - 100,000 spectra classified in a few minutes with ~90% accuracy.
  - Model works best for sources > 400 counts.
- Test other ML algorithms to increase accuracy and reduce size of training dataset.
- Apply algorithm for other types of sources.

References

- Evans I. N. et al. 2010, ApJS, 189, 37
- Getman K. V. et al 2005, ApJS, 160, 319
- Hebbar P. R. et al 2019, MNRAS, 485, 5604
- Park T. et al. 2006, ApJ 652, 610
- Tozzi P. et al. 2006, A&A, 451, 457

Acknowledgements

NSERC CRSNG



Pavan R Hebbar
hebbar@ualberta.ca
sites.ualberta.ca/~hebbar