Classifying X-ray Sources with Supervised Machine Learning

Hui Yang[1], Yichao Lin[2], Oleg Kargaltsev[2], Steven Chen[2], Jeremy Hare[3]

[1] Institut de Recherche en Astrophysique et Planétologie (IRAP), Toulouse, France
[2] Department of Physics, The George Washington University, Washington, DC, USA
[3] NASA Goddard Space Flight Center, Greenbelt, MD, USA

Supervised machine learning (ML) algorithms have been applied in X-ray astronomy over the last two decades at an increasing rate, especially as millions of serendipitous X-ray sources have been discovered by modern X-ray observatories like Chandra, XMM-Newton, and more recently eROSITA. In this talk, I will present the development of an automated and efficient multiwavelength ML pipeline, MUWCLASS, designed to classify large numbers of X-ray sources. This enables follow-up observations of interesting sources and population studies of various kinds. MUWCLASS has been applied to Chandra Source Catalog version 2.0 (CSCv2) and the 4XMM-DR13 catalog, augmented with multiwavelength properties obtained through cross-matching with surveys at other wavelengths. I will also discuss common pitfalls that we often encounter in supervised ML, as well as recent developments from our group to potentially address those issues and to improve the performance. Finally, I will present our recent efforts to use this pipeline to identify particle accelerator candidates among the unidentified Fermi-LAT sources.