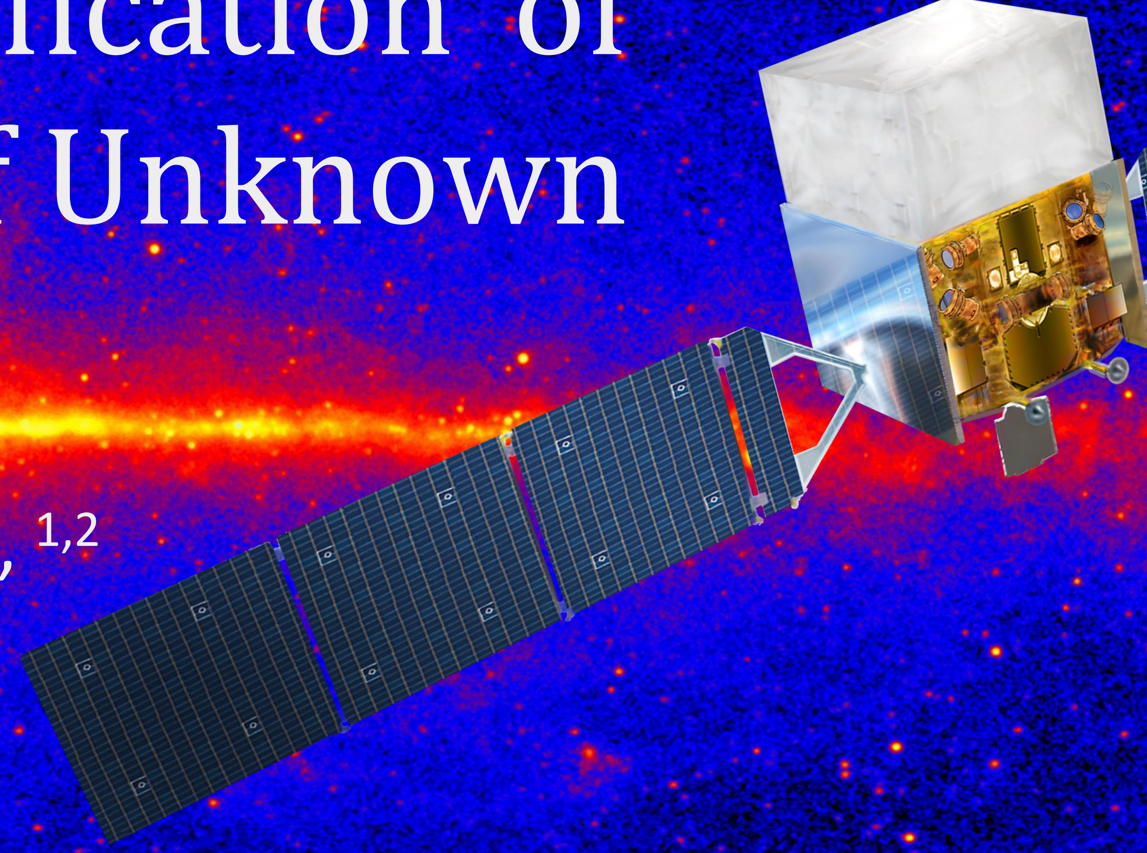
Artificial Neural Network Classification of the Fermi-LAT Catalog Blazars of Unknown Type and Unidentified Sources



[1,2] **Francesco Casini**, [1,2] Paolo Cristarella Orestano, [2] Sara Cutini, [1,2] Stefano Germani, [1,2] Gino Tosti

[1] **University of Perugia, Department of Physics and Geology, Italy**
[2] **Istituto Nazionale di Fisica Nucleare, Perugia, Italy**

francesco.casini@dottorandi.unipg.it

Abstract

The Fermi-LAT detected more than 7000 γ-ray sources in 14 years of operation. Many of these sources are still unassociated with counterparts in other wavelengths, others are associated to generic classes, such as blazar of unknown type, but their classification is still unclear. We consider a Machine Learning approach to the classification of Fermi-LAT unidentified sources and blazars of unknown type using multi-wavelength information. We present the artificial neural network method used to classify the blazars of unknown type and to find possible multi-wavelength counterparts for the Fermi-LAT unidentified γ-ray sources. We describe the multi-wavelength variables that characterized each source and the results obtained using them for the classification.

Introduction

The classification of identified sources and the identification of observed sources with counterparts at other wavelengths have been significant issues in γ-ray astrophysics since its beginnings. The *Fermi*-LAT has detected more than 7000 γ-ray sources in 14 years observation of which about one third are not associated with already known objects (UID), and approximately one fifth are associated with blazars of uncertain nature (BCU). We developed a machine learning method that uses an artificial neural network (ANN) trained with multi-wavelength data. We used this method to classify BCUs into BL Lacs (BLL) or Flat Spectrum Radio Quasar (FSRQ). Then we considered all the possible UIDs multi-wavelength counterparts and we implemented an ANN to find which one was the best candidate and to classify them in Blazar or Not-Blazar sources. To implement the ANN, we used the Python based *Keras* Application Programming Interface [1] and the *TensorFlow* platform [2].

UIDs Counterparts Selection

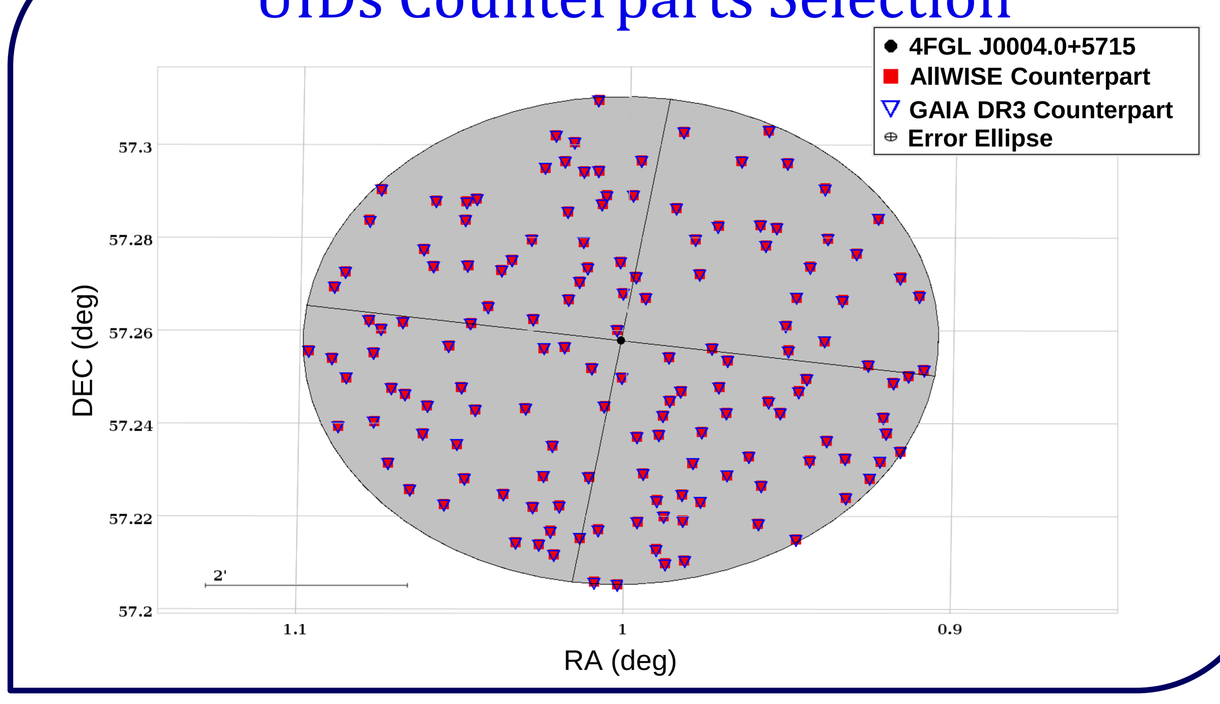


Fig 1: Plot of the counterparts from the AllWISE and GAIA DR3 catalogs that are in the Error Ellipse of the 4FGL J0004.0+5715 source.

Multi-wavelength Data

Sorted and normalized light curve / Spectral Energy Distribution

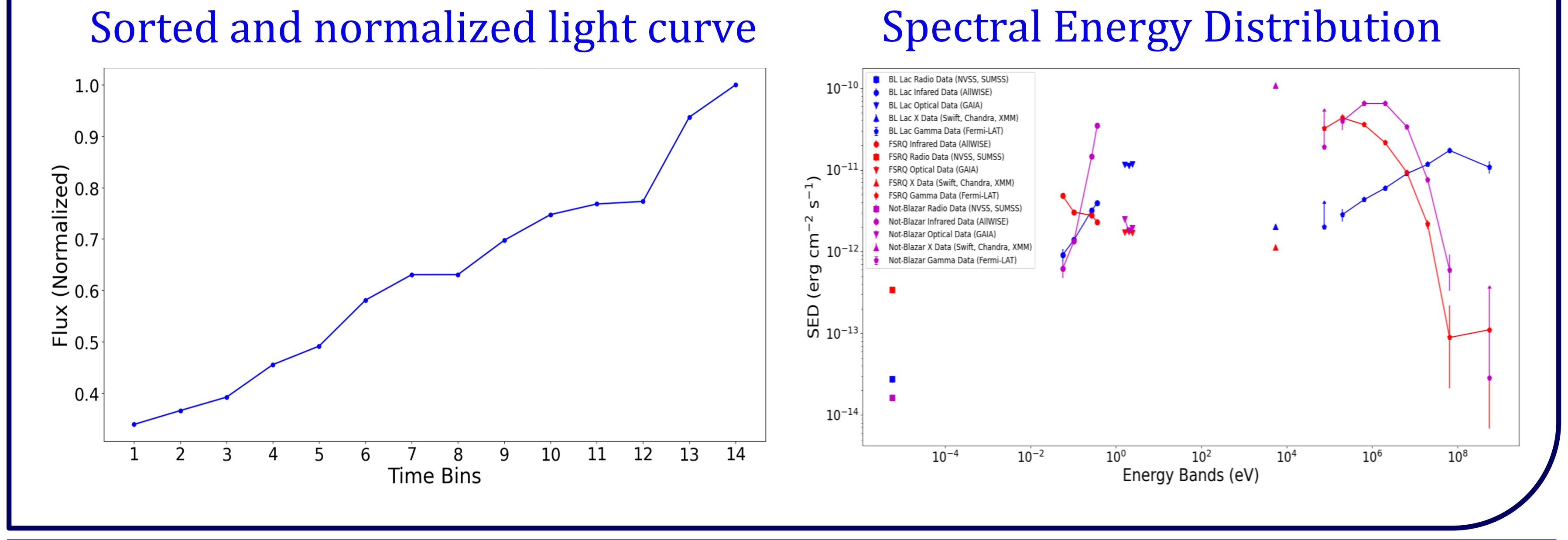


Fig. 2: Information used to build the input dataset **Left**: Sorted and normalized light curve extracted from the 4FGL-DR4 catalog. **Right**: Spectral Energy Distribution in 15 different energy bands from radio to γ-ray extracted from different catalogs.

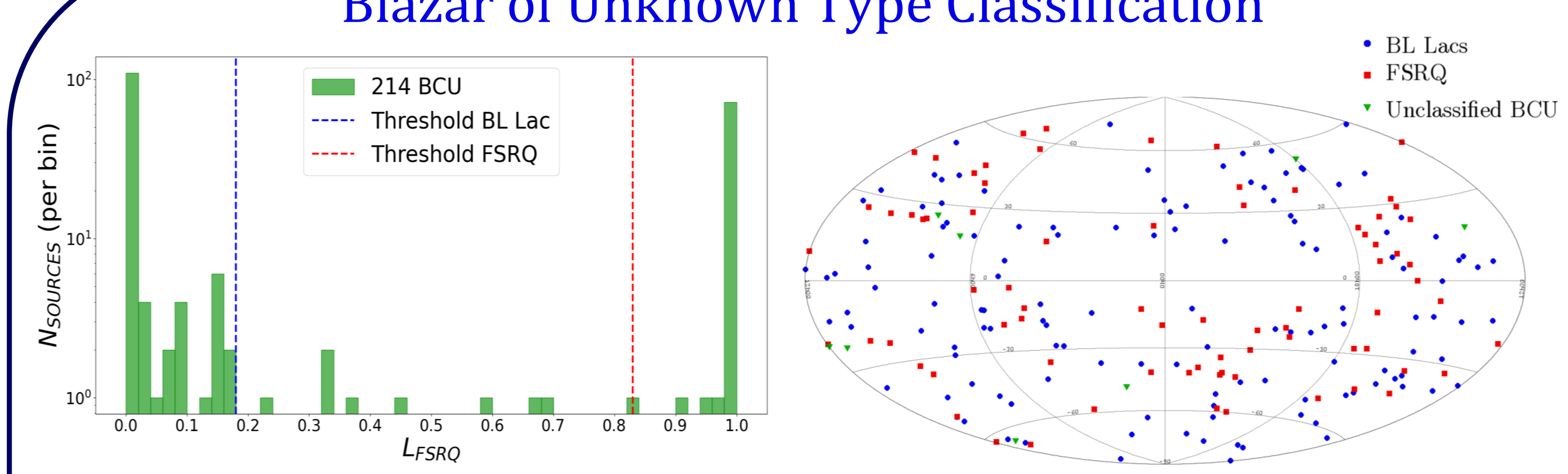
Multi-Wavelength Variables

To classify the sources an ANN must recognize the features that differentiate the classes. We select the physical information that can better distinguish the classes and we use them as input dataset for the ANN training. We considered NVSS [3], SUMSS [4] and MGPS [5] catalogs for radio data, GAIA DR3 catalog [6] for optical data, AllWISE catalog [7] for infrared data, 4XMM DR14 [8], Swift [9] and Chandra [10] catalogs for X-Ray data and 4FGL-DR4 [11] catalog for γ-ray data. We used for the analysis only the 4FGL-DR4 sources that have counterparts in all the other three catalogs. For the identified γ-ray sources we used the known counterparts in the cited catalogs. For the UID we considered all the counterparts from the catalogs that are in the 95% confidence Error Ellipse [Fig 1]. Each of these sources is then characterized by 28 parameters: 14 normalized annual fluxes sorted in ascending order and 15 SED values extracted from the aforementioned catalogs

ANN Classification

To evaluate the effectiveness of the model, we calculated various metrics using a test set. The first ANN is able to correctly identify 96% of the BLLs and 94% of the FSRQs in the test set, while the second one is able to correctly identify 98% of the Blazars and 89% of the Not-Blazars in the test set. The ANN calculates the likelihood of a source to belong to a class (e.g. the likelihood of being a BLL or a FSRQ). To effectively classify the BCUs and the UIDs, we applied thresholds on the likelihood that are found using the test sample; of 214 BCUs, 130 are classified as BL Lacs, 76 as FSRQs and 8 (i.e. 6%) remain unclassified [Fig 3a]. We did the same for the 152 UIDs: 105 are associated with a Blazar counterpart 44 are associated with a Not-Blazar counterpart and 3 remain unclassified [Fig 3c]. The distributions in the sky of the classified sources are shown in [Fig 3b] and [Fig 3d] in Aitoff projection in galactic coordinates.

Blazar of Unknown Type Classification



Unidentified Sources Classification

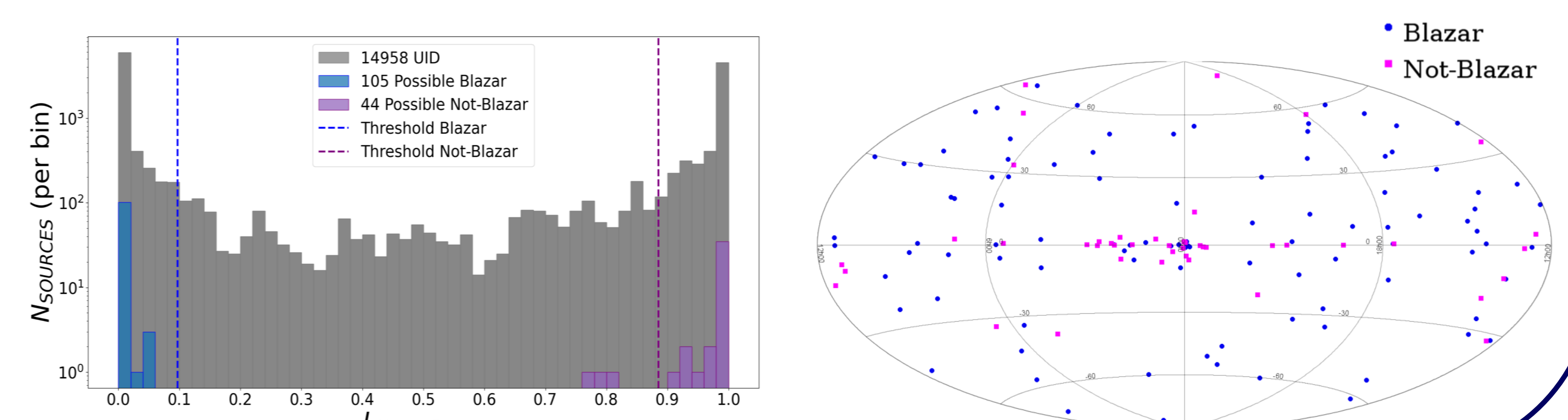


Fig 3: Results of the classification. **a**: likelihood of being a Flat Spectrum Radio Quasar source. The histogram is in logarithmic scale. **b**: distribution of the 214 classified BCUs in Aitoff projection in galactic coordinates. **c**: likelihood of being a Not-Blazar source. In grey are shown all the counterparts, while the best counterpart found by the ANN for each of the 152 UIDs is shown in blue (if classified as Blazar) and purple (if classified as Not-Blazar). The histogram is in logarithmic scale **d**: distribution of the 152 classified UIDs counterparts in Aitoff projection in galactic coordinates.

Conclusions

Using the ANNs trained with this multi-wavelength dataset we are able to achieve both goals described in the introduction. The first ANN is able to classify 96% of the 214 BCUs considered. Then, using the second ANN on the 152 considered UIDs, we can propose a multi-wavelength counterpart for 98% of the UIDs and to classify them as Blazar or Not-Blazar. The distribution of possible UIDs counterparts follows the expected pattern in galactic coordinates although no positional information is given to the ANN. These potential multi-wavelength UIDs counterparts can be further studied through dedicated follow up campaigns from other observatories.

Bibliography

[1] Francois Chollet et al. Keras. 2015. **[2]** Martin Abadi et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems **[3]** J. J. Condon et al. “The NRAO VLA Sky Survey”. **[4]** T. Mauch et al. SUMSS: a wide-field radio imaging survey of the southern sky - II. The source catalogue **[5]** T. Murphy et al. The second epoch Molonglo Galactic Plane Survey: compact source catalogue **[6]** A. Vallenari et al. Gaia Data Release 3: Summary of the content and survey properties **[7]** Edward L. Wright et al. “The wide-field infrared survey explorer (WISE): mission description and initial on-orbit performance” **[8]** Webb N. A. et al. The XMM-Newton serendipitous survey. IX. The fourth XMM-Newton serendipitous source catalogue **[9]** Evans P. A. et al. 2SXPS: An Improved and Expanded Swift X-Ray Telescope Point-source Catalog **[10]** Ian N. Evans et al. The Chandra Source Catalog Release 2 Series **[11]** J. Ballet et al. Fermi Large Area Telescope Fourth Source Catalog Data Release 4 (4FGL-DR4).