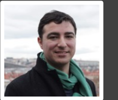
Radio Galaxy prediction with multi-survey data and ensemble Machine Learning

Rodrigo Carvajal[a,b], I. Matute[a,b], J. Afonso[a,b], R. P. Norris[c,d], K. J. Luken[c,e], P. Sánchez-Sáez[f], P. Cunha[g,h], A. Humphrey[g,i], H. Messias[j,k], S. Amarantidis[l,a], D. Barbosa[a,b], H. A. Cruz[m], H. Miranda[a,b], A Paulino-Afonso[g], C. Pappalardo[a,b].

(a) Instituto de Astrofísica e Ciências do Espaço, Universidade de Lisboa. Portugal, (b) Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Portugal, (c) School of Science, Western Sydney University, Australia. (d) CSIRO Space & Astronomy, Australia. (e) CSIRO Data61, Australia. (f) ESO, Germany. (g) Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, Portugal. (h) Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Portugal. (i) DTx CoLAB, University of Minho, Portugal. (j) ALMA Observatory, Chile. (k) ESO, Chile. (l) Institut de Radioastronomie Millimétrique (IRAM), Spain. (m) Closer Consultoria Ltda. Portugal.



rcarvajal@ciencias.ulisboa.pt

Our Goal:
Train and test a Machine Learning model to predict detection of Radio Galaxies and redshift. (Carvajal+ submitted)

Introduction:
High-z AGN can help understanding conditions and processes during Epoch of Reionisation (EoR).

Radio observations allow to study central engines of AGN without major host galaxies' contamination. 10-30% of detected AGN are radio loud[1].

Models and simulations expect large number of AGN detected in the radio at EoR[2,3]. Current observations do not match expectations. No clear cause is known[4].

The use of traditional methods for detection and characterisation (e.g. SED fitting) might not be well suited for incoming data volumes from large surveys (e.g. LSST, SKA, VLASS, ngVLA, etc,). Thus, new approaches are needed. Machine Learning (ML) can help achieving this in viable running times in different areas of the sky.

Data:
Models were trained with **multi-wavelength photometry** from NIR-detected sources in the **HETDEX Spring** field[5].

Pipeline was validated in HETDEX field and in multi-wavelength photometry from the **Stripe 82** field[6].

Methods:
We created three consecuitive models. One to classify between AGN and galaxies. The second to predict radio detection on AGN (i.e. Radio Galaxies, RG). And, the third, to predict redshift values for RG.

Radio detection is a flag that might be tuned for specific sensitivity.

Redshift prediction has been tested before[7].

Results:
For test stages, we obtained **recall** of **96%** for HETDEX and **94%** for Stripe 82 with AGN/Galaxy classfier. For radio detection model, HETDEX offers **recall** of **52%**, while in Stripe 82, it is **58% (better than the 10% fraction of radio-loud QSO)**. Redshift prediction in our pipeline delivers a **NMAD** of **0.07** and an **outlier fraction** of **19%** for HETDEX and **0.09** and **22%** for Stripe 82.

Application of pipeline to undefined sources in both fields (~19M) creates >90k new RG candidates.

From most important features, it is possible to create **AGN selection criterion** (W1-W2 vs. g-r) [Fig. 3].

Prospects:
Our pipeline can be used with future large-area surveys to deliver probable radio detections in a short time.

References:
[1] Macfarlane, S. et al., 2021, MNRAS, 506, 5888
[2] Amarantidis, S. et al., 2019, MNRAS, 485, 2694
[3] Inayoshi, K., Visbal, E. & Haiman, Z., 2020, ARAA, 58:1
[4] Afonso, J. et al., 2015, PoS(AASKA14) 071
[5] Hill, G. J. et al., 2008, Panor. Views Galaxy Form. Evol., 399, 115
[6] Hodge, J. A. et al., 2011, AJ, 142, 3
[7] Carvajal, R. et al., 2021, Galaxies, 9, 86

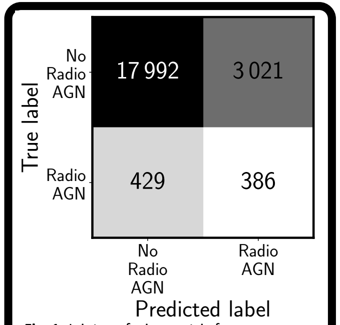


Fig. 1: Joint confusion matrix for AGN/Galaxy and radio detection predictions in data from Stripe 82

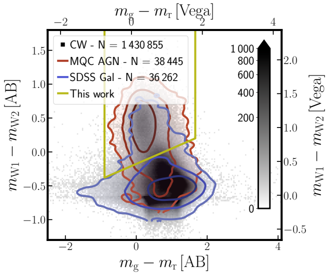


Fig. 3: Colour-colour criterion for AGN detection derived from ML model.

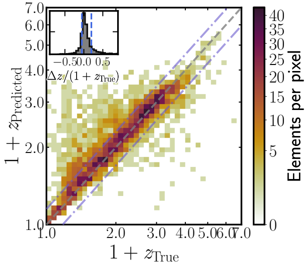


Fig. 2: True and predicted redshifts in data from Stripe 82

Acknowledgements: This work has been supported by the Fundação para a Ciência e a Tecnologia (FCT) through the Fellowship PD/BD/150455/2019 (PhD::SPACE Doctoral Network PD/00040/2012) and POCH/FSE (EC), and through research grants PTDC/FIS-AST/29245/2017, UID/FIS/04434/2019, UIDB/04434/2020 and UIDP/04434/2020.