COFINANCIAMENTO / COFINANCING
ia instituto de astrofísica e ciências do espaço
fct Fundação para a Ciência e a Tecnologia
COMPETE 2020
PORTUGAL 2020
UNIÃO EUROPEIA Fundo Europeu de Desenvolvimento Regional
PD-F FCT PhD PROGRAMMES
Ciências ULisboa

Credit: ESA/Hubble & NASA, Y. Chu

EXTRACTING PHYSICAL RULES FROM ENSEMBLE MACHINE LEARNING FOR THE SELECTION OF RADIO AGN

Rodrigo Carvajal (IA-FCUL, Portugal), I. Matute, J. Afonso, R. P. Norris, K. J. Luken, P. Sánchez-Sáez, P. A. C. Cunha, A. Humphrey, H. Messias, S. Amarantidis, D. Barbosa, H. A. Cruz, H. Miranda, A. Paulino-Afonso, and C. Pappalardo.

This work was supported by the Fundação para a Ciência e a Tecnologia (FCT) through the grant UIDP/04434/2020, UIDB/04434/2020, and the PhD Fellowship PD/BD/150455/2019 (PhD::SPACE Doctoral Network PD/00040/2012) and POCH/FSE (EC).

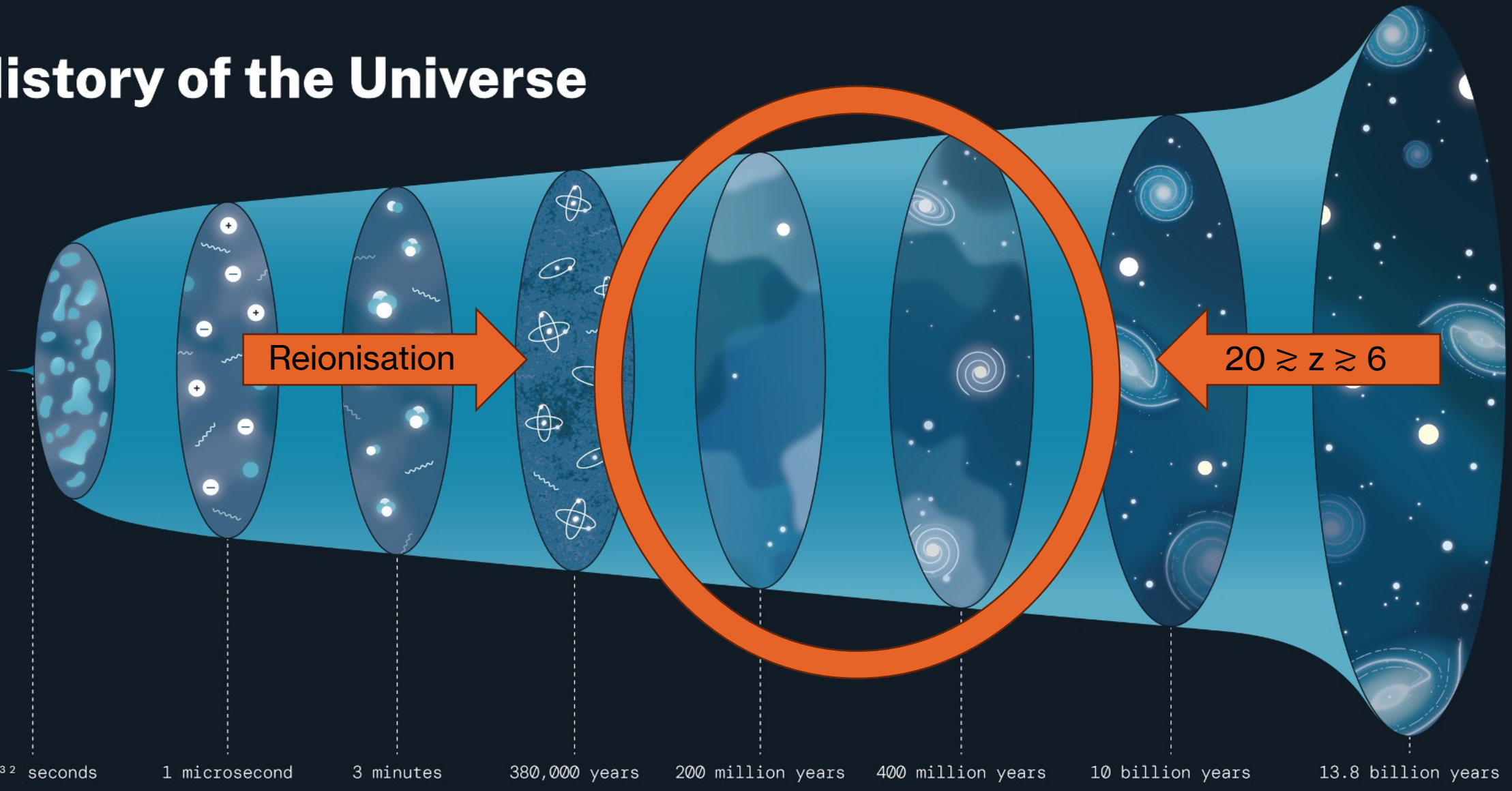
Results mostly from...

‘Selection of powerful radio galaxies with machine learning’

Carvajal et al. 2023 (A&A)

arXiv:2309.11652

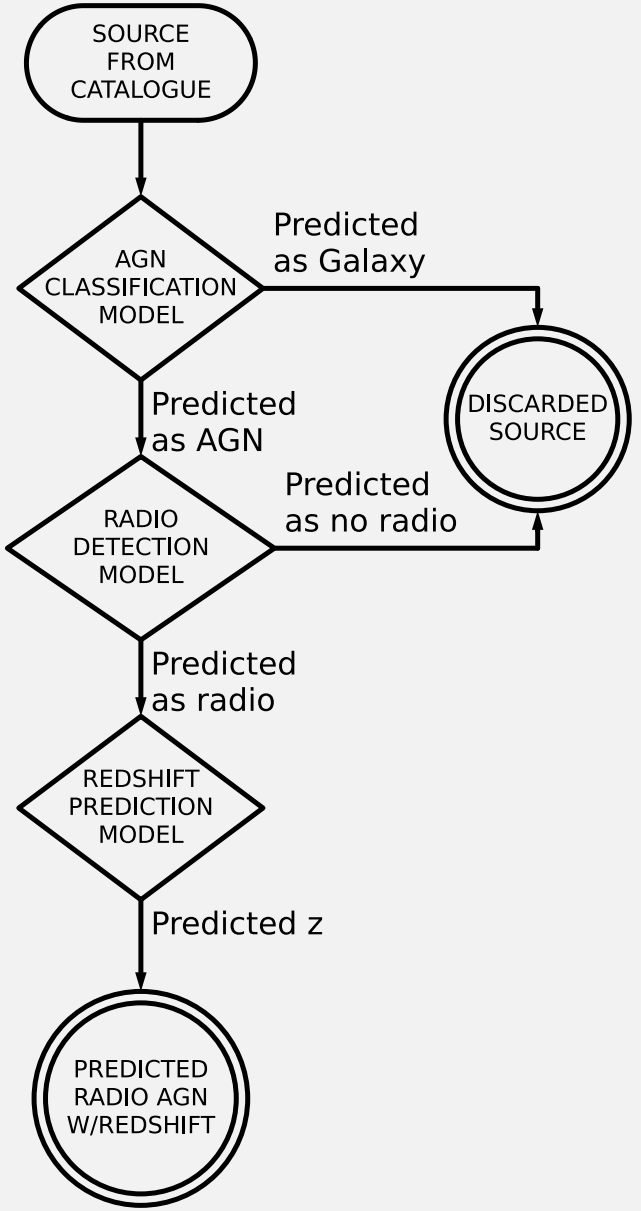
History of the Universe



Credit: NASA 2022

Prediction Pipeline

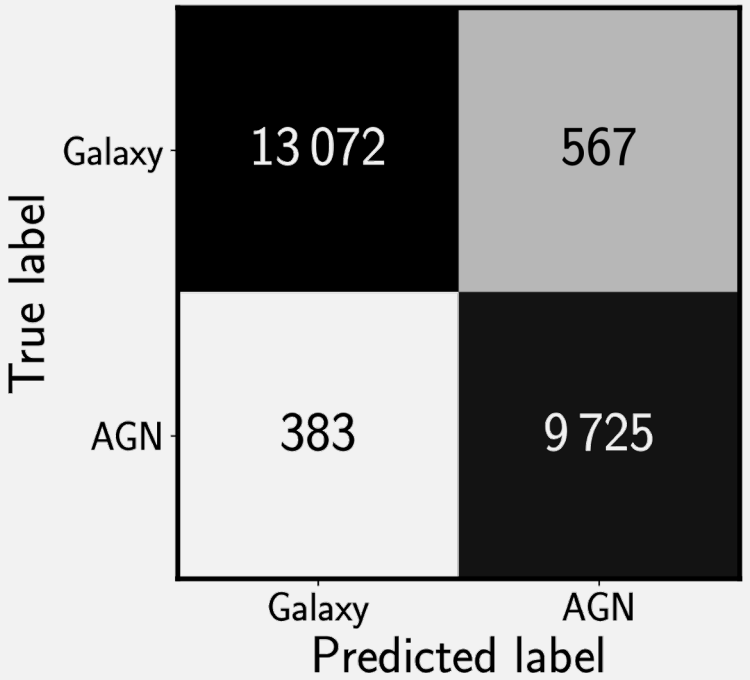
- AGN/Galaxy + Radio-detection + Redshift.
- Model stacking: CatBoost, XGBoost, RF, ET, GBC/GBR (tabular data).
- Training: Photometry from IR-detected sources in HETDEX Spring Field.
- WISE + Pan-STARRS + 2MASS + LoTSS detection.
- Validation: HETDEX & Stripe 82

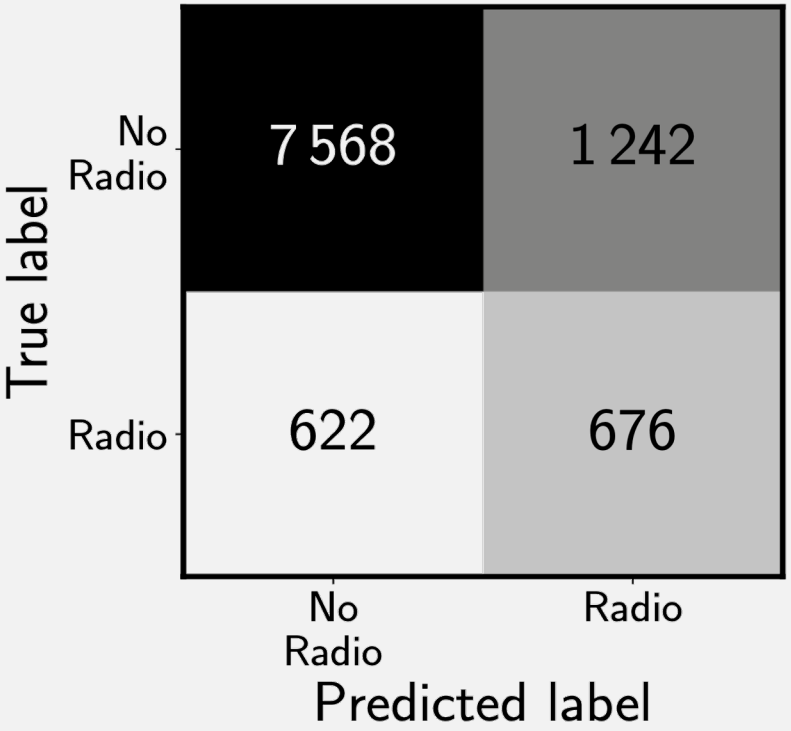


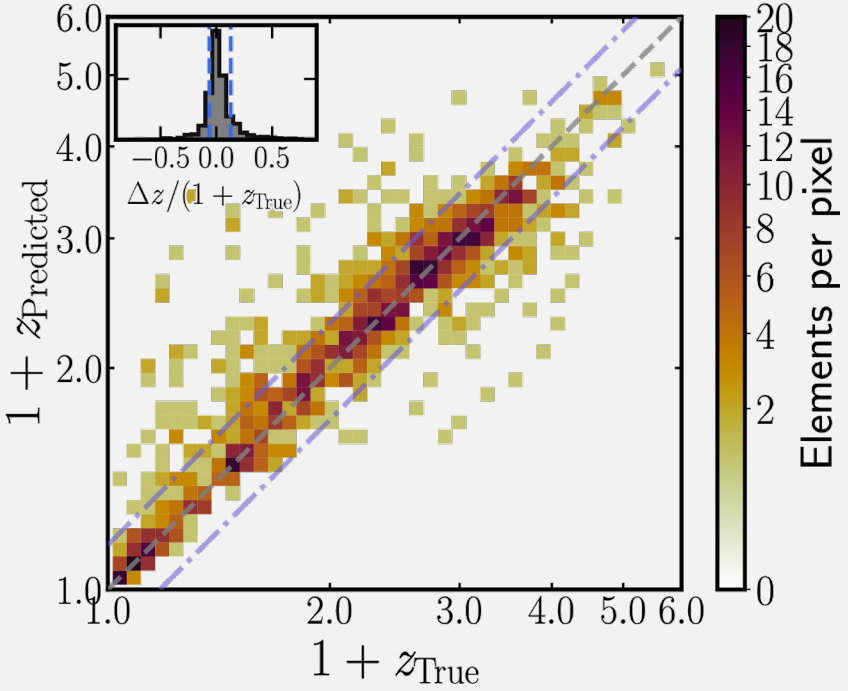
Carvajal et al., 2023

RESULTS

Carvajal et al., 2023







Completeness: 96%
Precision: 95%

Completeness: 52%
Precision: 35%

σ_{NMAD}: 7%
Outlier fraction: 19%

Feature Importances: SHAP values

- Properties with higher predicting power.
- Possible for individual sources.
- Example with AGN/Galaxy classification: W1-W2 & g-r.

AGN-Galaxy model

Feature	SHAP value	Feature	SHAP value	Feature	SHAP value
W1_W2	32.458	i_y	5.086	z_y	1.591
g_r	11.583	y_W1	4.639	H_W3	1.048
W1_W3	8.816	band_num	4.050	W4mag	0.514
r_i	7.457	y_W2	3.228	H_K	0.466
i_z	6.741	z_W2	2.348	W3_W4	0.466
r_J	6.613	y_J	1.718	J_H	0.178

Radio detection model

Feature	SHAP value	Feature	SHAP value	Feature	SHAP value
g_i	14.120	z_W1	6.751	W4mag	2.691
W2_W3	13.201	r_i	5.577	band_num	2.661
g_r	12.955	r_z	5.161	K_W4	0.939
y_J	8.224	i_z	4.512	H_K	0.719
K_W3	7.441	z_y	4.121	J_H	0.190
W1_W2	6.874	y_W1	3.864		

Redshift prediction model

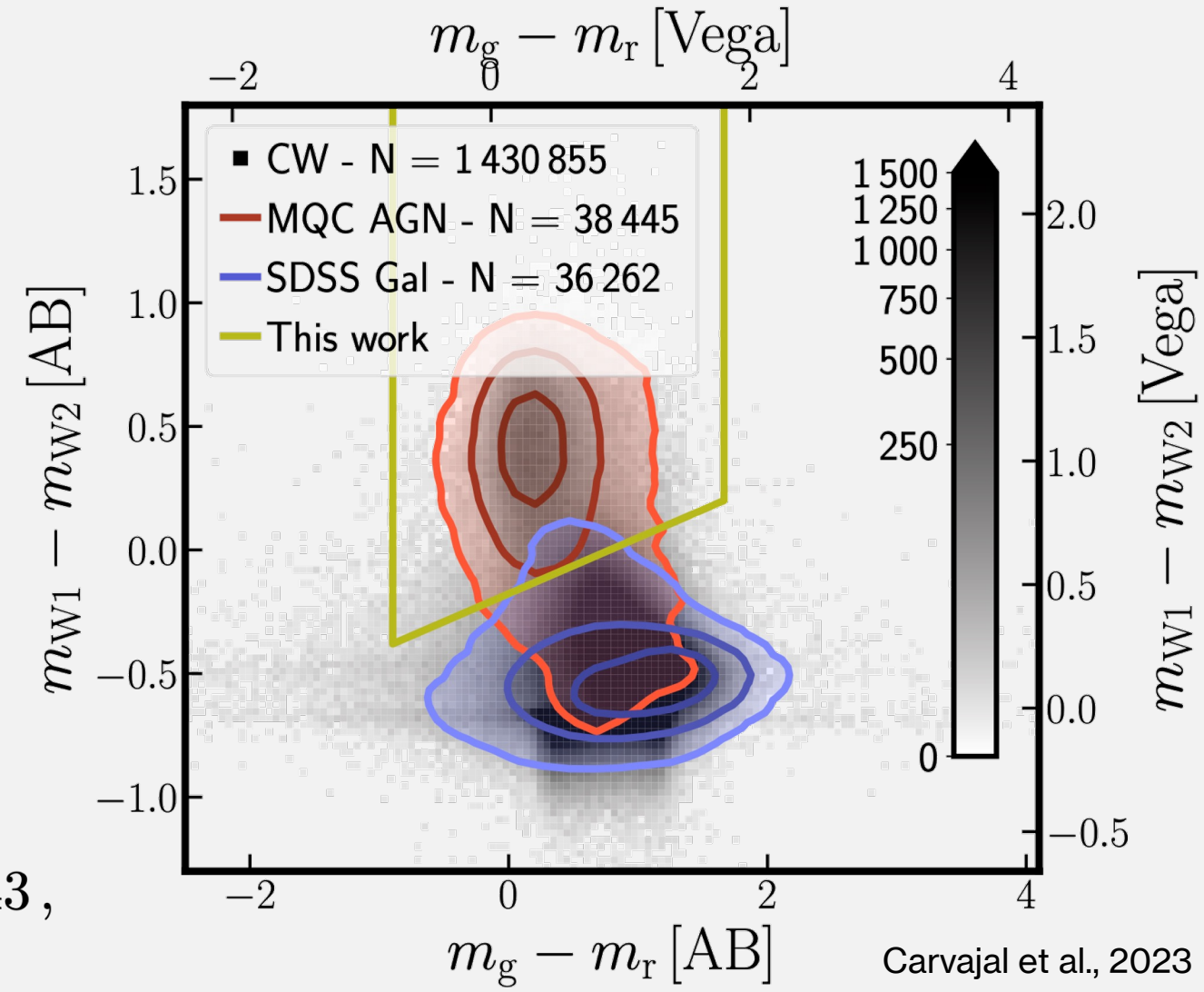
Feature	SHAP value	Feature	SHAP value	Feature	SHAP value
g_r	32.594	z_y	3.557	W4mag	1.639
y_W1	20.770	y_J	3.010	g_W3	1.479
W2_W3	12.462	band_num	2.595	K_W3	0.853
W1_W2	5.692	i_y	2.381	K_W4	0.451
r_i	4.381	H_K	2.230	J_H	0.146
r_z	3.755	i_z	2.005		

Carvajal et al., 2023

A NEW COLOUR-COLOUR CRITERION

As efficient as previous IR colour-colour criteria.

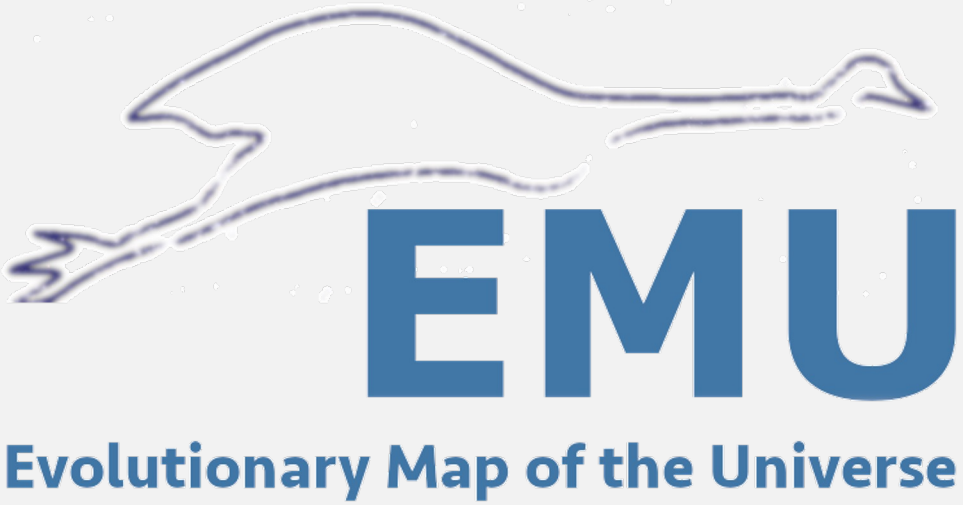
$$
\begin{aligned}
g - r &> -0.76\,, \\
g - r &< 1.8\,, \\
W1 - W2 &> 0.227 \times (g - r) + 0.43\,,
\end{aligned}
$$



Carvajal et al., 2023

To summarise...

- (Astro-) Physical rules can be extracted from traditional ML models.
- One step towards AGN (radio) – Galaxy (IR+optical) connection.
- Possible to use in SKA and Pathfinders footprints and more.



Thank you!

‘Selection of powerful radio galaxies with machine learning’

R. Carvajal et al. 2023 (A&A)

arXiv:2309.11652

racarvajal@ciencias.ulisboa.pt

racarvajal.github.io