

Machine learning AGN-SFG classifier for extragalactic radio surveys

Berta Margalef Bentabol
Jesper Karsten

Astronomy & Astrophysics manuscript no. aanda
June 9, 2023

©ESO 2023

A multi-band AGN-SFG classifier for extragalactic radio surveys using machine learning

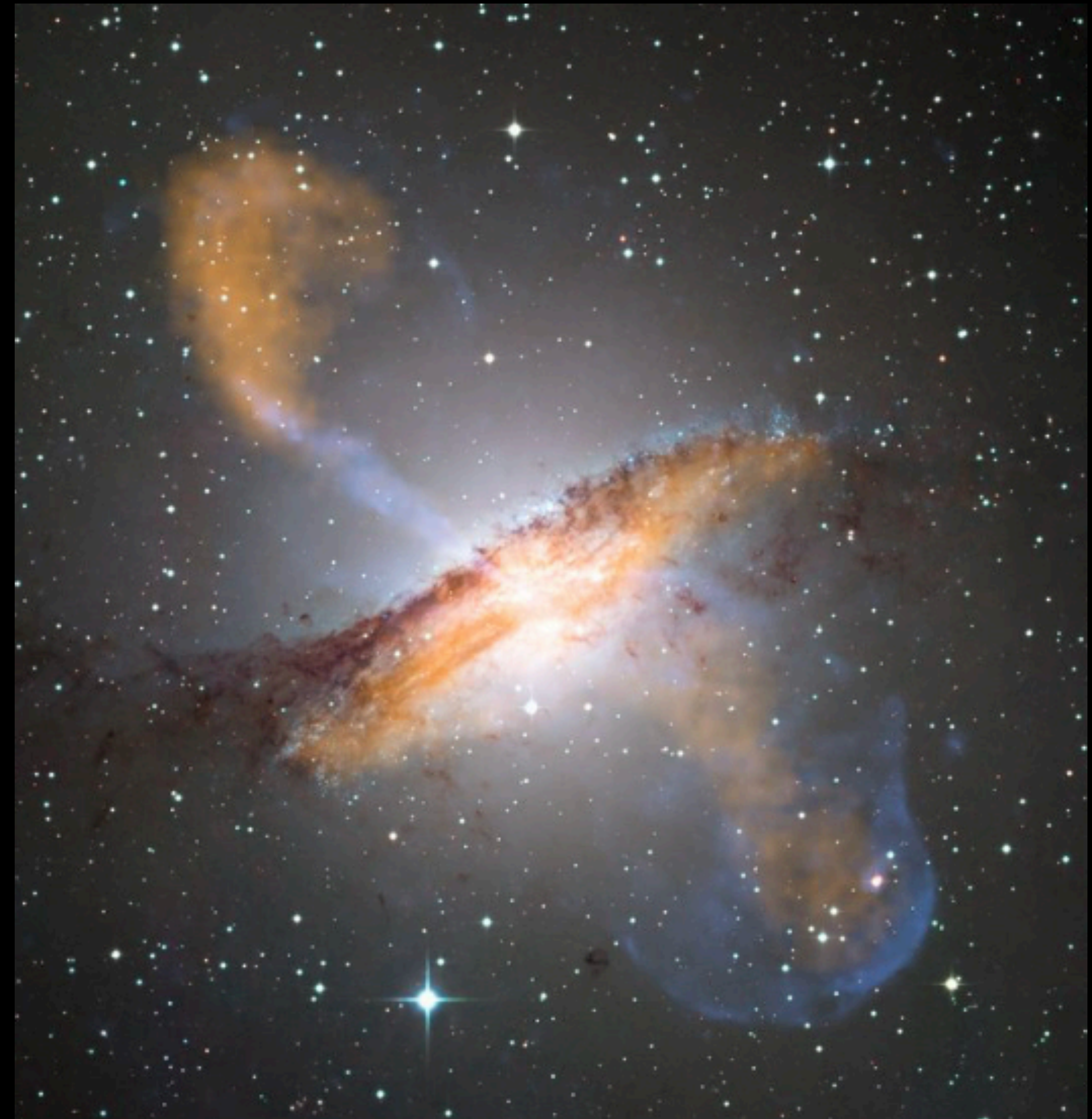
J. Karsten¹, L. Wang^{1,2}, B. Margalef-Bentabol², P. N. Best³, R. Kondapally³, A. La Marca^{1,2}, R. Morganti^{1,4}, H.J.A. Röttgering⁵, M. Vaccari^{6,7,8}, and J. Sabater^{3,9}

SRON

Netherlands Institute for Space Research

Radio surveys

- Need to separate SFGs from AGNs.
- A lot of effort from experts in LOFAR Deep Fields.
- Can we use machine learning to reproduce this classification?



Radio surveys

Classes

		Energetic output	
		Low-excitation (Jets)	High-excitation (Electromagnetic radiation)
Radio output	Radio-quiet	SFG	RQ
	Radio-loud	LERG	HERG

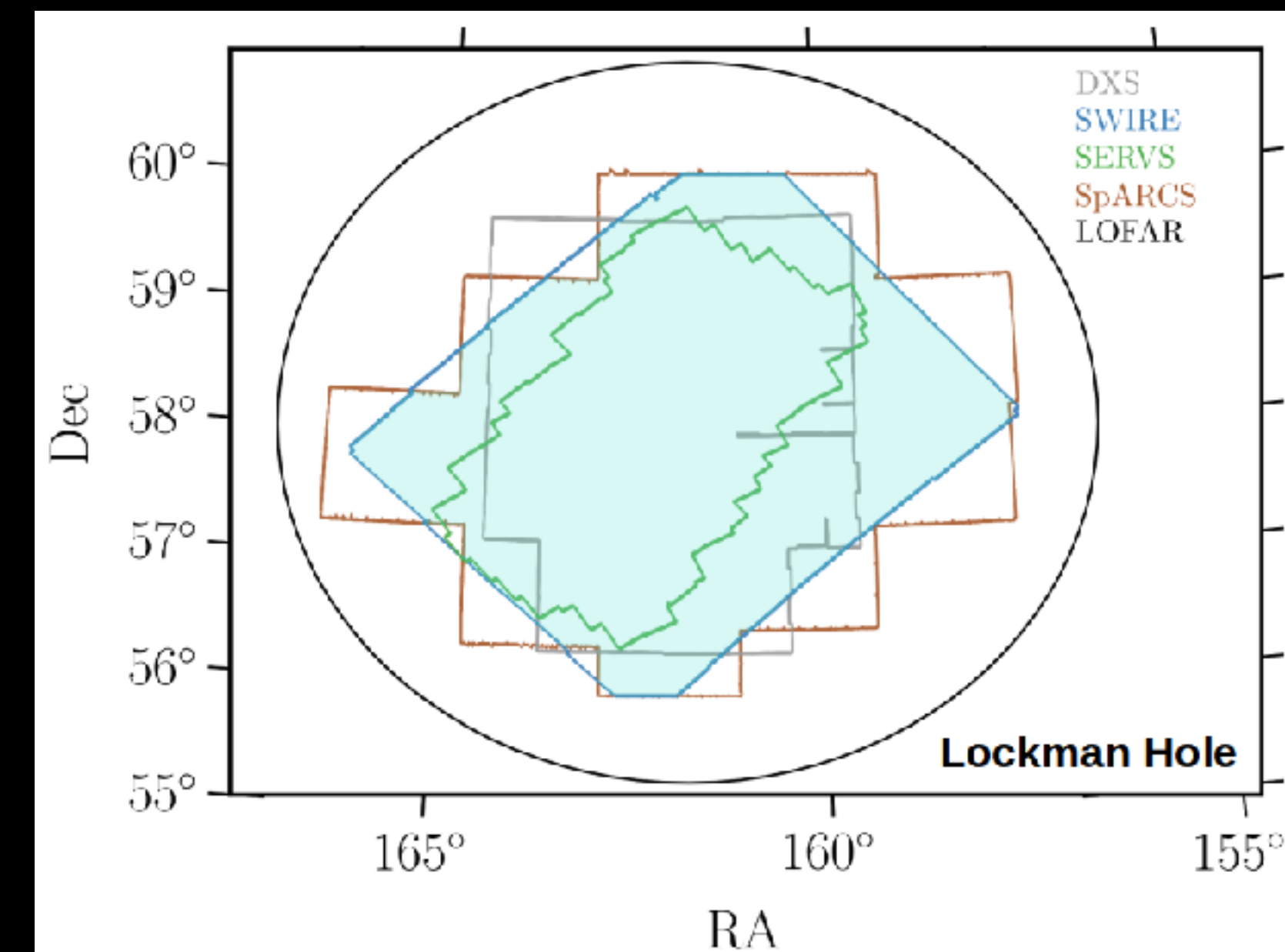
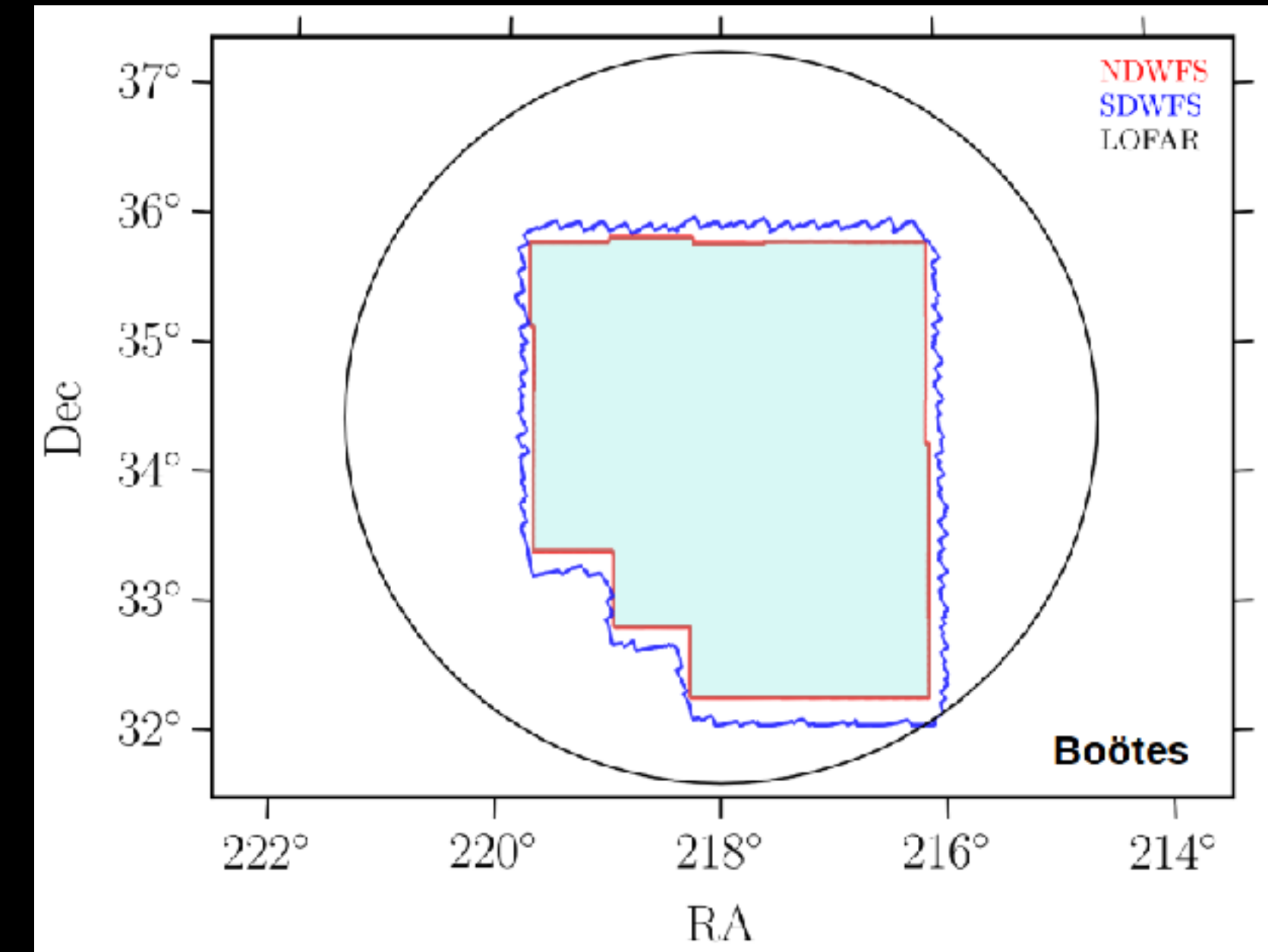
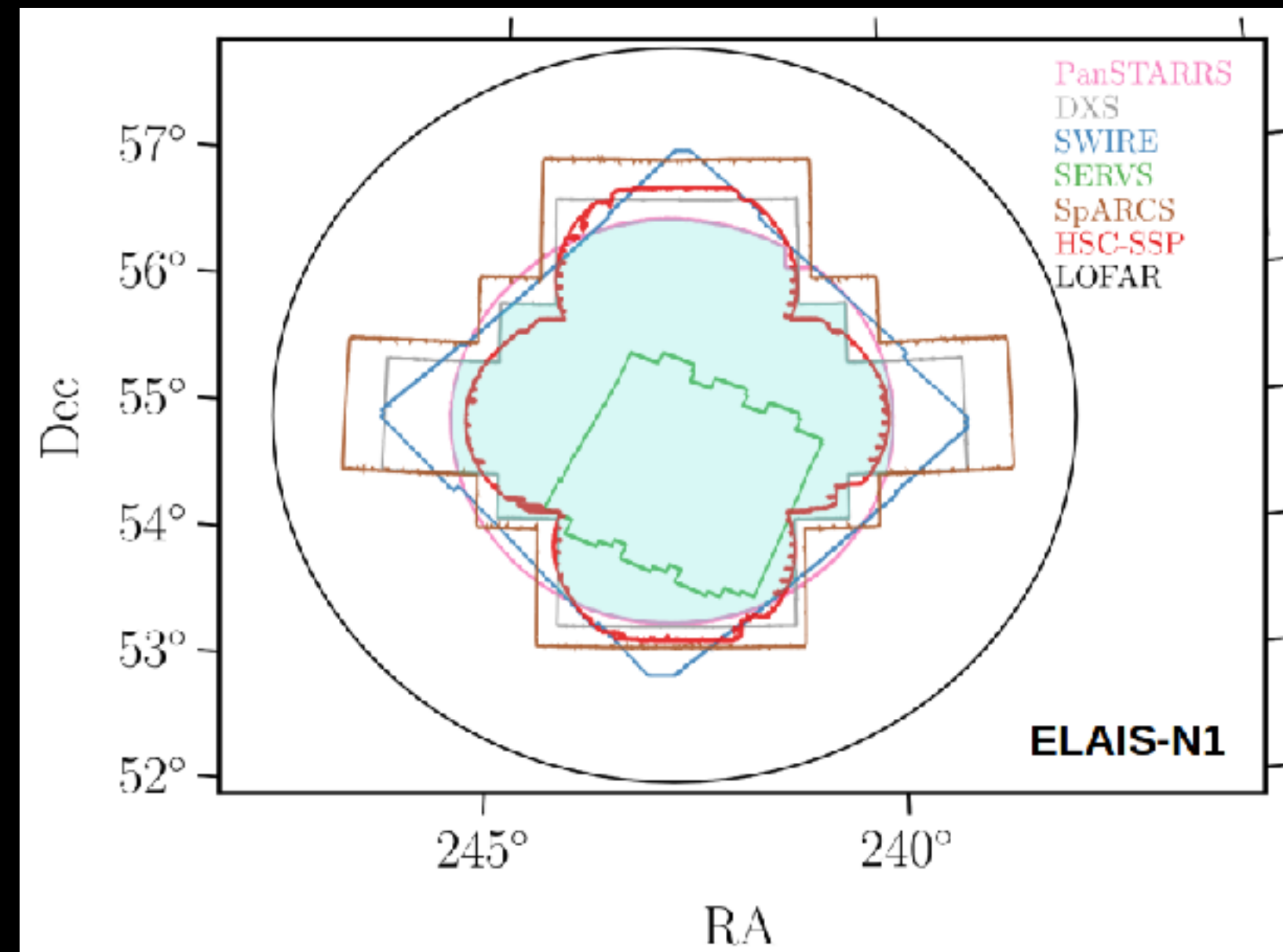
Data

LOFAR deep fields

3 LOFAR deep fields:

- ELAIS-N1
- Boötes
- Lockman Hole

Radio data from
LOFAR (150 MHz)



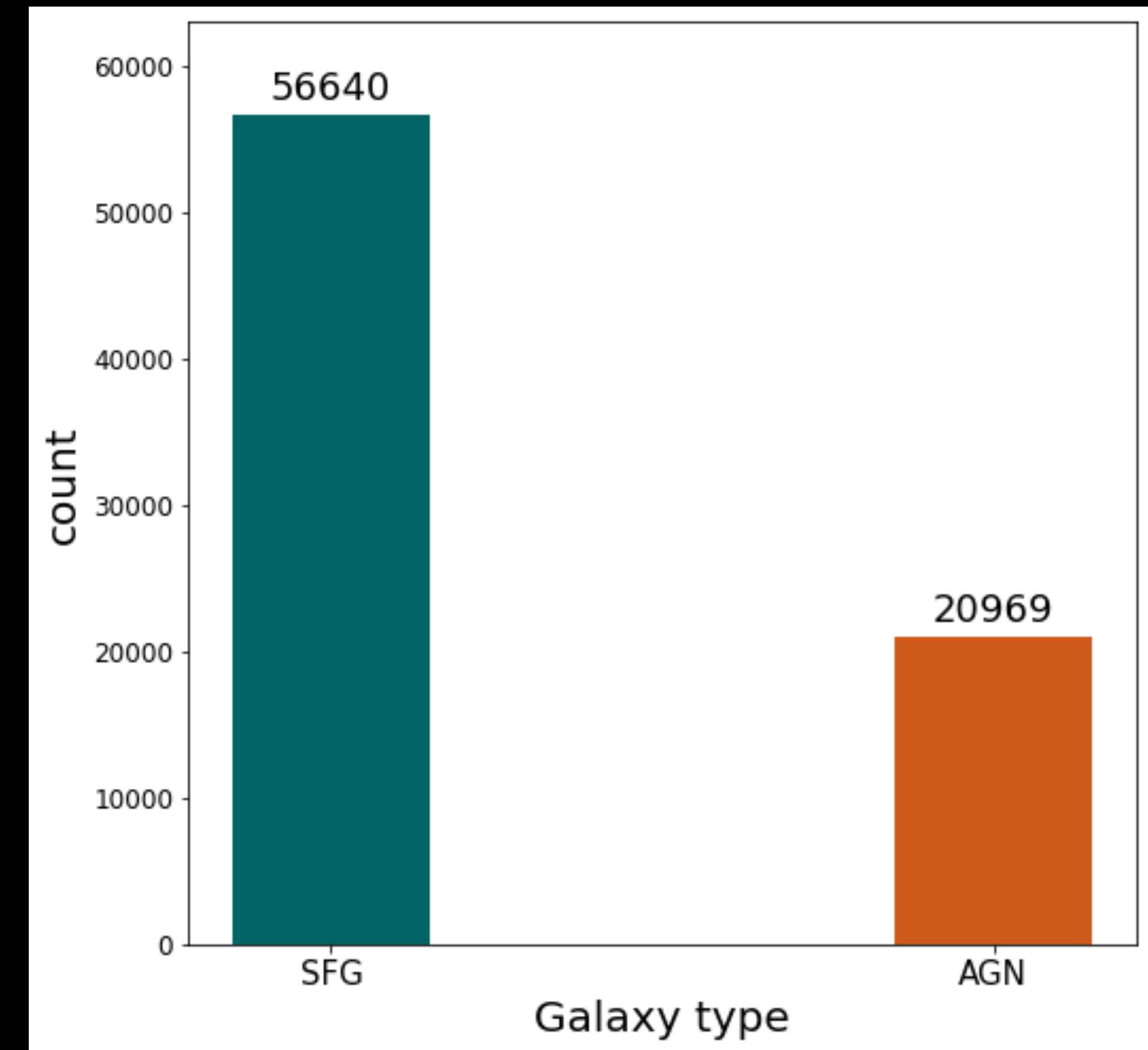
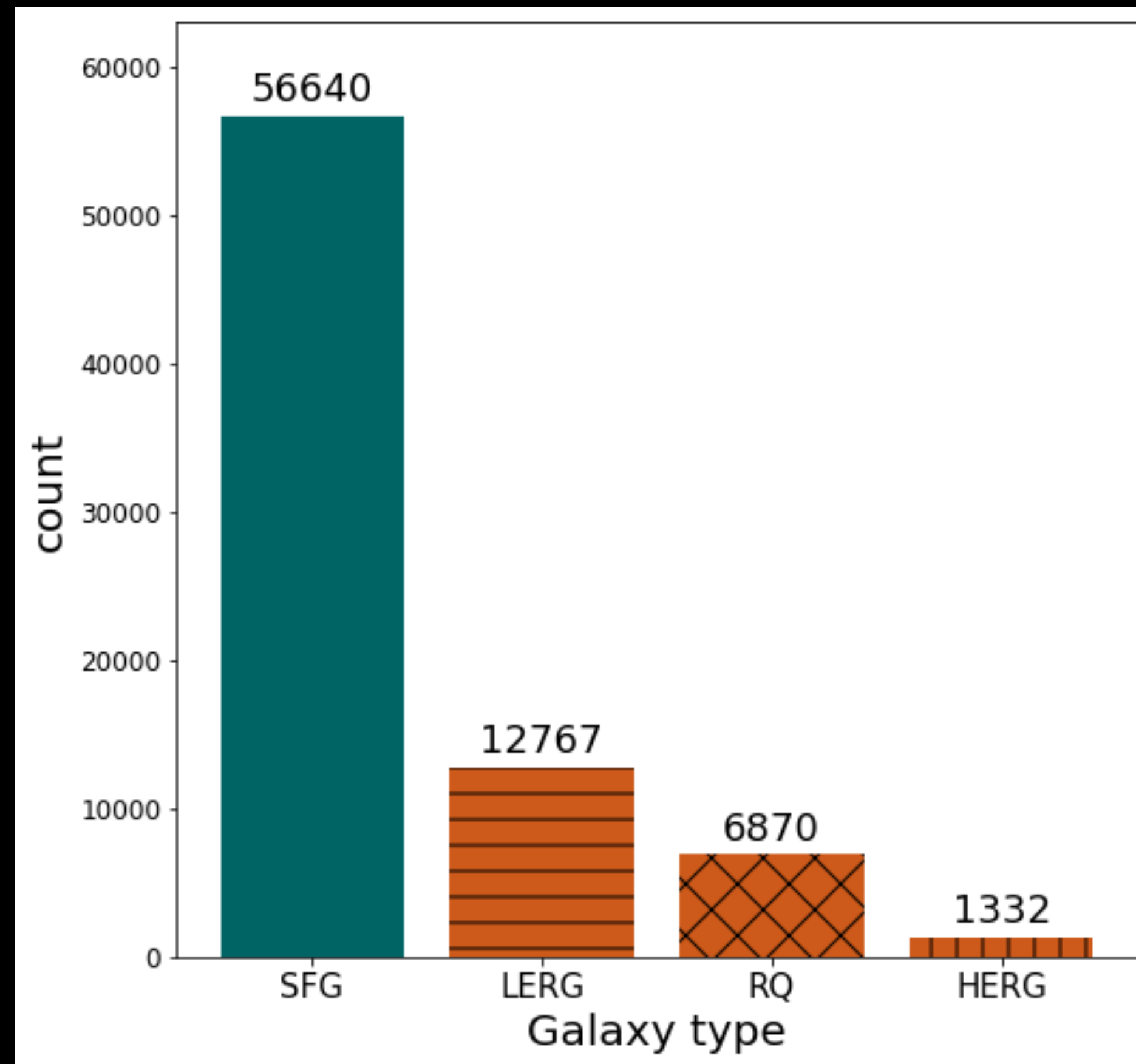
Data

Additional data

- Flux densities from the UV to the FIR (and radio) -> Sometimes different filters in each fields.
- Redshifts.
- Classifications derived from SED analysis (AGN vs SFG).

Data

LOFAR deep fields



Methods

Supervised Machine Learning

Why machine learning?

- No need for tedious classification from experts.
- Can reproduce expert's classification.
- Quickly classify different datasets.

Our approach

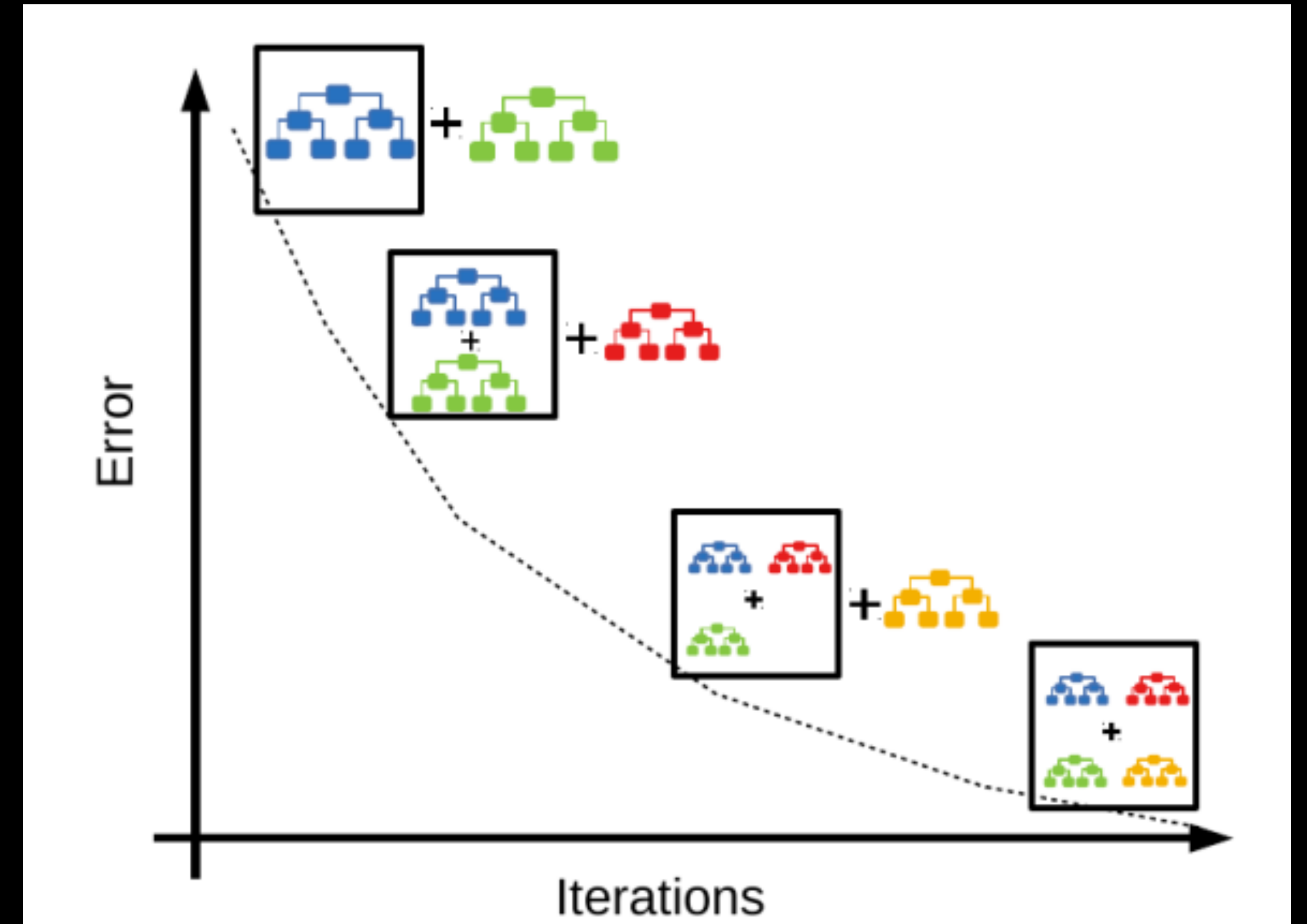
- Focus on binary classification (AGN vs SFG).
- Not enough data for 4-class classification.

Methods

LightGBM

- Supervised ML method.
- Based on decision trees.
- Automatic missing value handling.
- Bayesian optimisation.

Trained on LOFAR deep fields.



Methods

Metrics

Precision: $P = \frac{TP}{TP + FP}$ (Reliability)

Recall: $R = \frac{TP}{TP + FN}$ (Completeness)

F1-score: $F_1 = \frac{2}{R^{-1} + P^{-1}}$

TP: True Positives

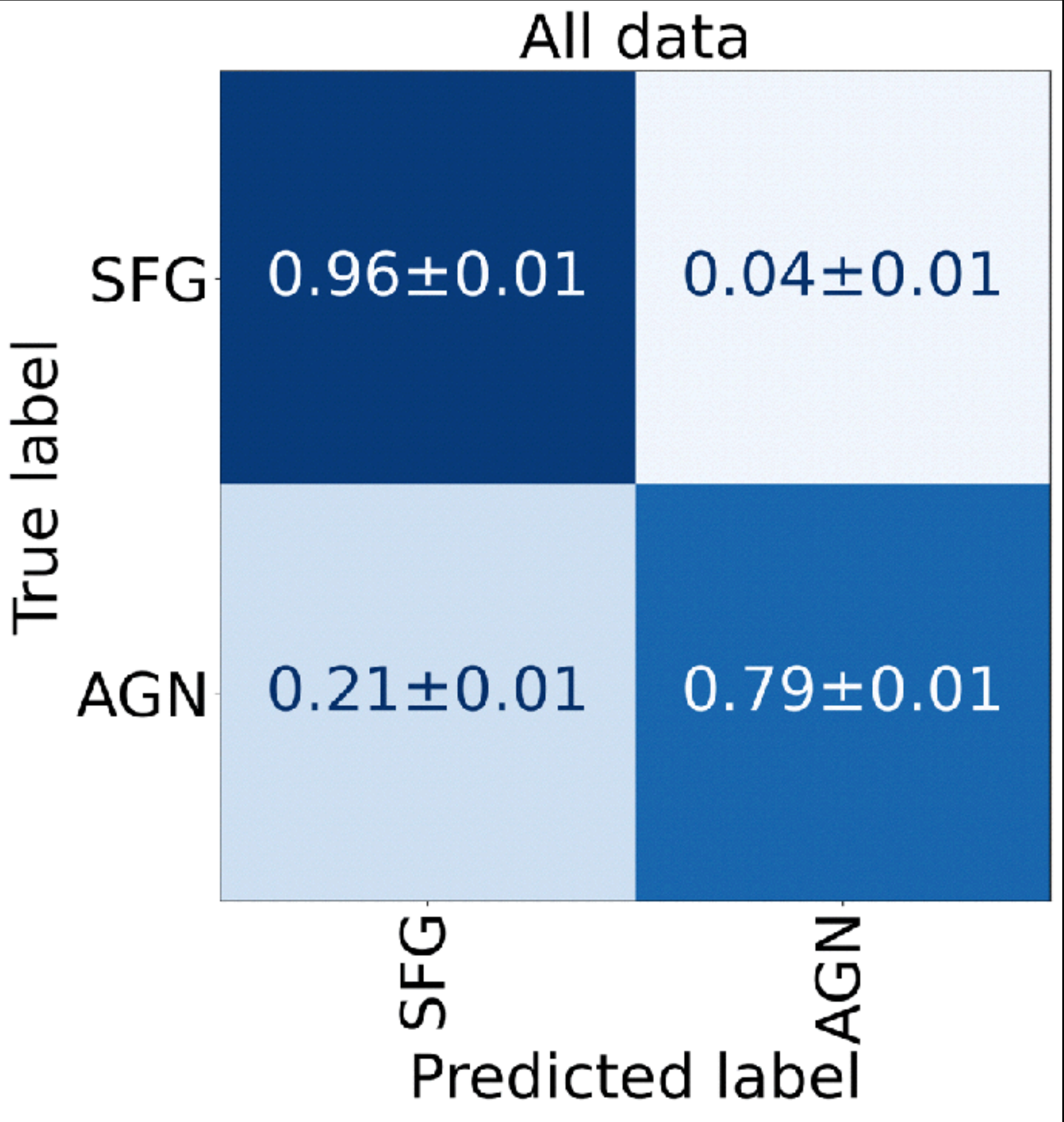
FP: False Positives

FN: False Negatives

Results

Binary classification

Performance on LOFAR deep fields

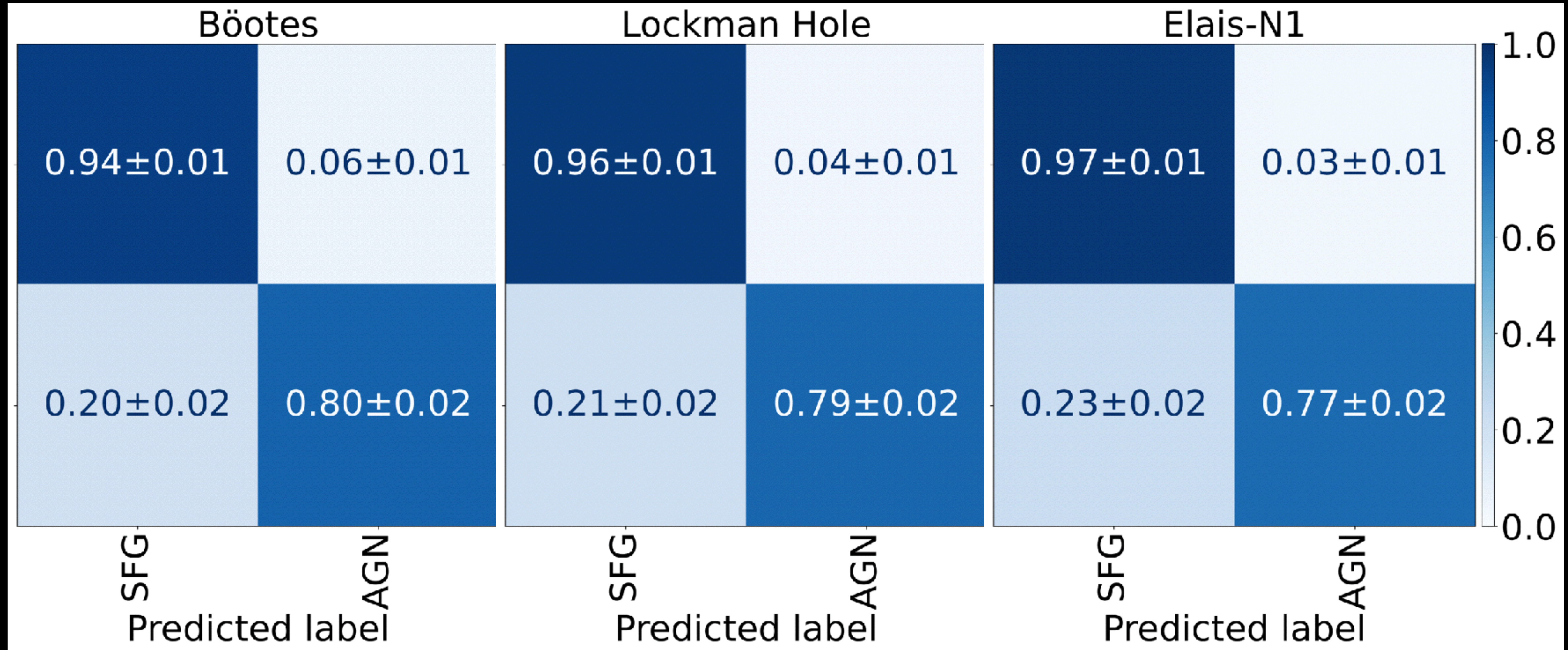


	P	R	F1-score
SFG	92%	96%	94%
AGN	87%	79%	83%
All	90%	87%	88%

Results

Binary classification

Performance on LOFAR deep fields



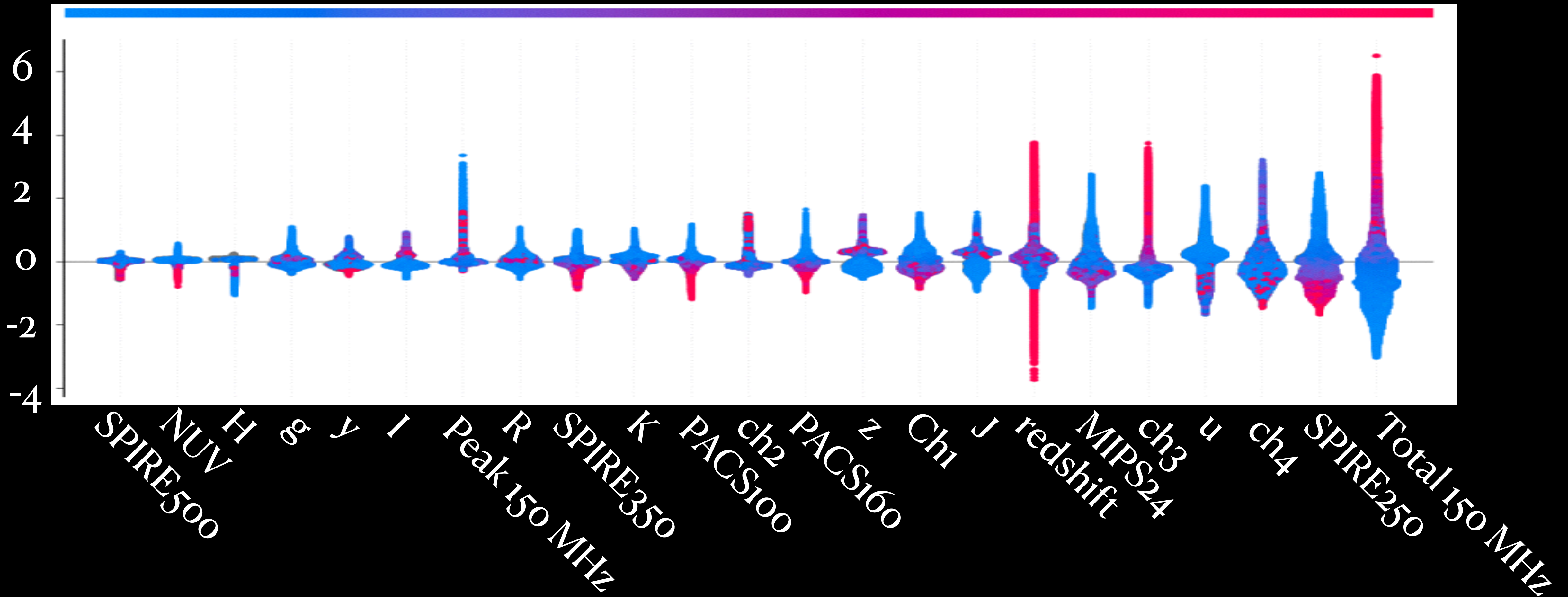
Results

Feature relevance

Low

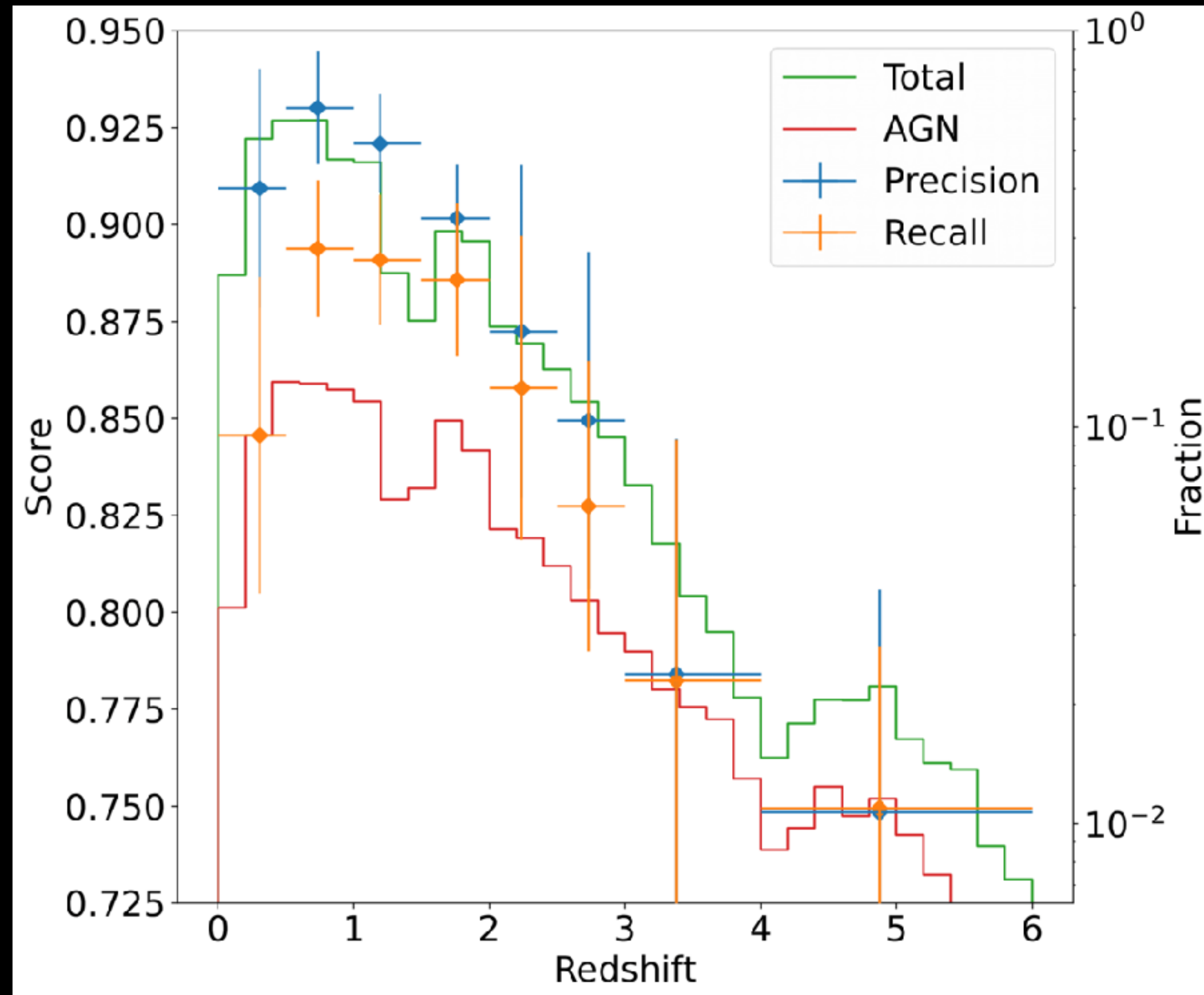
High

SHAP



Results

Dependance on sample size



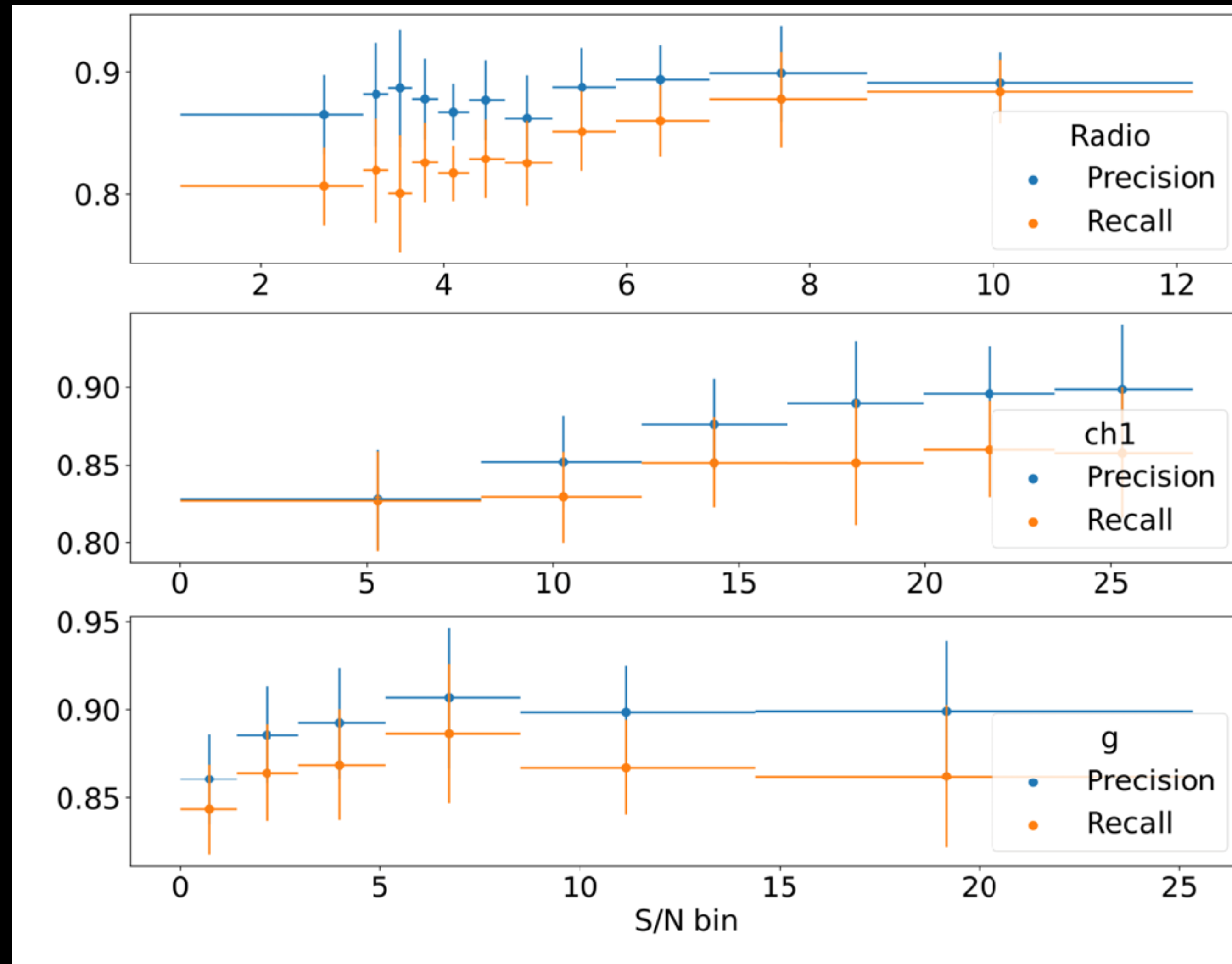
Results

Dependance on SED sampling

	P (SFG)	R (SFG)	P (AGN)	R (AGN)	F1-score
NUV, U, grippy, J, H, K, ch1-ch4, MIPS, PACS, SPIRE	95%	89%	69%	83%	83%
NUV, U, grizy, J, H, K, MIPS, PACS, SPIRE	94%	87%	62%	81%	80%
NUV, U, grizy, J, H, K, 150 MHz	94%	84%	53%	77%	76%
grizy	94%	79%	34%	71%	66%

Results

Dependance on signal-to-noise



Conclusions

- Supervised LightGBM model can classify AGN vs SFG.
- SED sampling important for performance.
- Higher S/N for some bands results in better performance.
- 1.4 GHz radio data can also be used
- Model available at <https://github.com/Jesper-Karsten/M>



Thank you!

✉ B.Margalef.Bentabol@sron.nl

