Machine learning applications & outlook for radio astronomy

O. Ivy Wong (CSIRO) | 15 June 2023

Australasian Leadership Computing Syposium, Canberra

Acknowledgement of country

I acknowledge and pay my respect to

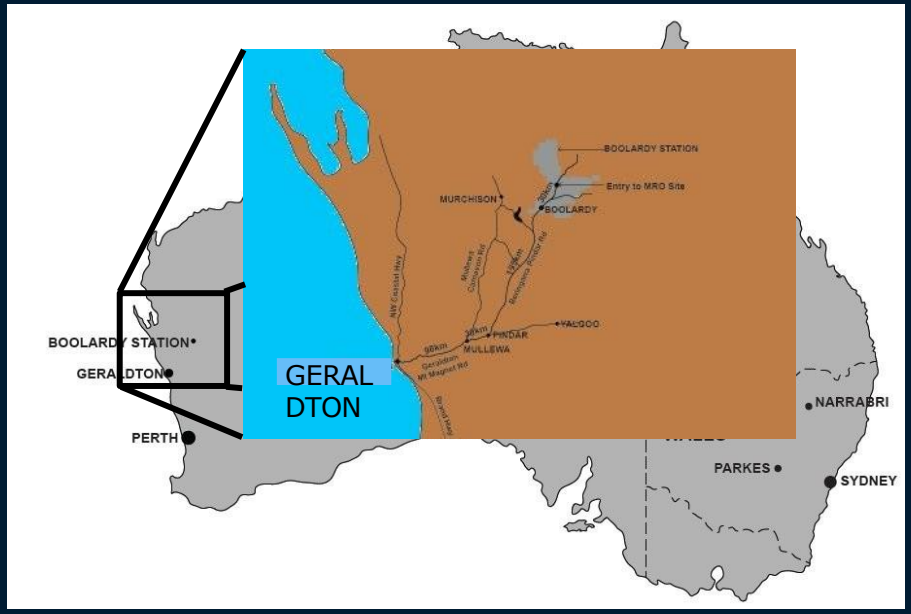
the Whadjuk Noongar people, the traditional custodians of the lands on which I live and work,

and

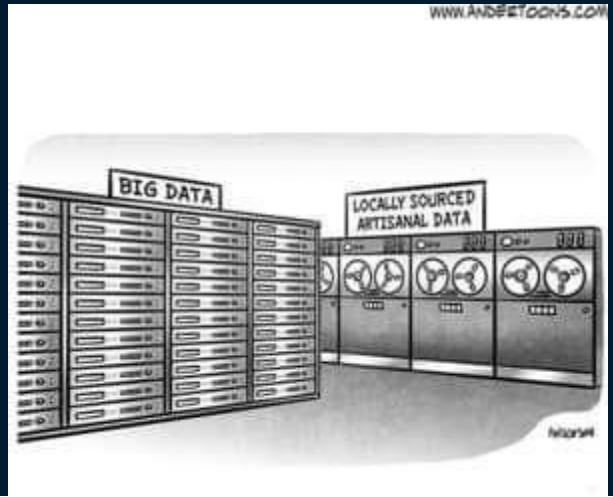
the Ngunawal & Ngambri people, the Traditional Owners of this region where we are meeting

Radio Astronomy's big data era

e.g. Australian SKA Pathfinder (ASKAP)



- Data rate (MRO → Perth):
  - 77 GB/hr – 22 TB/hr, <100 PB/yr
- Data rates & volume exceeding current processing methods

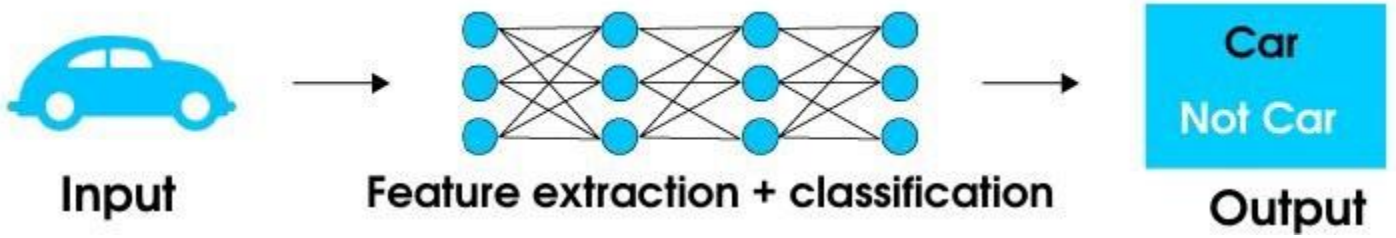


Why do we want this?

To survey the Universe deeper, faster & at higher resolution than before (both spatially & spectrally)

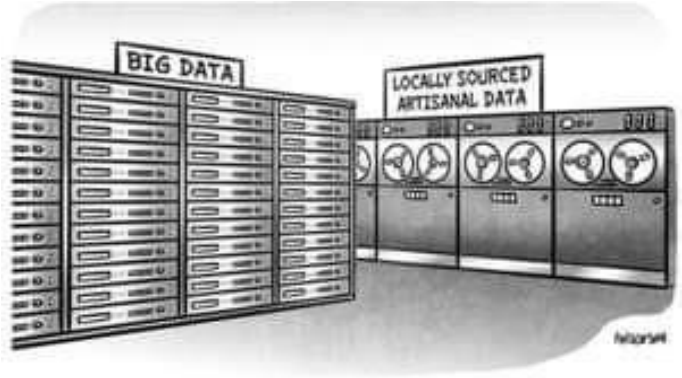
To further knowledge on the formation & evolution of galaxies throughout cosmic time

Why machine learning?



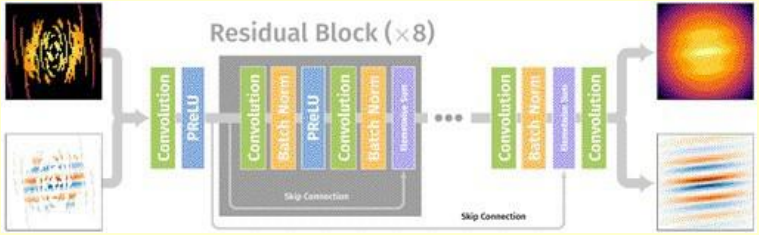
Need greater automation

(esp. where simpler algorithms are less effective)



Overview of 3 themes of applications

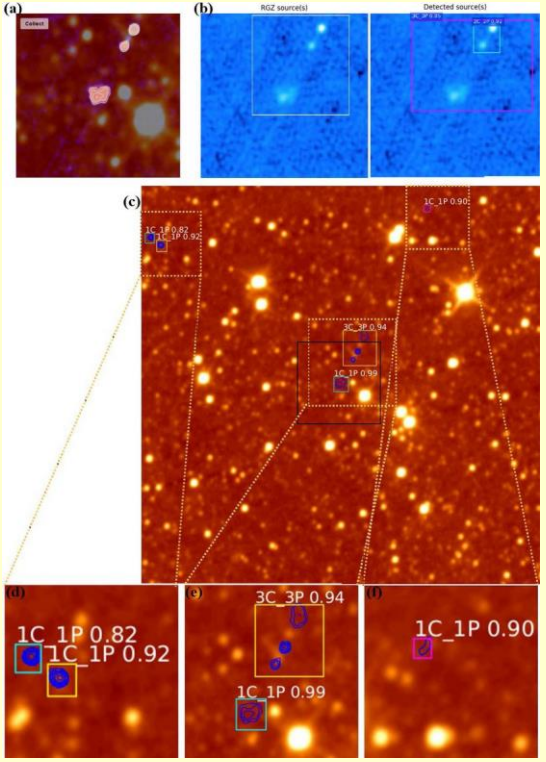
1) Processing interferometric observations



Schmidt,K.+ 2022

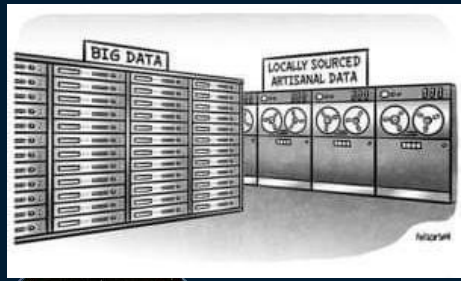
2) Radio source classification

3) Finding the unknown



Wu, OIW.+ 2019

Traditional ('artisanal') → New (AI/ML) methods



Hard Yakka

Science question(s) / problem(s)

Reduce/replace human input for automated processing

Data collecting (observations)

Data processing (calibration/ image reconstruction)

Data analysis (classification, statistics, model)

Science result(s) → publish/ profit

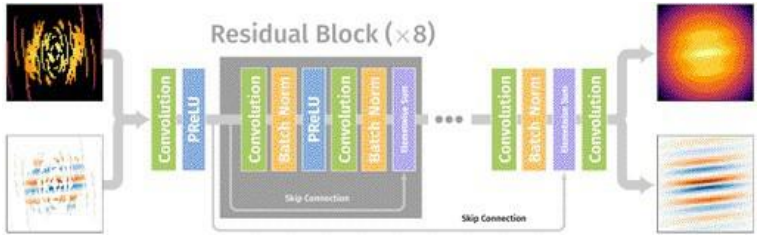
Grad student/ astronomer

No

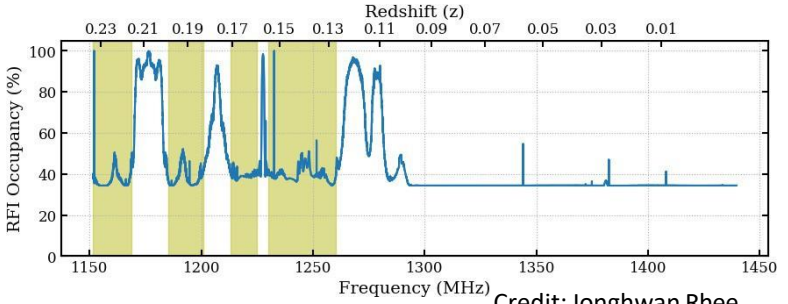
Yes

ML solutions for interferometric observations

- Reduce noise artefacts in reconstructed images by filling the gaps in the Fourier plane → bypass typically bespoked imaging decisions (Schmidt + 2022)
- RFI removal from observations using (e.g. Vinsen+ 2019; Yang+2020; Sadr+2020; Mesarcik+2022)
- Diagnose system health of large telescope arrays such as LOFAR (NL; Mesarcik + 2020)

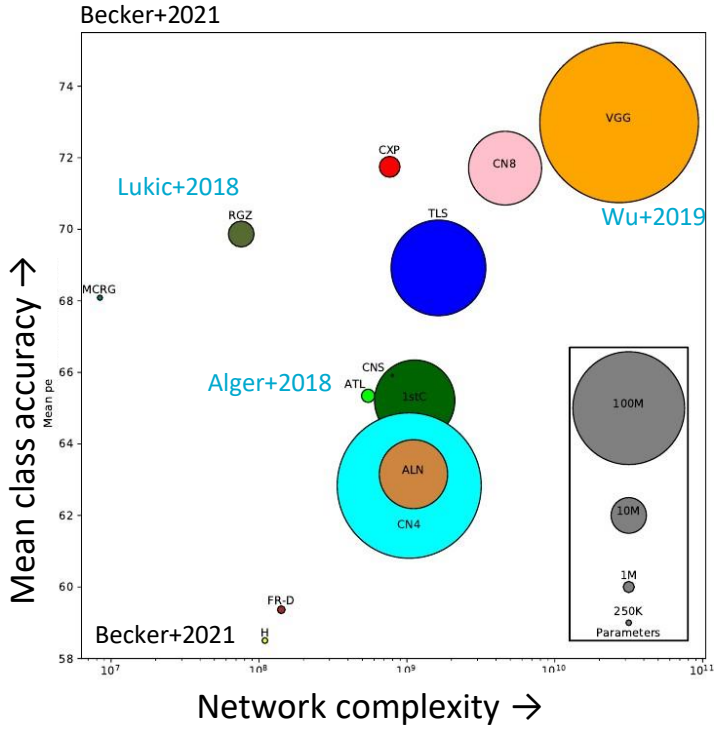


Schmidt,K.+ 2022

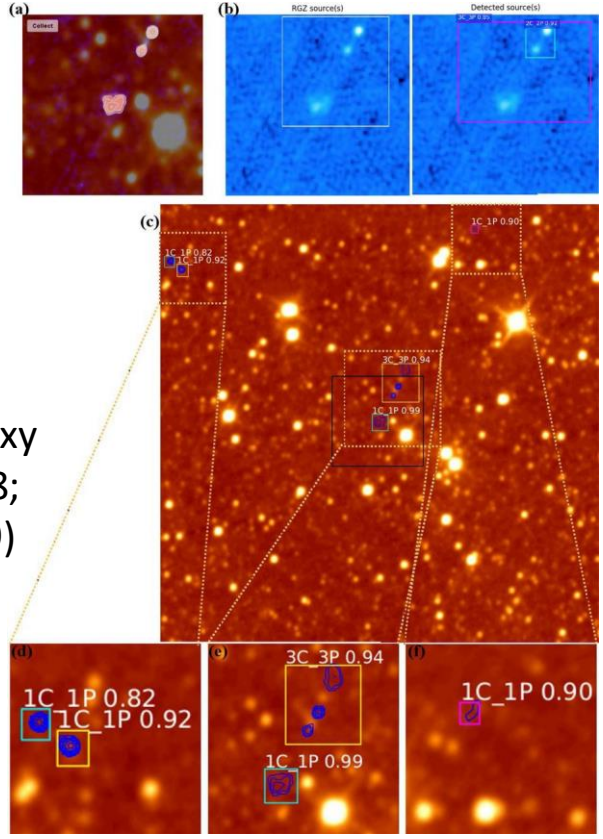


Credit: Jonghwan Rhee

Radio source classifications



- Supervised models
  - – based on Radio Galaxy Zoo labels (Lukic+2018; Alger+2018; Wu+2019)

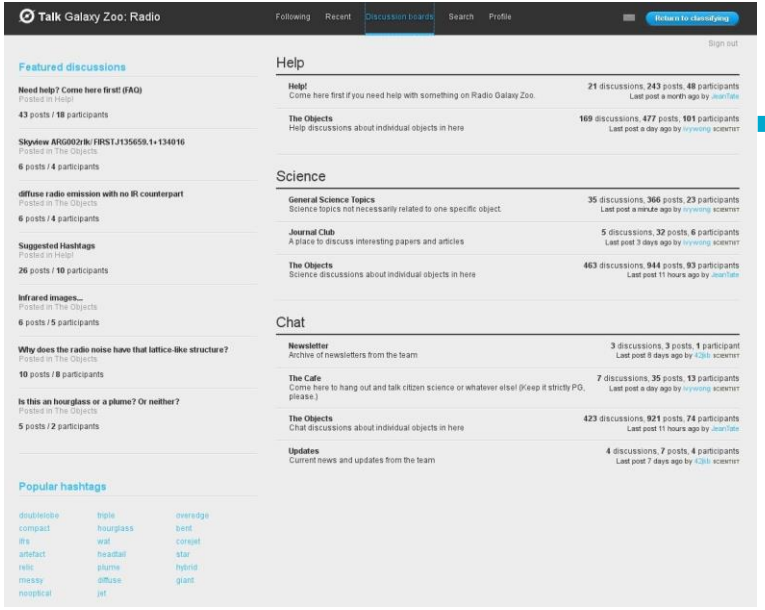


Radio Galaxy Zoo: CLARAN – a deep learning classifier for radio morphologies

Wu, OIW+2019

Multimodal classification

Using hashtags from RGZ chat forum + images

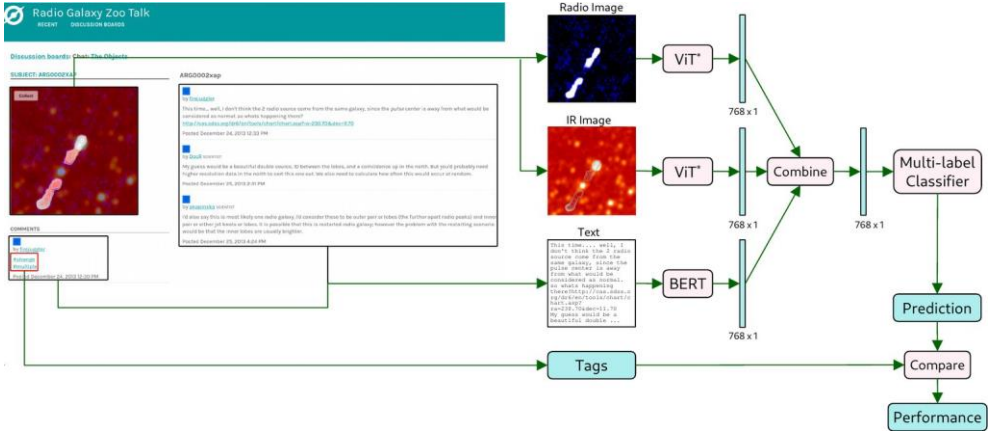


Publications of the Astronomical Society of Australia (2023), 1–12
doi:10.1017/pasa.2020.32

RESEARCH PAPER

Radio Galaxy Zoo: Tagging Radio Subjects using Text

Dawei Chen,[1,2] Vinay Kerai,[1,3] Matthew J. Alger,[4] O. Ivy Wong,[5,6] and Cheng Soon Ong[7,2]



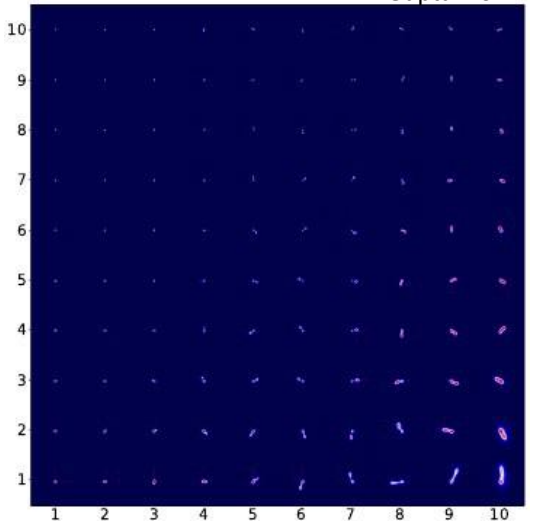
Chen + 2023 submitted to PASA

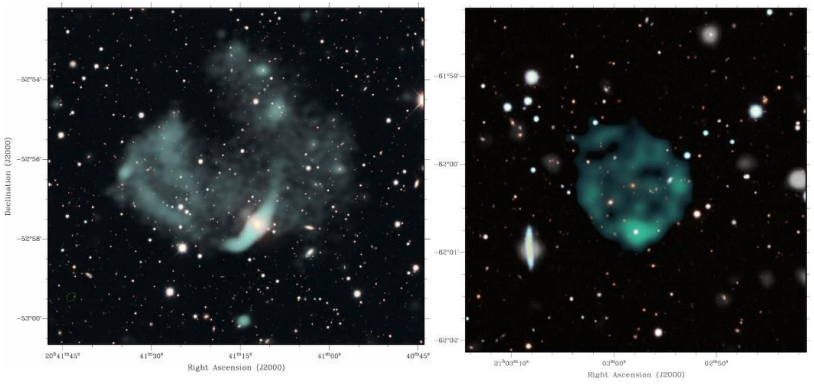
Finding the unknown

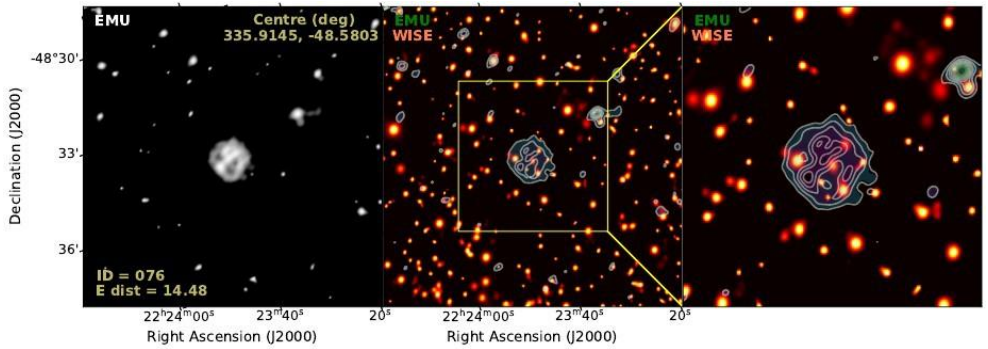
- Unsupervised models
  - Self-organising maps of images & feature vectors
  - (Galvin+2019; Ralph+2019; Gupta+2022)

- Tags & text semantics may be useful for finding unknowns from
- next-generation citizen science projects – Radio Galaxy Zoo: EMU

Gupta+2022





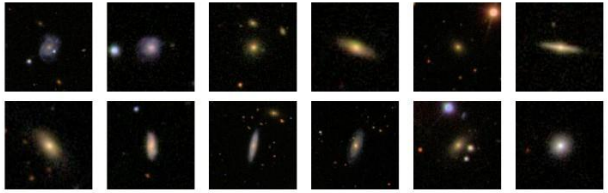


Enable discoveries with ASKAP EMU survey!

Astronomaly: active learning for anomaly detection

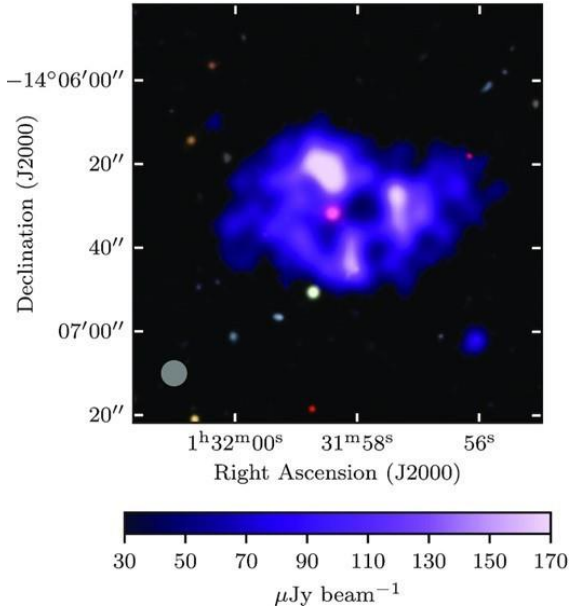
(a): ASTRONOMALY: Top twelve most anomalous objects after applying active learning, showing how most artefacts are removed.

(b): No active learning: Top twelve most anomalous objects using isolation forest (artefacts are highlighted with a red border).



(c): Random examples.

Effective for differentiating between artefacts & true anomalies



Discovery of new ORC Lochner + 2023

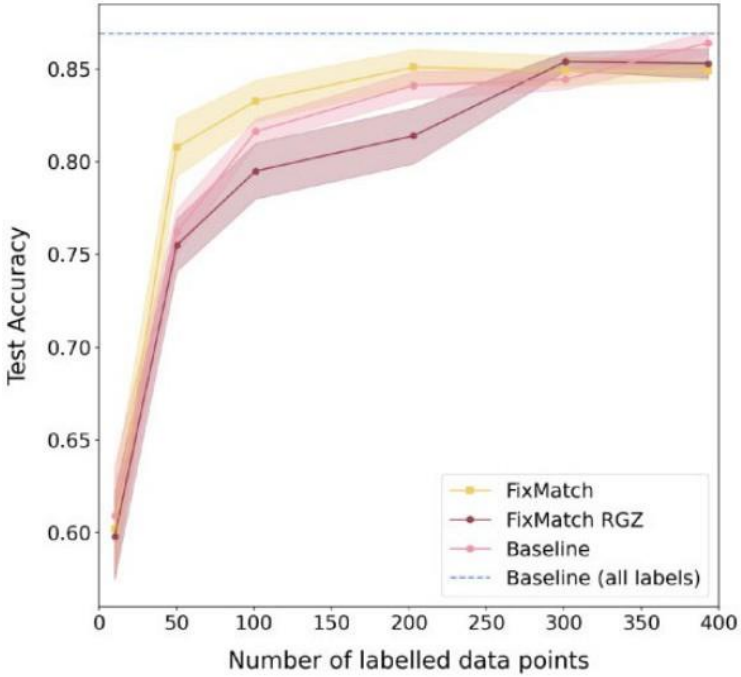
Lochner & Bassett 2020; https://github.com/MichelleLochner/astronomaly

Semi-supervised learning for source classification

Potentially leverage labelled datasets for mapping to unlabelled ones

- Example: Bowles+2022 used FixMatch SSL method
- Class imbalance in datasets (& high variance between datasets) negatively affect performance
- Overall, there is insufficient robustness to replace train-test cycle

Bowles+2022

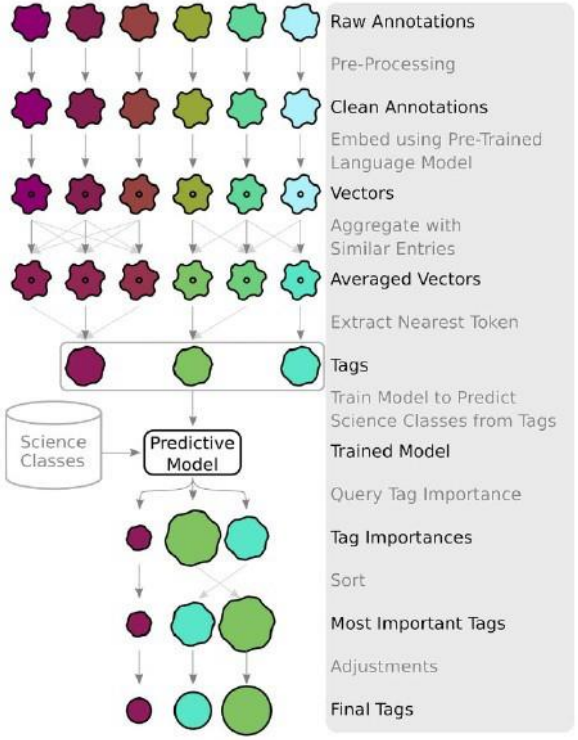


(a) Accuracy on the test set.

Using NLP to derive a taxonomy of tags

Derive a taxonomy of tags from plain English descriptions of source morphologies (free of restrictions from obfuscating technical terminology)

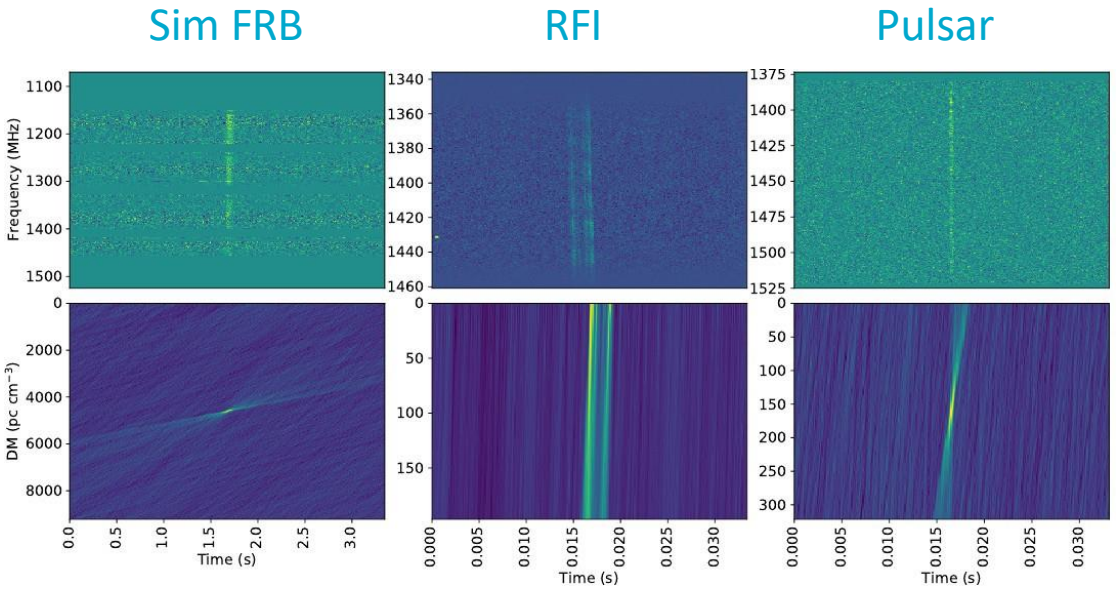
- In preparation for RGZ: EMU
- Input set of ~8500 plain English annotations
- Simple framework which produced ~22 unique semantic tags



Bowles+2023

Radio transient classification

e.g. Fast Extragalactic Transient Candidate Hunter (FETCH): DL classification of RFI vs true transients



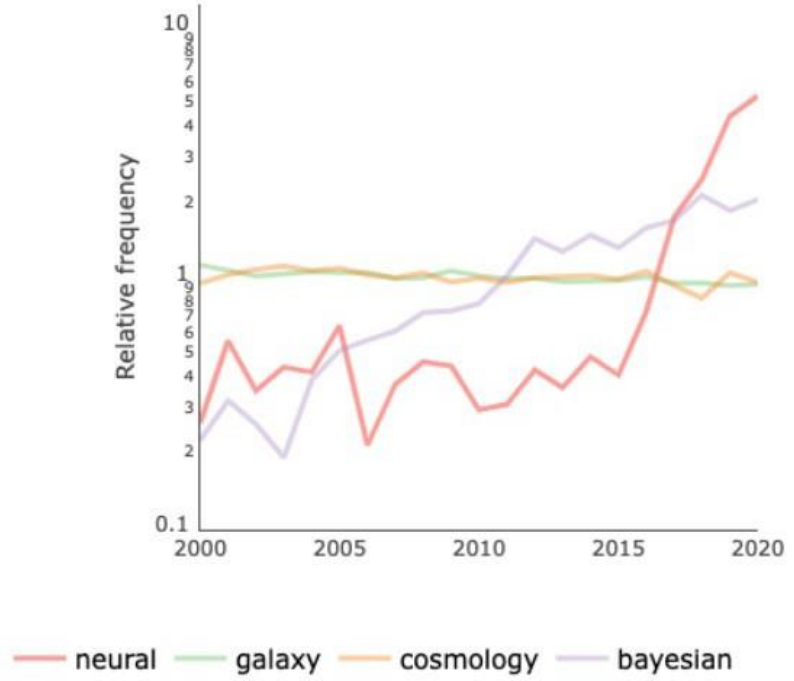
FT Model	Val Acc (%)
VGG19 (4)	99.78
VGG16 (4)	99.40
DenseNet169 (11)	95.40
DenseNet201 (7)	94.05
DenseNet121 (4)	88.23

DMT Model	Val Acc (%)
VGG16 (2)	99.92
Xception (21)	99.87
VGG19 (0)	99.73
InceptionV3 (31)	99.46
InceptionResNetV2 (34)	99.35

Agarwal, D+2020; https://github.com/devanshkv/fetch

ML in Radio Astronomy is here to stay

- ML applications in radio astronomy has matured greatly in the last 5 years
- End-to-end data flow from telescope – science results benefit from ML approaches
- Large diversity of methods:
  - unsupervised – supervised, simpler – deeper learning methods → *dependent on problem being solved*
- ML applications necessary for maximising science from “big data” era radio astronomy



Huertas-Company & Lanusse 2022

Thank you

CSIRO Space & Astronomy
Ivy Wong
ivy.wong@csiro.au