



# Machine learning applications & outlook for radio astronomy

O. Ivy Wong (CSIRO) | 15 June 2023

Australasian Leadership Computing Symposium, Canberra

Australia's National Science Agency





# 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*

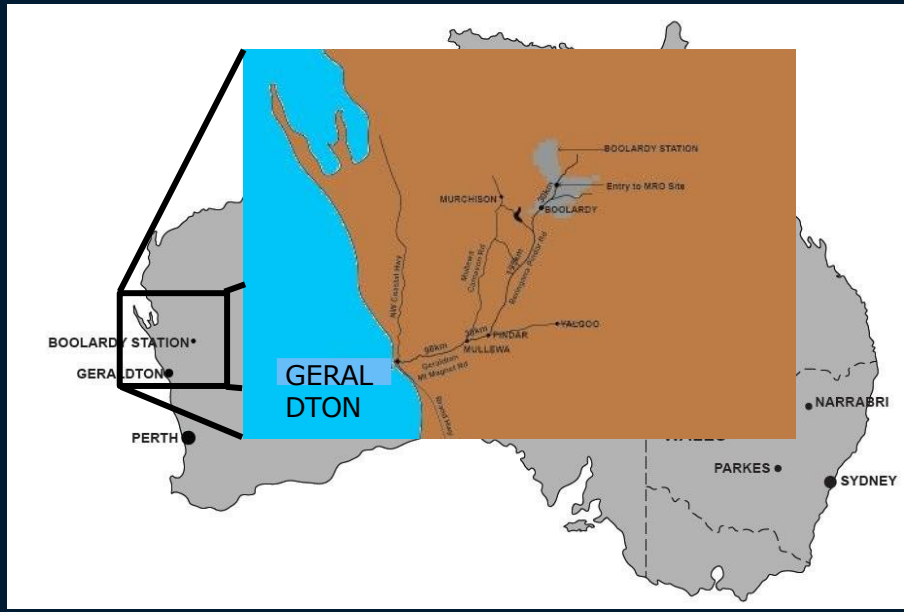
*the Ngunawal & Ngambri people, the Traditional  
Owners of this region where we are meeting*



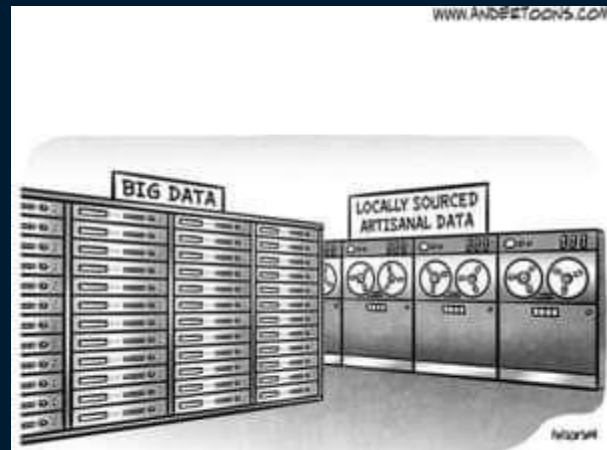
# Radio Astronomy's big data era



## 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





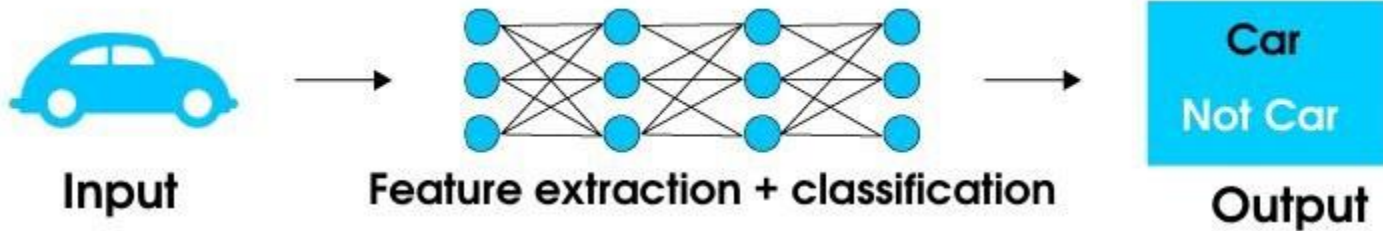
# Why do we want this?

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

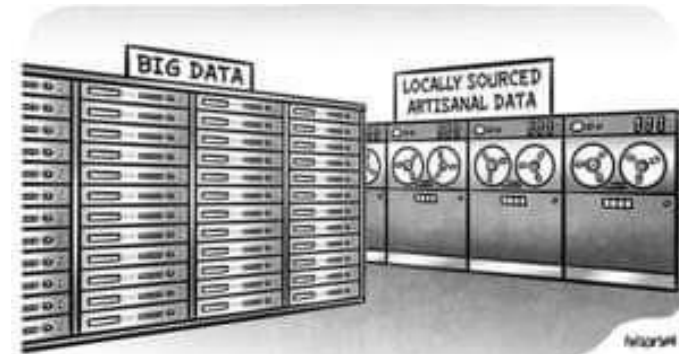
To further knowledge on the formation & evolution of galaxies throughout cosmic time



# Why machine learning?

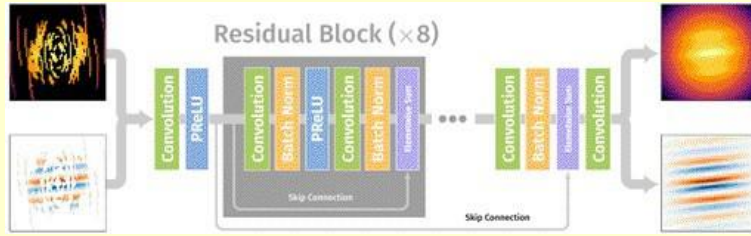


**Need greater automation**  
*(esp. where simpler algorithms are less effective)*



# Overview of 3 themes of applications

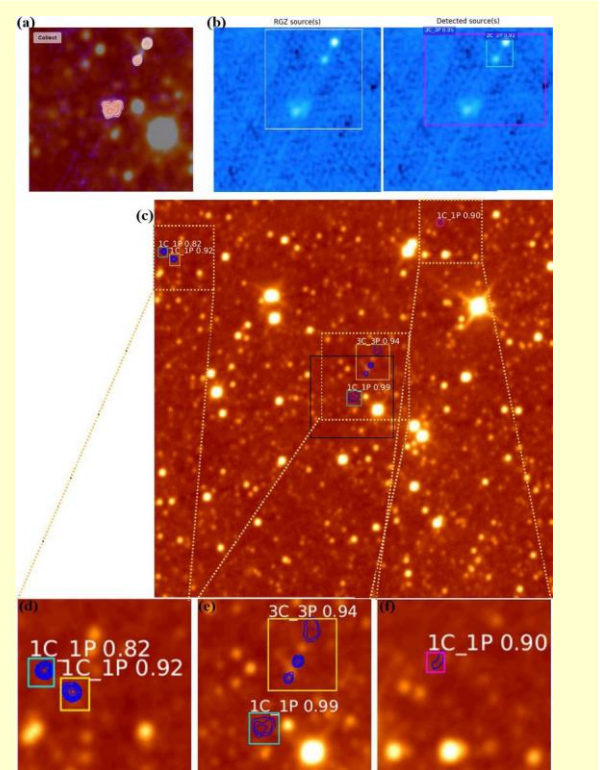
## 1) Processing interferometric observations



Schmidt, K.+  
2022

## 2) Radio source classification

## 3) Finding the unknown





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

**Hard Yakka**

Science question(s) / problem(s)

Data collecting (observations)

Data processing (calibration/ image reconstruction)

Data analysis (classification, statistics, model)

Science result(s) → publish/ profit

Reduce/replace human input for automated processing

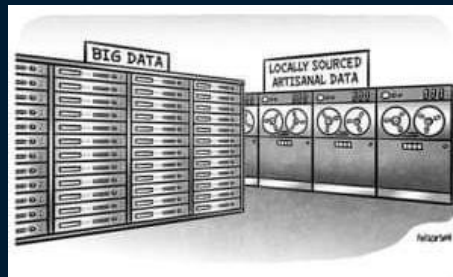


Grad student/ astronomer



No

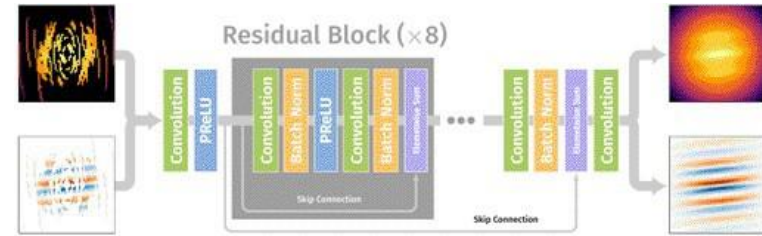
Yes



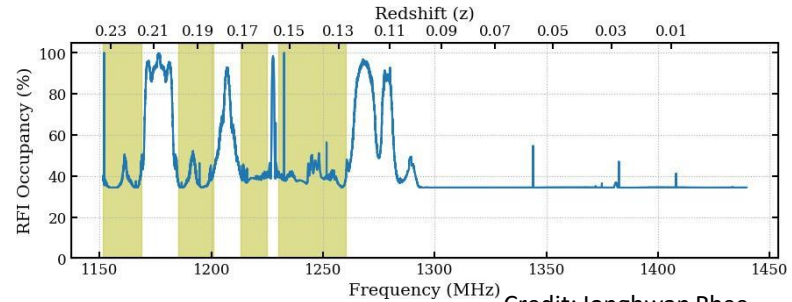


# ML solutions for interferometric observations

- Reduce noise artefacts in reconstructed images by filling the gaps in the Fourier plane → bypass typically bespoke 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

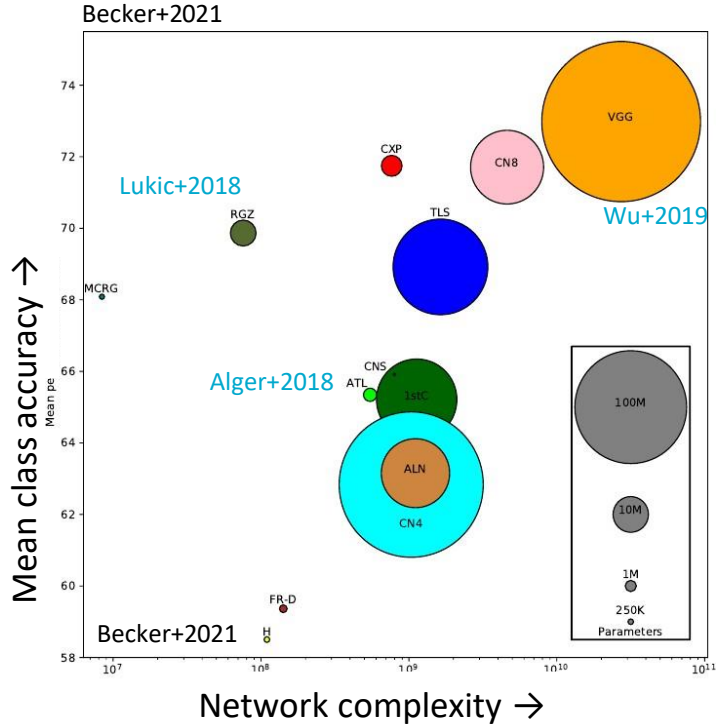


Credit: Jonghwan Rhee

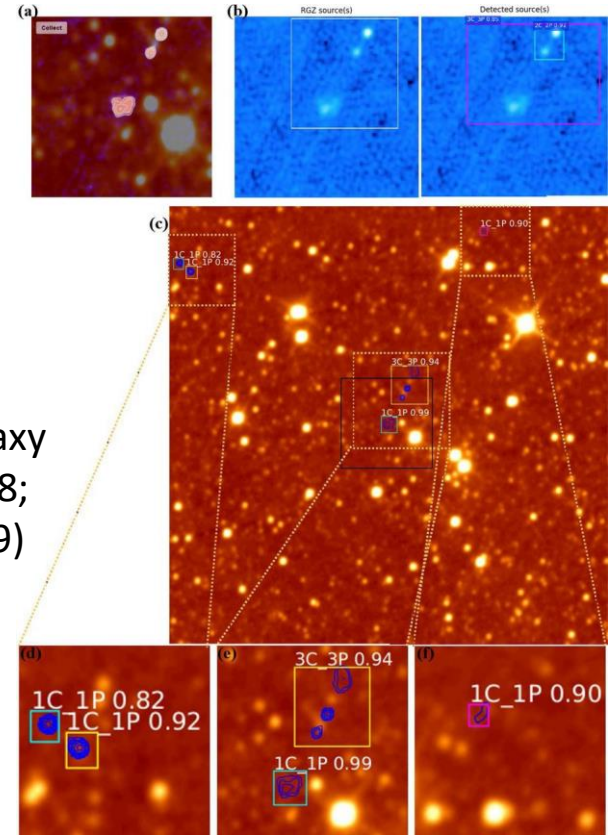




# Radio source classifications



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





# Multimodal classification

## Using hashtags from RGZ chat forum + images

**Talk Galaxy Zoo: Radio**

Following Recent **Discussion boards** Search Profile Follow to classify your Sign out

**Featured discussions**

- Need help? Come here first! (FAQ)**  
Posted in Help  
43 posts / 18 participants
- Skyview ARG02Rb/FRSTJ135659.1-134016**  
Posted in The Objects  
6 posts / 4 participants
- diffuse radio emission with no IR counterpart**  
Posted in The Objects  
6 posts / 4 participants
- Suggested Hashtags**  
Posted in Help  
26 posts / 10 participants
- Infrared images...**  
Posted in The Objects  
6 posts / 5 participants
- Why does the radio noise have that lattice-like structure?**  
Posted in The Objects  
10 posts / 8 participants
- Is this an hourglass or a plume? Or neither?**  
Posted in The Objects  
5 posts / 2 participants

**Popular hashtags**

doublelobe	triple	overedge
compact	hourglass	base
IR	wat	current
apofact	headtail	star
radio	plume	hybrid
messy	diffuse	giant
neopical	jet	

**Help**

- Help**  
Come here first if you need help with something on Radio Galaxy Zoo.  
21 discussions, 243 posts, 48 participants  
Last post 6 months ago by [Jarlaine](#)
- The Objects**  
Help discussions about individual objects in here  
169 discussions, 477 posts, 101 participants  
Last post 6 days ago by [Vinyong Kerai](#)

**Science**

- General Science Topics**  
Science topics not necessarily related to one specific object.  
35 discussions, 366 posts, 23 participants  
Last post 9 months ago by [Vinyong Kerai](#)
- Journal Club**  
A place to discuss interesting papers and articles  
5 discussions, 32 posts, 6 participants  
Last post 3 days ago by [Vinyong Kerai](#)
- The Objects**  
Science discussions about individual objects in here  
463 discussions, 944 posts, 93 participants  
Last post 11 hours ago by [Jarlaine](#)

**Chat**

- Newsletter**  
Archive of newsletters from the team  
3 discussions, 3 posts, 1 participant  
Last post 9 days ago by [Jarlaine](#)
- The Cafe**  
Come here to hang out and talk citizen science or whatever else! (Keep it strictly PG, please.)  
7 discussions, 35 posts, 13 participants  
Last post 4 days ago by [Vinyong Kerai](#)
- The Objects**  
Chat discussions about individual objects in here  
423 discussions, 821 posts, 74 participants  
Last post 11 hours ago by [Jarlaine](#)
- Updates**  
Current news and updates from the team  
4 discussions, 7 posts, 4 participants  
Last post 7 days ago by [Jarlaine](#)

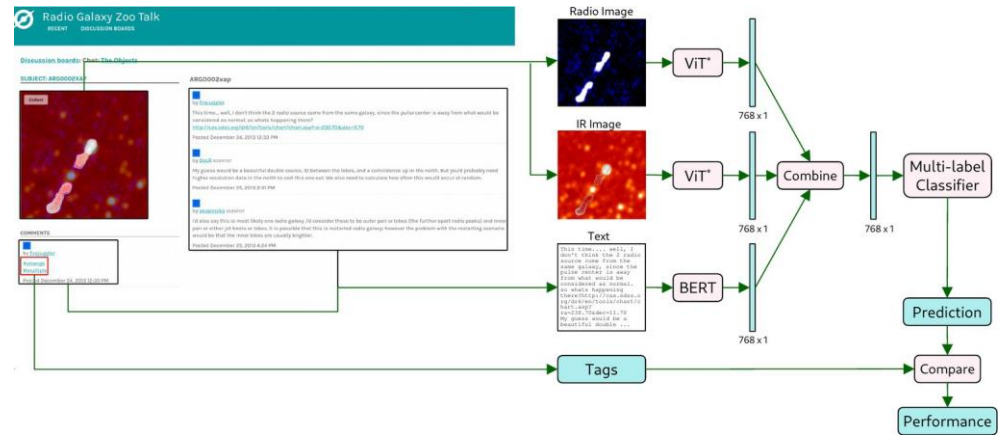


Publications of the Astronomical Society of Australia (2023), 1–12  
doi:10.1017/pasa.2020.32

### RESEARCH PAPER

## Radio Galaxy Zoo: Tagging Radio Subjects using Text

Dawei Chen,<sup>1,2</sup> Vinay Kerai,<sup>1,3</sup> Matthew J. Alger,<sup>4</sup> O. Ivy Wong,<sup>5,6</sup> and Cheng Soon Ong<sup>7,2</sup>

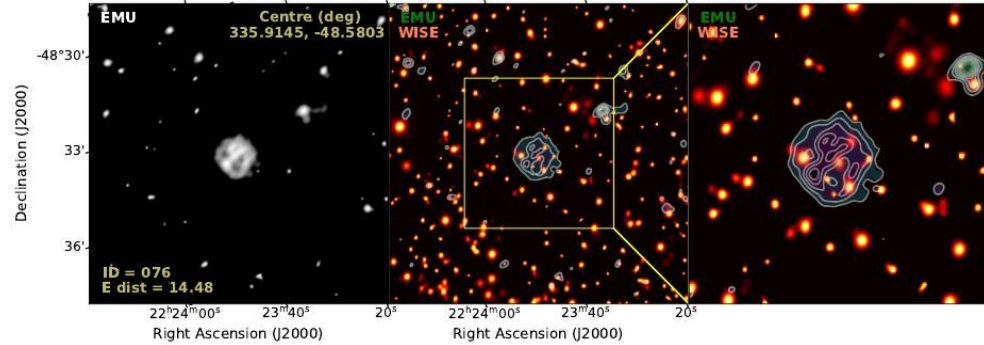
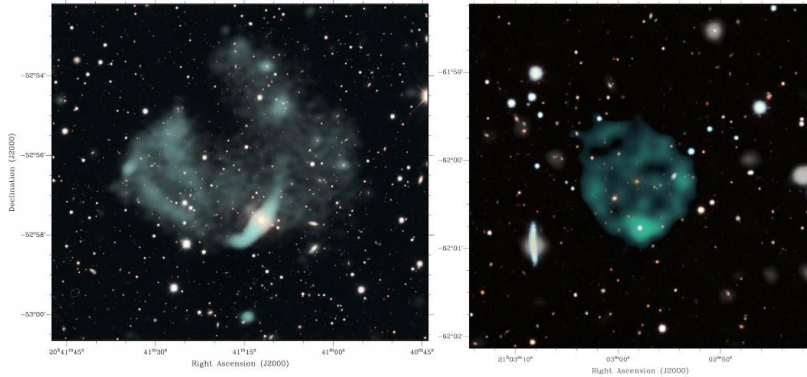
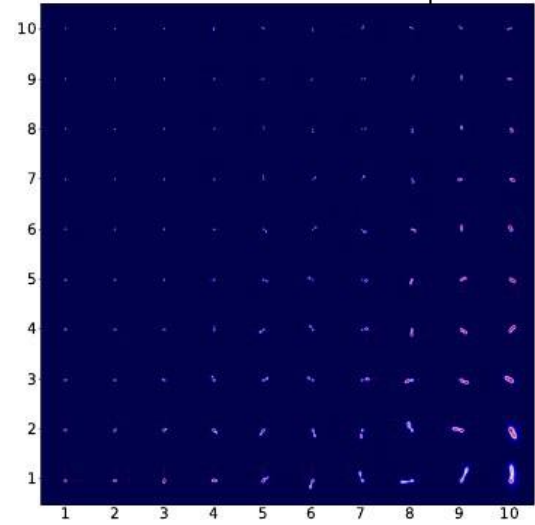




# Finding the unknown

Gupta+2022

- 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





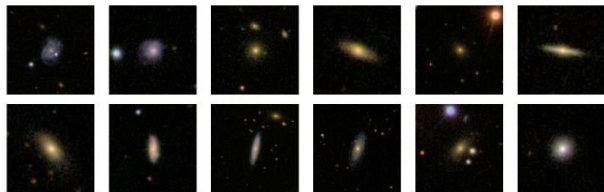
# Astronomy: active learning for anomaly detection



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

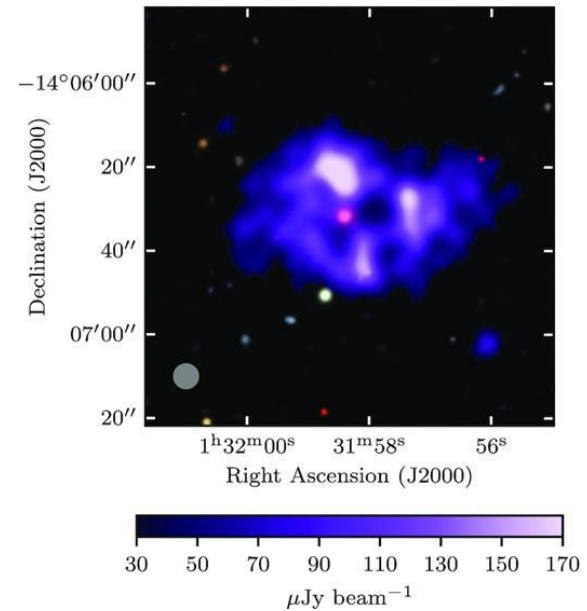


(b): No active learning: Top twelve most anomalous objects using isolation forest (artefacts are highlighted with a red border).



(c): Random examples.

Effective for differentiating between artefacts & true anomalies



*Discovery of new ORC*

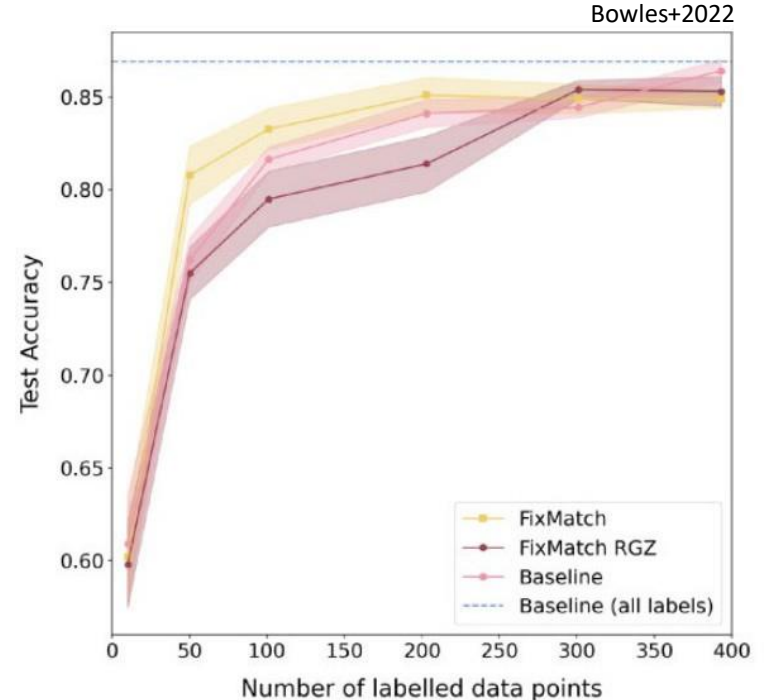
Lochner + 2023



# 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



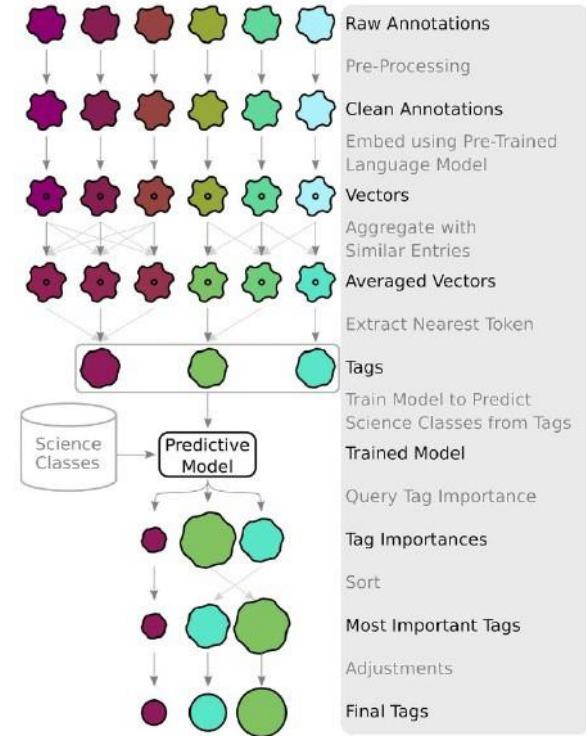
(a) Accuracy on the test set.



# 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)***

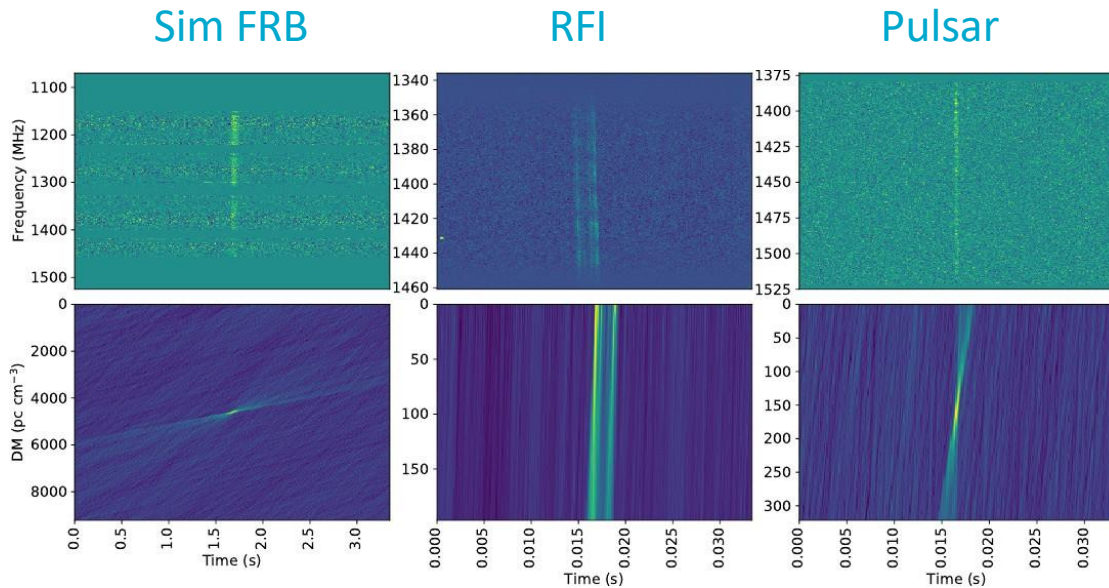
- In preparation for RGZ: EMU
- Input set of ~8500 plain English annotations
- Simple framework which produced ~22 unique semantic tags





# 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

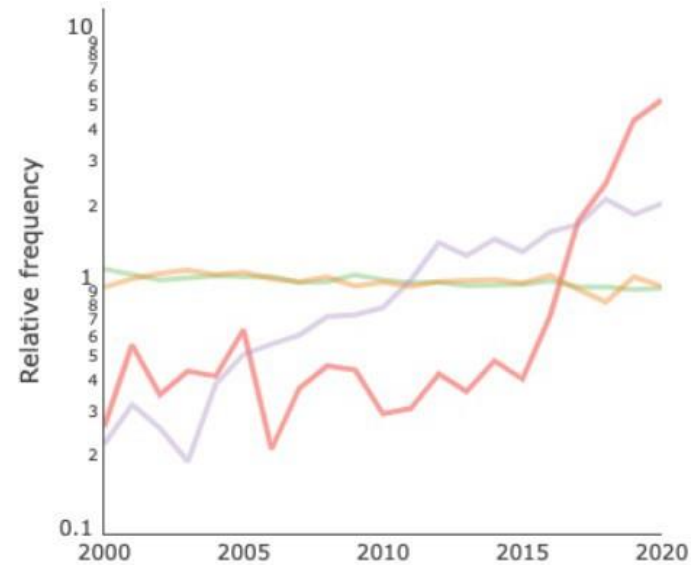
---

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



# 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



— neural — galaxy — cosmology — bayesian

Huertas-Company & Lanusse 2022





# Thank you

CSIRO Space & Astronomy

Ivy Wong

[ivy.wong@csiro.au](mailto:ivy.wong@csiro.au)

Australia's National Science Agency

