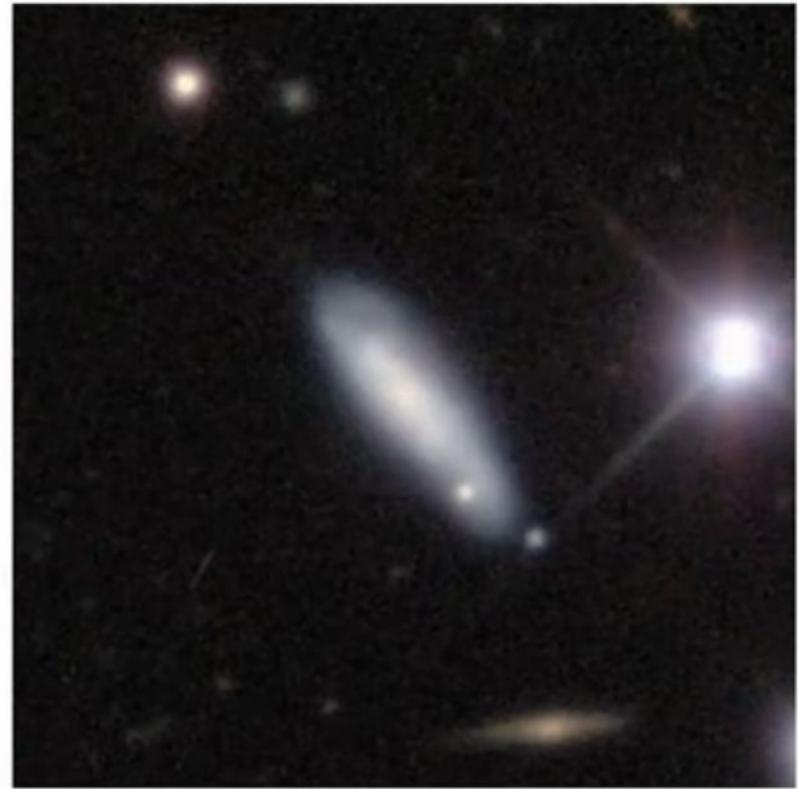
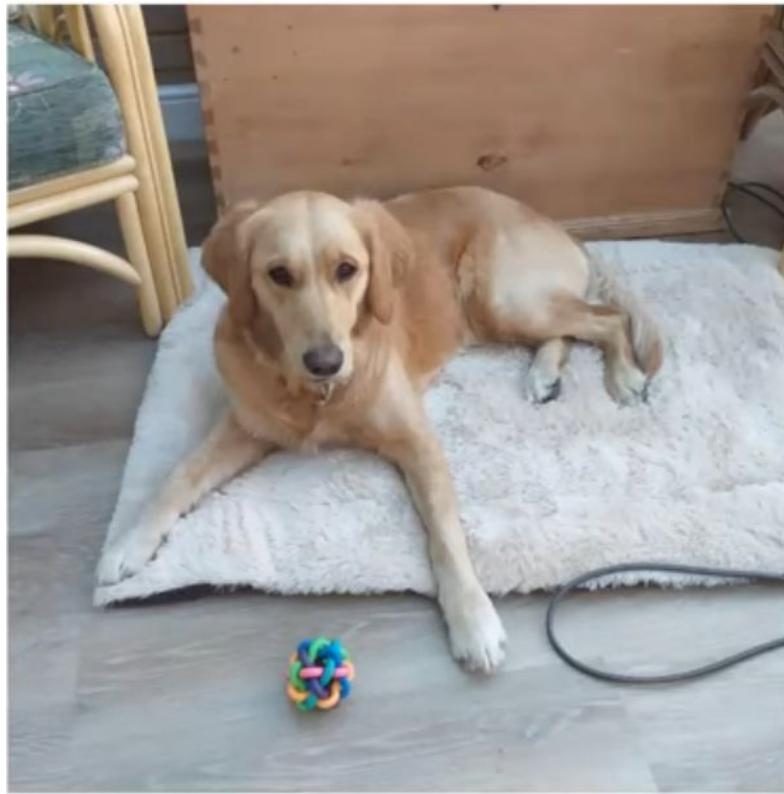

Finding Radio Bursts Fast

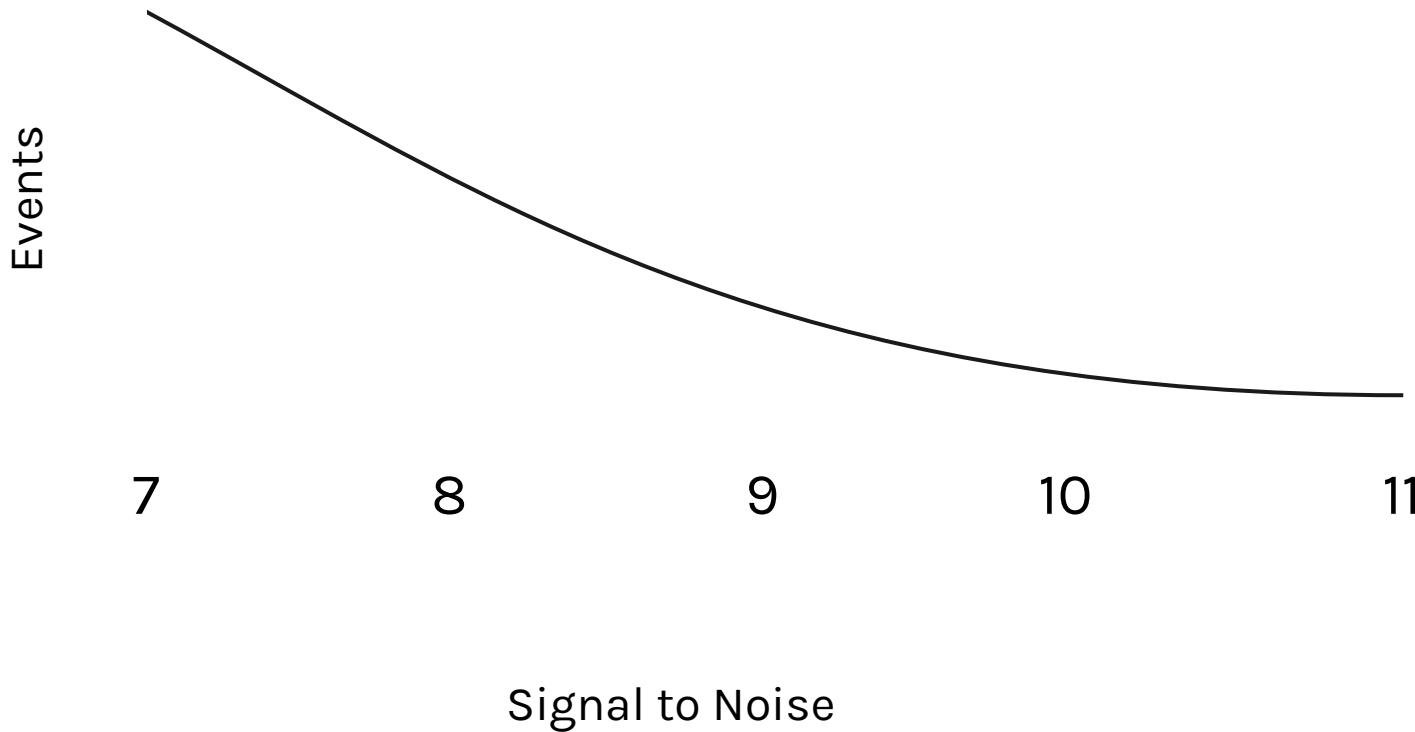
Mike Walmsley, with CHIME and Galaxy Zoo collaborations

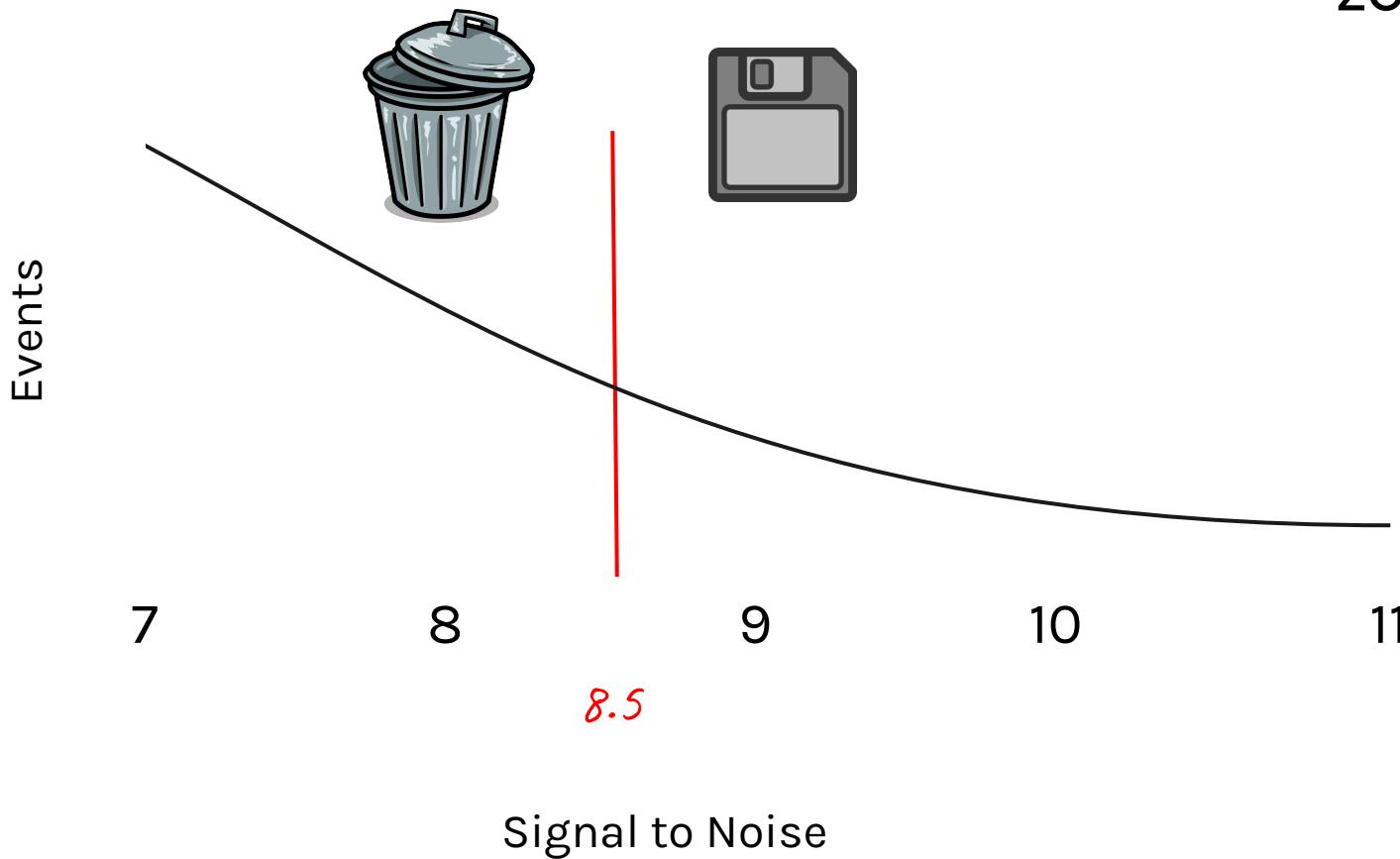
University of ~~Oxford~~ Manchester! Thanks Anna!

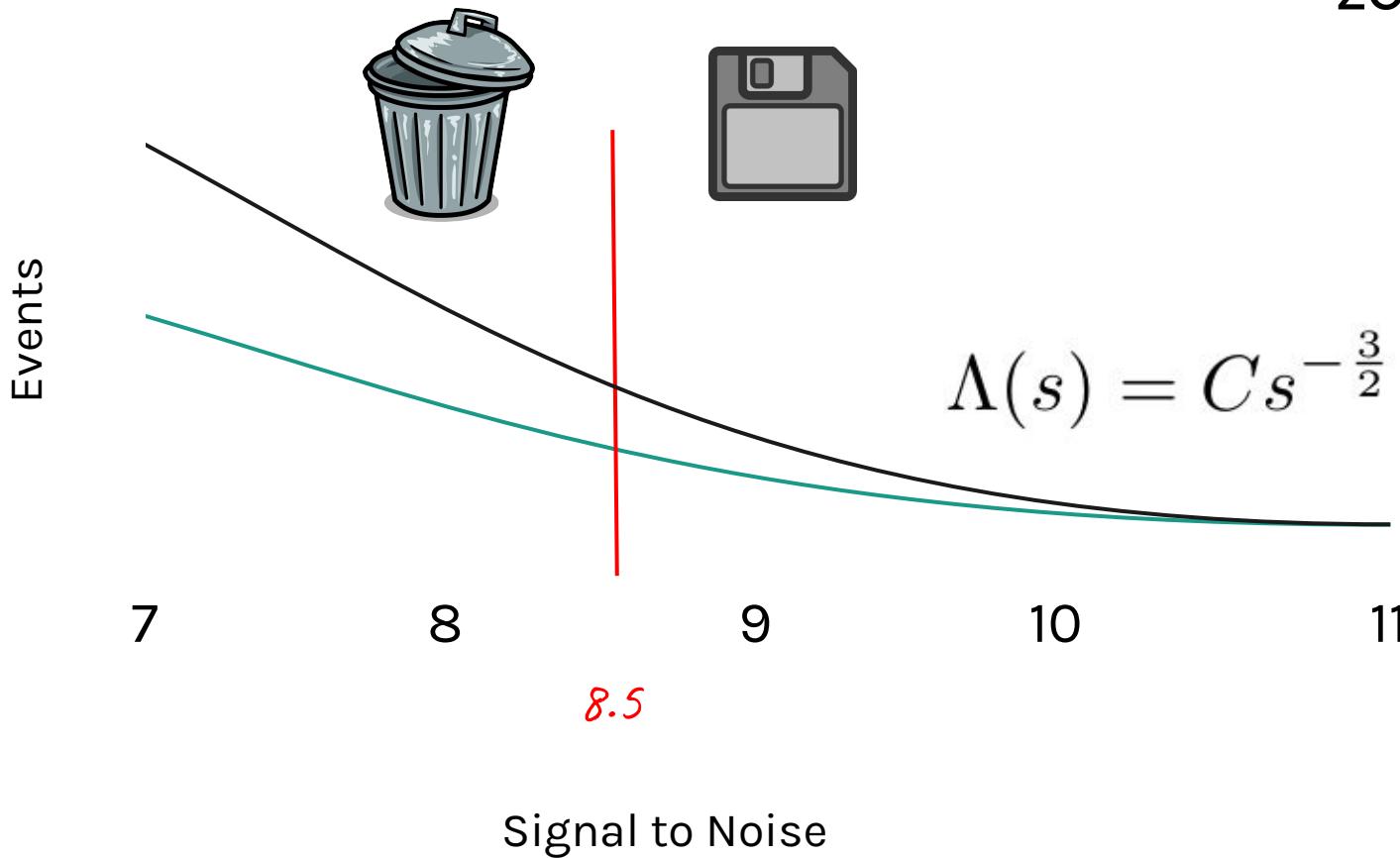
Why do we need an ML Club?

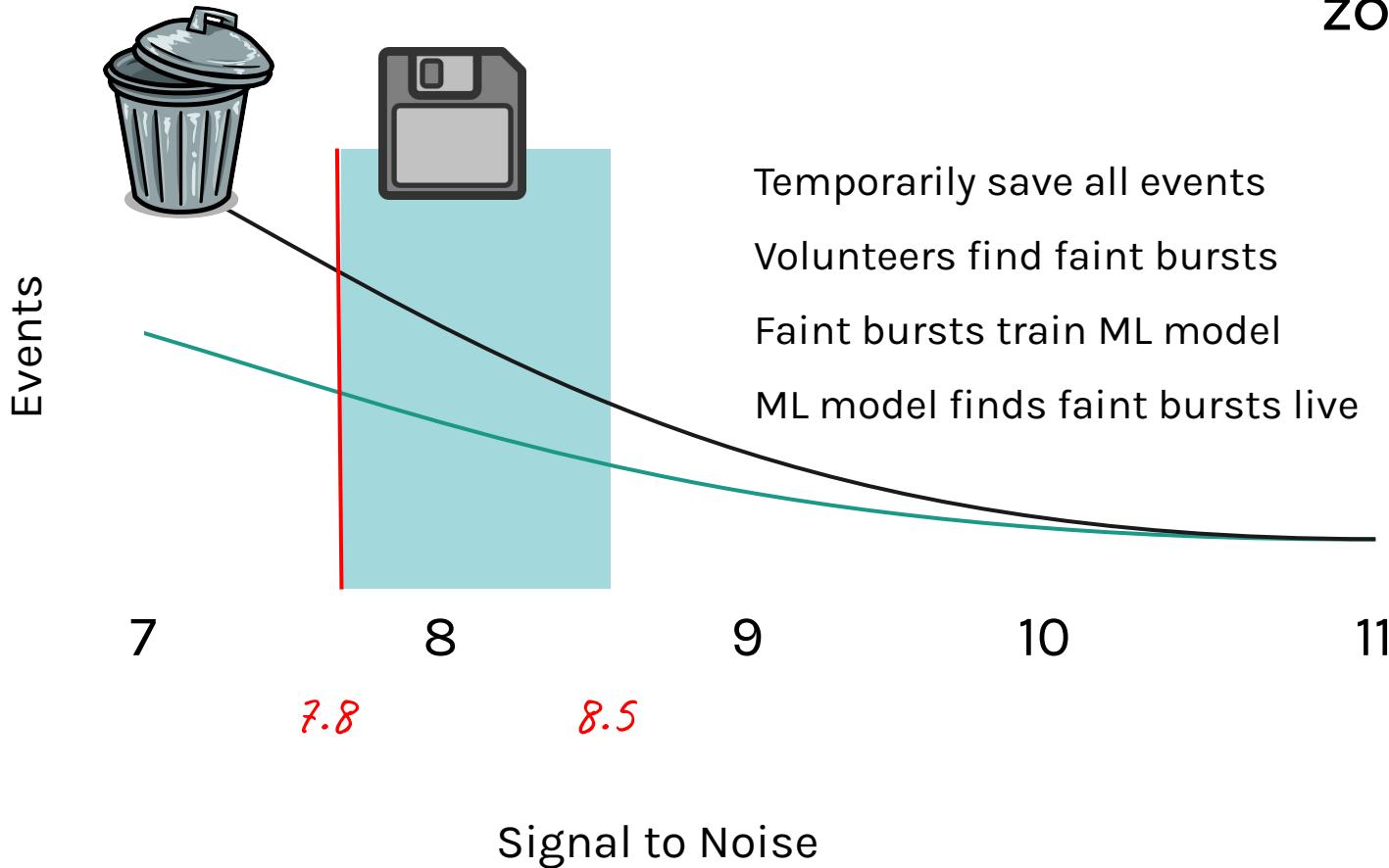


- 1. CHIME FRB: Shortcuts**
- 2. Galaxy Zoo: Uncertainty and Information**
- 3. Now what?**









The screenshot shows the landing page for the 'Bursts from Space' project. At the top, there's a navigation bar with a user icon, 'UNDER REVIEW', and the project name 'Bursts from Space'. Below the navigation are links for 'ABOUT', 'CLASSIFY', 'TALK', 'COLLECT', 'RECENTS', and 'LAB'. A blue banner at the top says 'Beta complete! Any more feedback? <https://forms.gle/TkB7oUVLN9sMkvx7>'. The main content area features a dark background with a starry galaxy image. Centered text reads 'Hunt mysterious bursts from space!'. Below this are two buttons: 'Learn more' and 'Get started'. At the bottom, a teal footer bar contains the text: 'This project has been built using the Zooniverse Project Builder but is not yet an official Zooniverse project. Queries and issues relating to this project directed at the Zooniverse Team may not receive any response.'

The screenshot shows the classification interface for the 'Bursts from Space' project. At the top right are 'TASK' and 'TUTORIAL' buttons. The main area features a spectrogram titled 'Is this burst from space, or humans?'. The y-axis is 'Frequency (MHz)' from 400 to 800, and the x-axis is 'Time (ms) (dummy)' from 0 to 250. The spectrogram shows a dense field of blue dots with several prominent horizontal bands of higher intensity. To the right of the spectrogram is a 'FIELD GUIDE' section with the text: 'Space bursts are smooth, brief, and cover many frequencies, while human bursts are messy and often only at a few specific frequencies.' Below this are two buttons: 'Space' and 'Humans'. Further down are 'NEED SOME HELP WITH THIS TASK?' buttons for 'Done & Talk' and 'Done', along with a settings gear icon. At the bottom is a 'SWITCH TO DARK THEME' link.

<http://bit.ly/burstsfromspace>

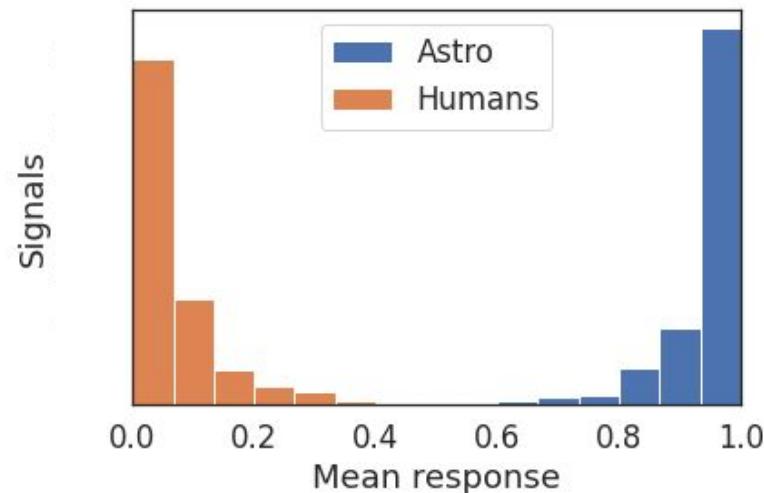
Beta Results - Performance

13,000 classifications from 116 volunteers

- Average of more than 100 per person

99.3% agreement with CHIME experts when asking 15+ volunteers

- 99.1% with 10 volunteers
- 98.6% with 5 volunteers



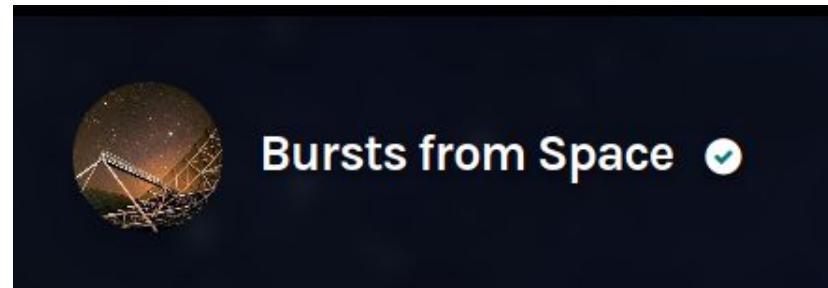
Project Launched - Oct. 30th 2020

New Data Weekly

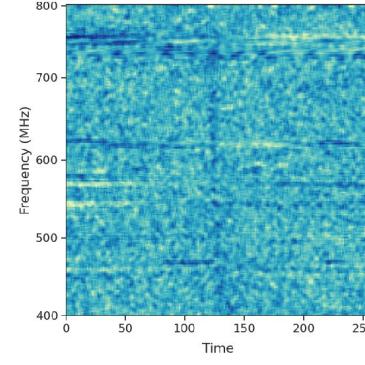
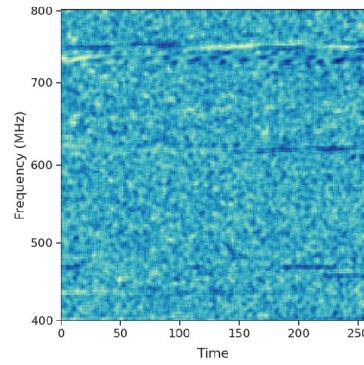
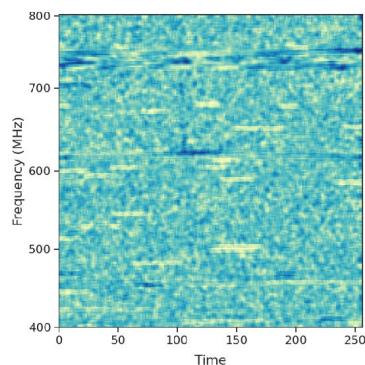
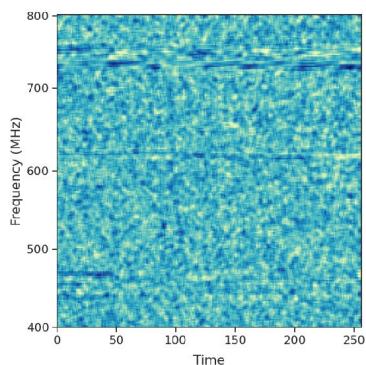
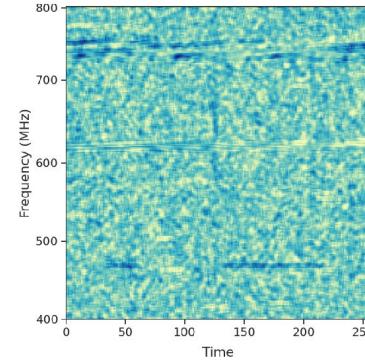
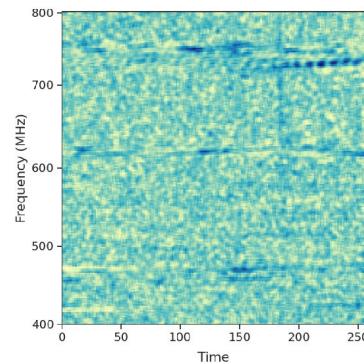
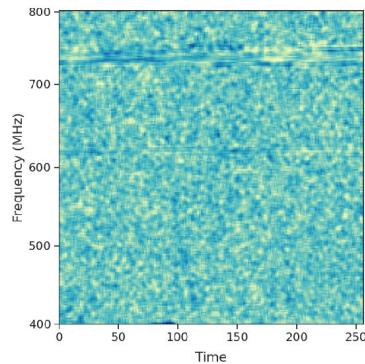
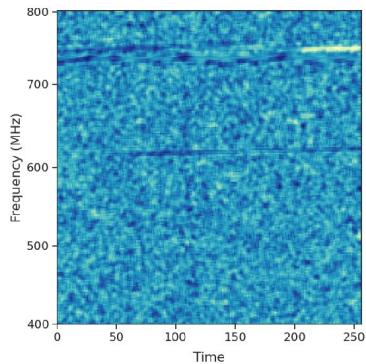
19,689 Candidates

304,071 Classifications

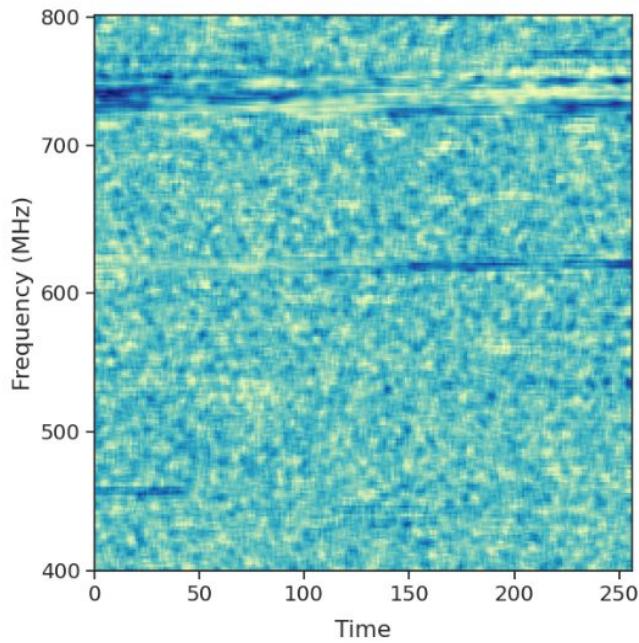
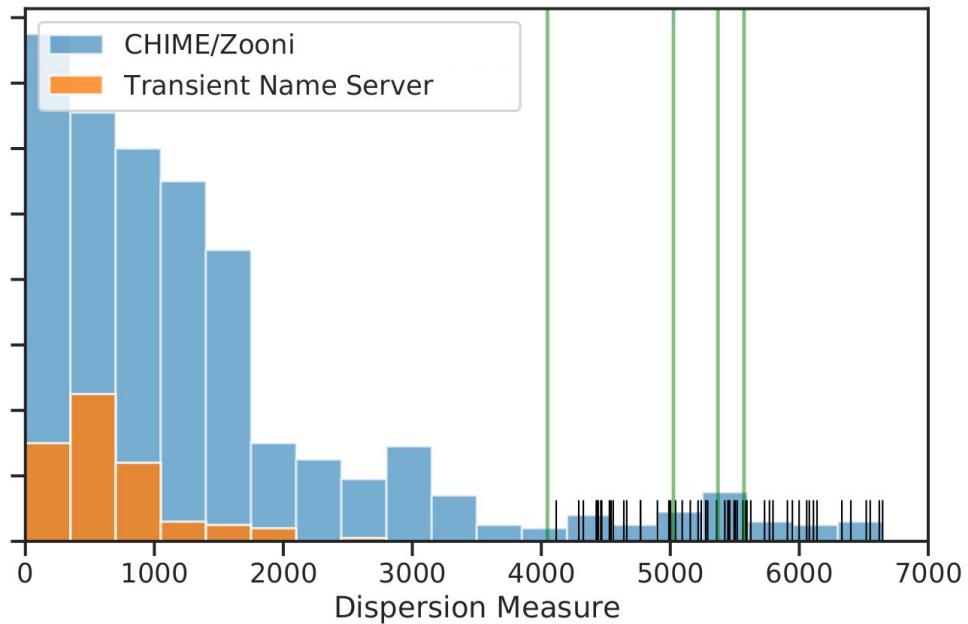
Many hundreds of FRB-like signals



bit.ly/burstsfromspace

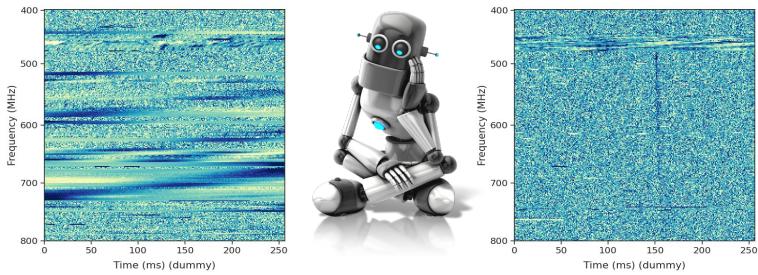


Num. Published Bursts

Space fraction: 0.80, DM: 5369 pc cm⁻³

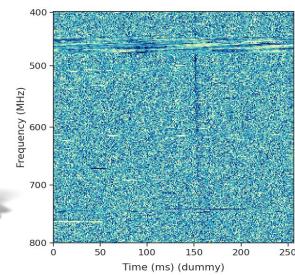
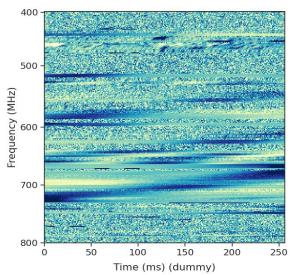
Can ML identify faint FRBs?

How would you classify?

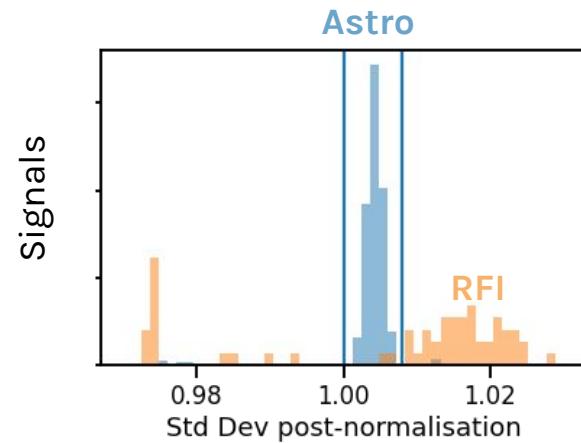


CHIME's classifier - **99% accurate**

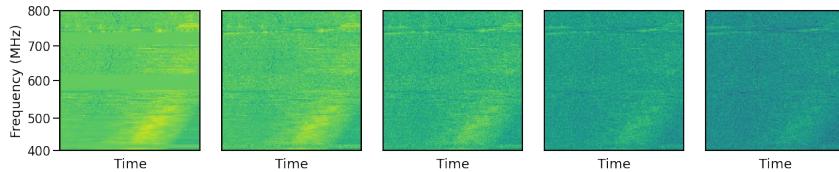
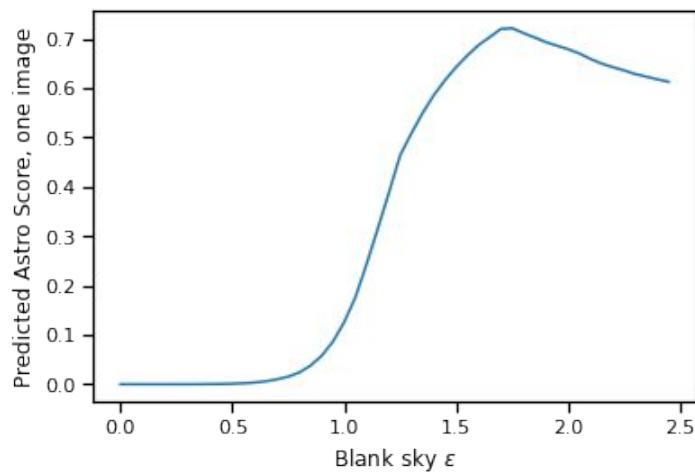
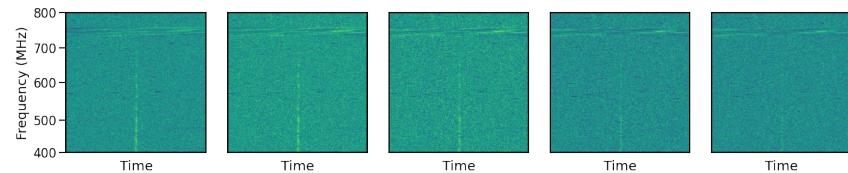
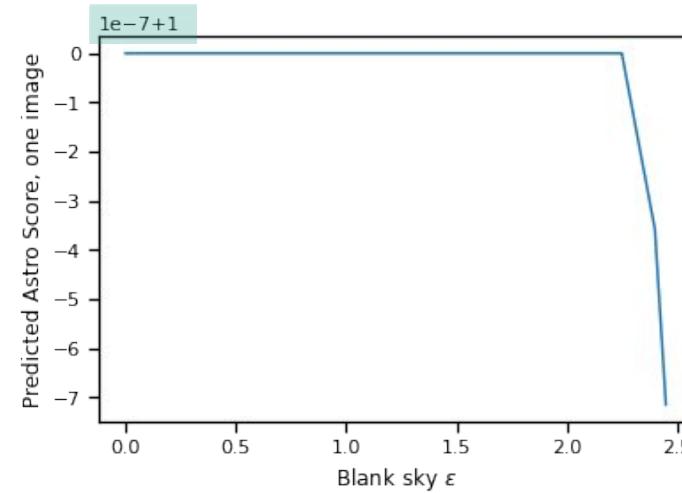
How would you classify?

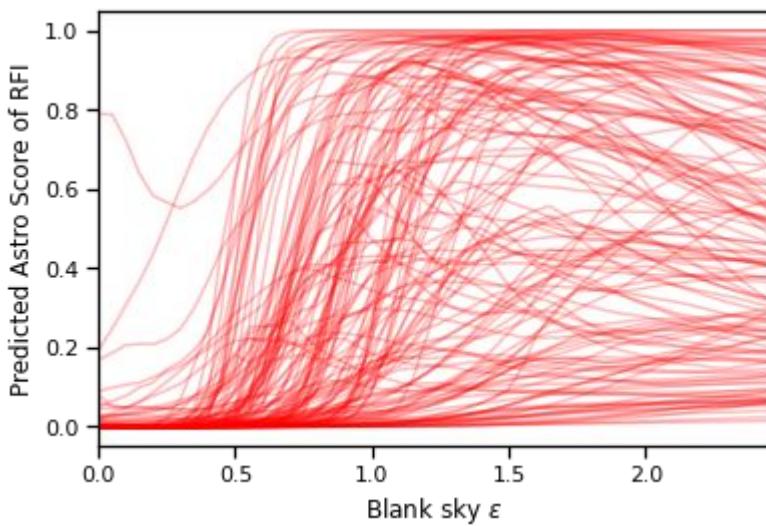
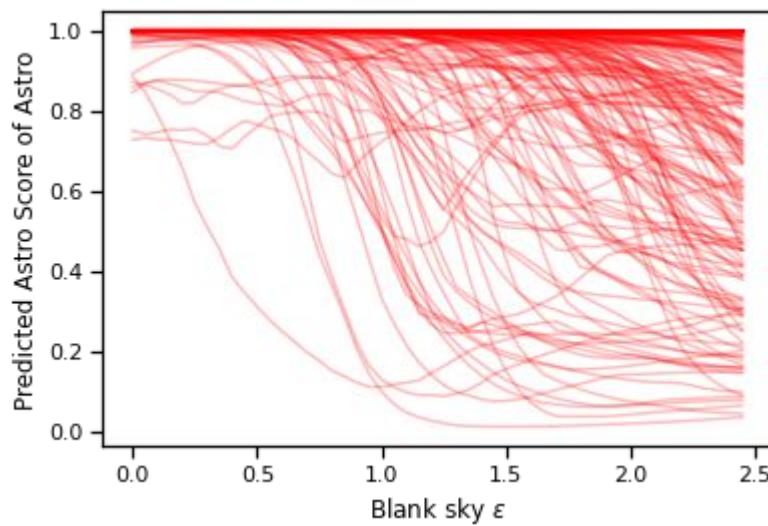


CHIME's classifier - **99% accurate**



Standard deviation **only** - **93% accurate**

RFI + ε Blank SkyAstro + ε Blank Sky

RFI + ε Blank SkyAstro + ε Blank Sky

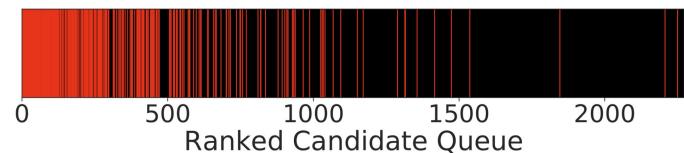
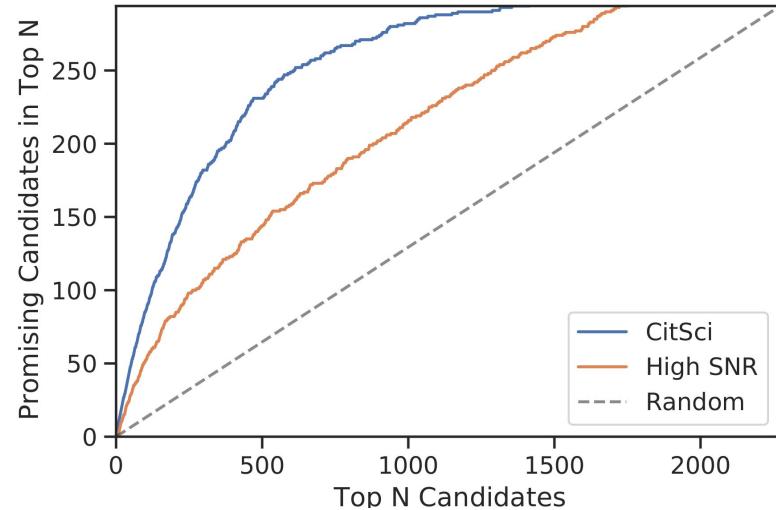
Retraining

Classifier calls faint signals astro

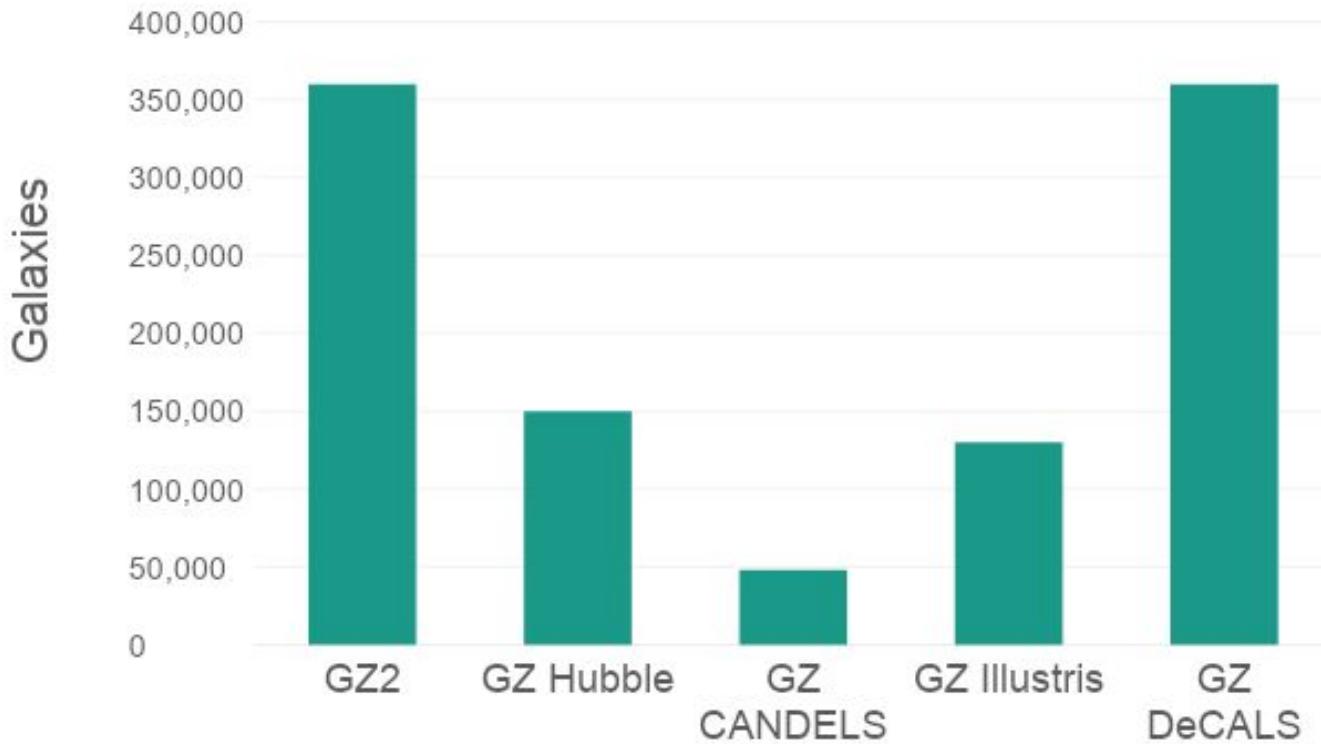
Fails to generalise

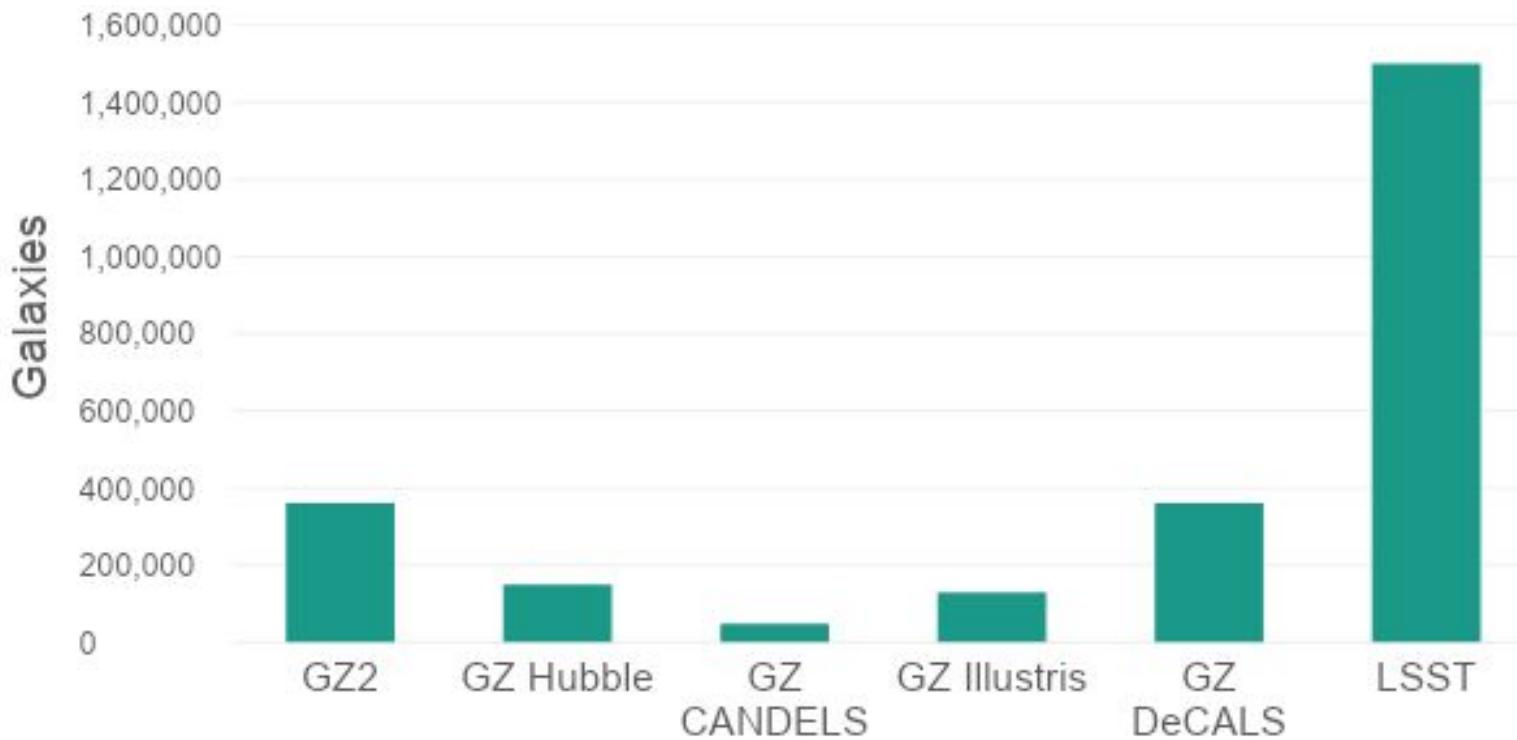
Retrain on citizen scientist labels

Performs far better



1. Fast Radio Bursts: Shortcuts
2. Galaxy Zoo: Uncertainty and Information
3. Now what?

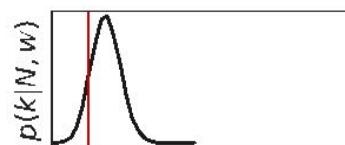
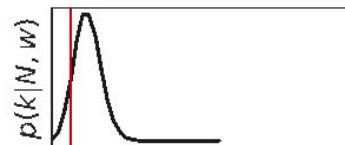
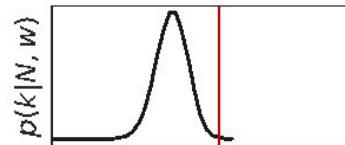
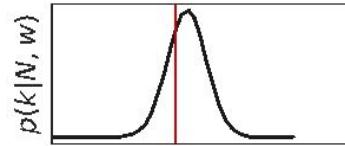






Posteriors for Votes

- Our CNN can learn from uncertain labels and make probabilistic predictions $p(k|w)$



0 20 40
 k 'Bar' votes, of N total

1 Model

Probabilistic CNN

N volunteers and k responses \approx N trials and k successes

How fair might
the coin be? $\text{Beta}(\rho|\alpha, \beta)$

Toss N times,
get k heads $\text{Bin}(k|\rho, N)$

How likely is each ρ given observed k, N ?

$$\mathcal{L} = \int \text{Beta}(\rho|\alpha, \beta) \text{Bin}(k|\rho, N) d\alpha d\beta$$

Predict $f^w(x) = \alpha, \beta$ and maximise the likelihood of α, β

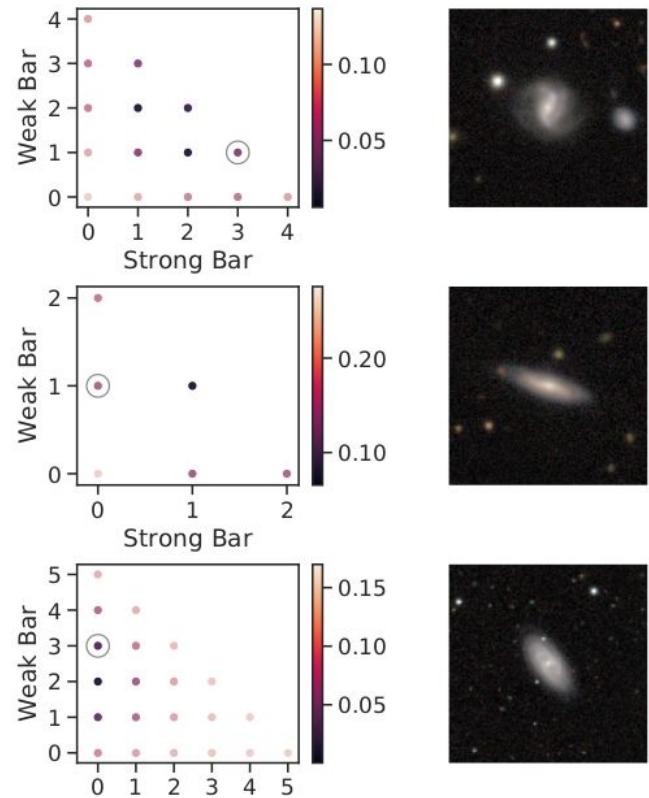
Volunteers N
Responses k
Typical vote prob. ρ
Galaxy x
CNN output $f^w(x)$

Multiple Answers

$$\mathcal{L} = \int \text{Beta}(\rho|\alpha, \beta) \text{Bin}(k|\rho, N) d\alpha d\beta$$

Add a few dimensions...

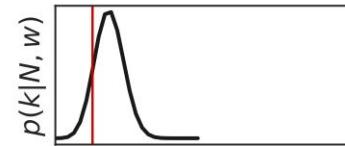
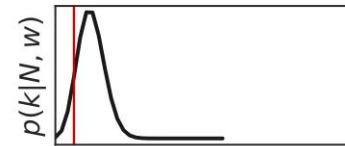
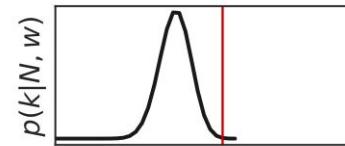
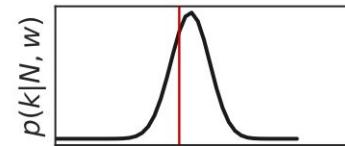
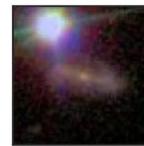
$$\mathcal{L}_q = \int \text{Dirichlet}(\vec{\rho}|\vec{\alpha}) \text{Multi}(\vec{k}|\vec{\rho}, N) d\vec{\alpha}$$



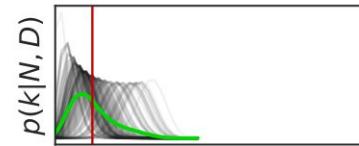
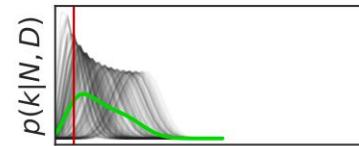
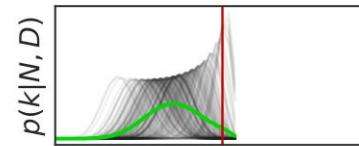
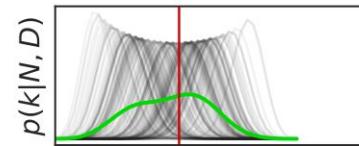
Posteriors for Votes

- Our CNN can learn from uncertain labels and make probabilistic predictions $p(k|w)$
- Marginalising over weights (BCNN) lets us predict votes over all CNN we might have trained

$$p(k|D) = \int p(k|w) p(w|D) dw$$



For details on BCNN, see Y. Gal (2016)



1 Model

30 “Models” (BCNN)

Probabilistic to Bayesian CNN

Galaxy x
 CNN weights w
 Training data D_{train}
 CNN output $f^w(x)$
 Dropout dist. q_θ^*
 Forward pass t of T

What about the models we might have trained, but didn't?

$$p(y = c | x, D_{train}) = \int f^w(x) p(w | D_{train}) dw$$

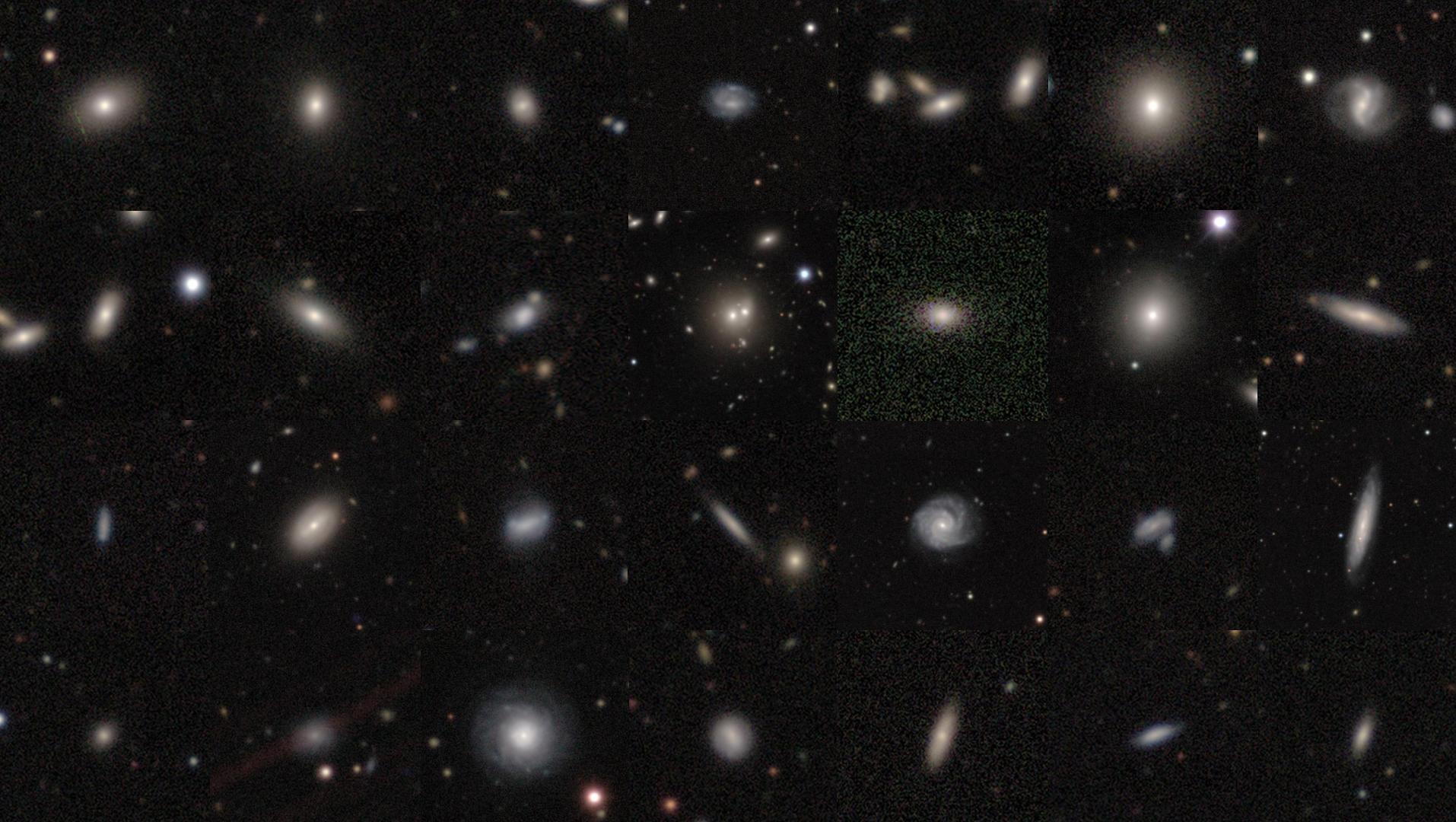

Unknown!

Approximate $p(w | D_{train})$ with Dropout

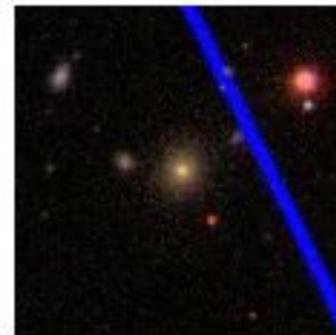
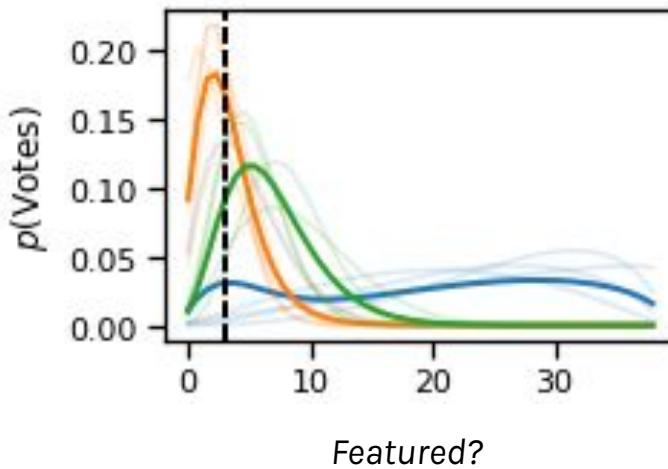
$$\approx \int q_\theta^*(w) dw$$

$$\approx \frac{1}{T} \sum_{t=1}^T f^{w_t}(x)$$

See Y. Gal et al
 (2016)







Pick galaxies where the models **confidently** disagree.

$$I = - \int H[p(k|w)] p(w|D) dw + H \left[\int p(k|w) p(w|D) dw \right]$$

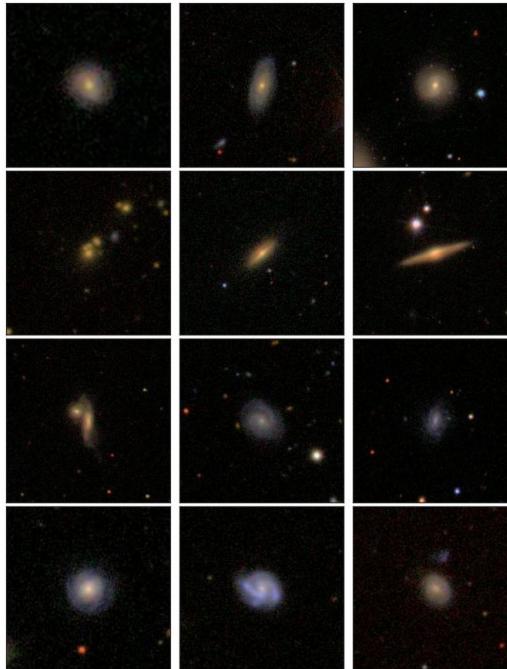
↑

Each model is
confident...

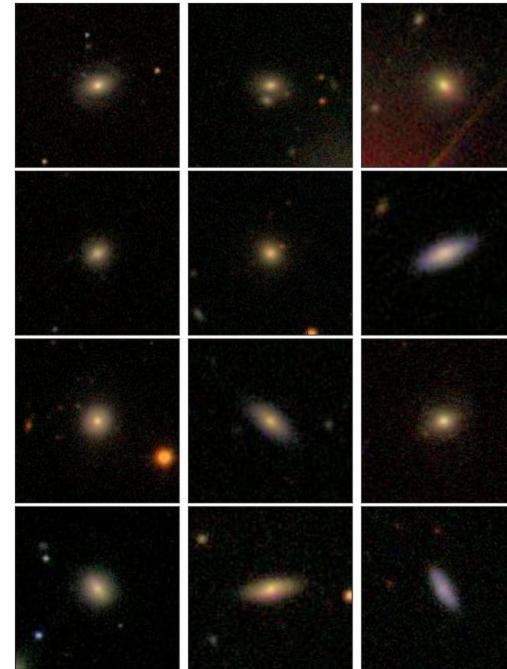
...but they give different
answers

Mutual Information I
Entropy H
Votes k
Weights w
Training data D

Selected Galaxies for “Smooth?”

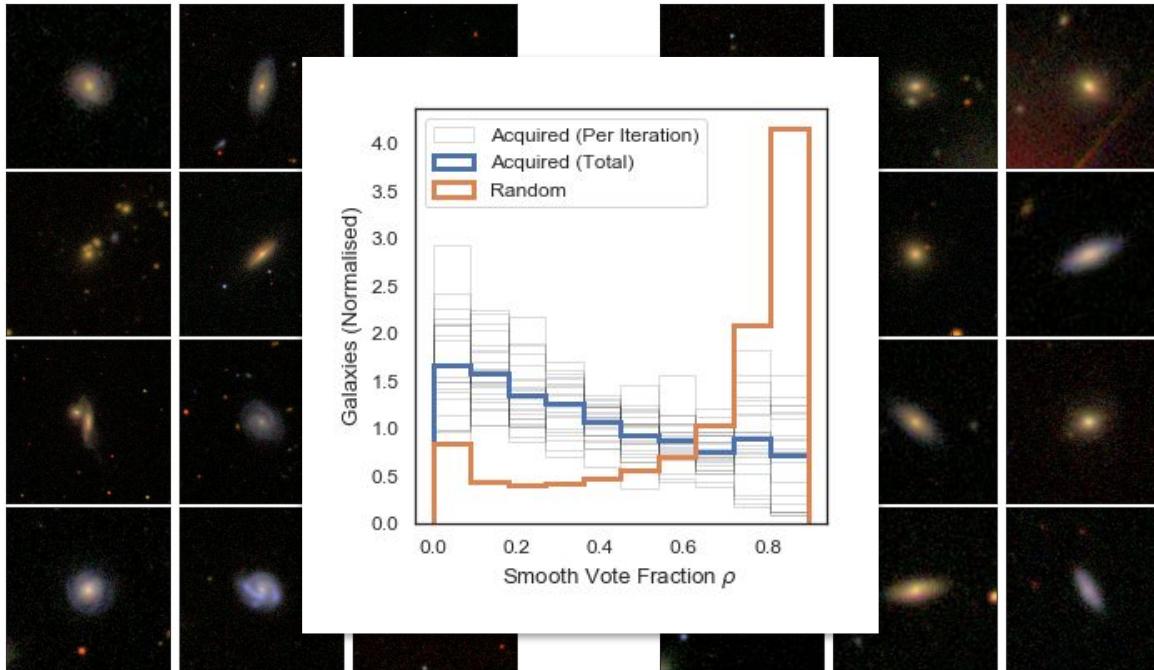


High mutual information



Low mutual information

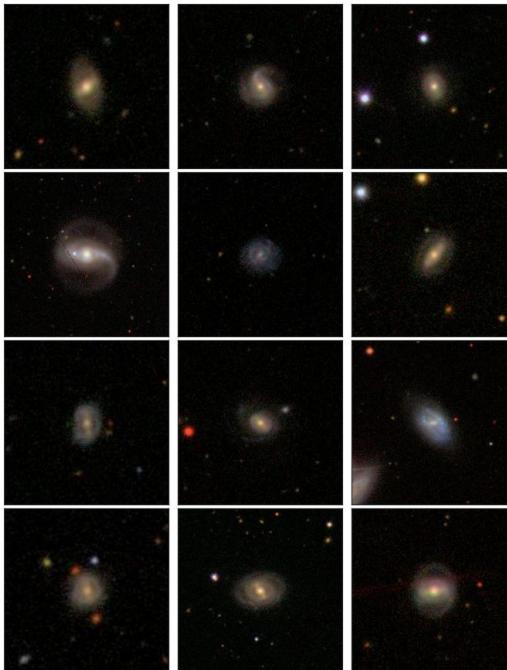
Selected Galaxies for “Smooth?”



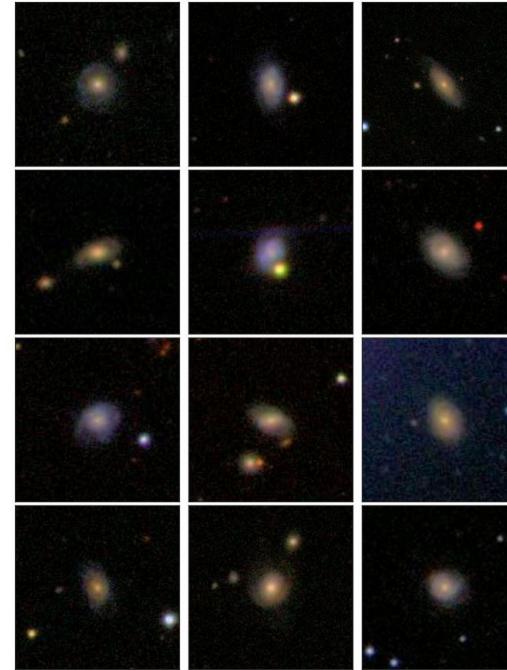
High mutual information

Low mutual information

Selected Galaxies for “Bar?”

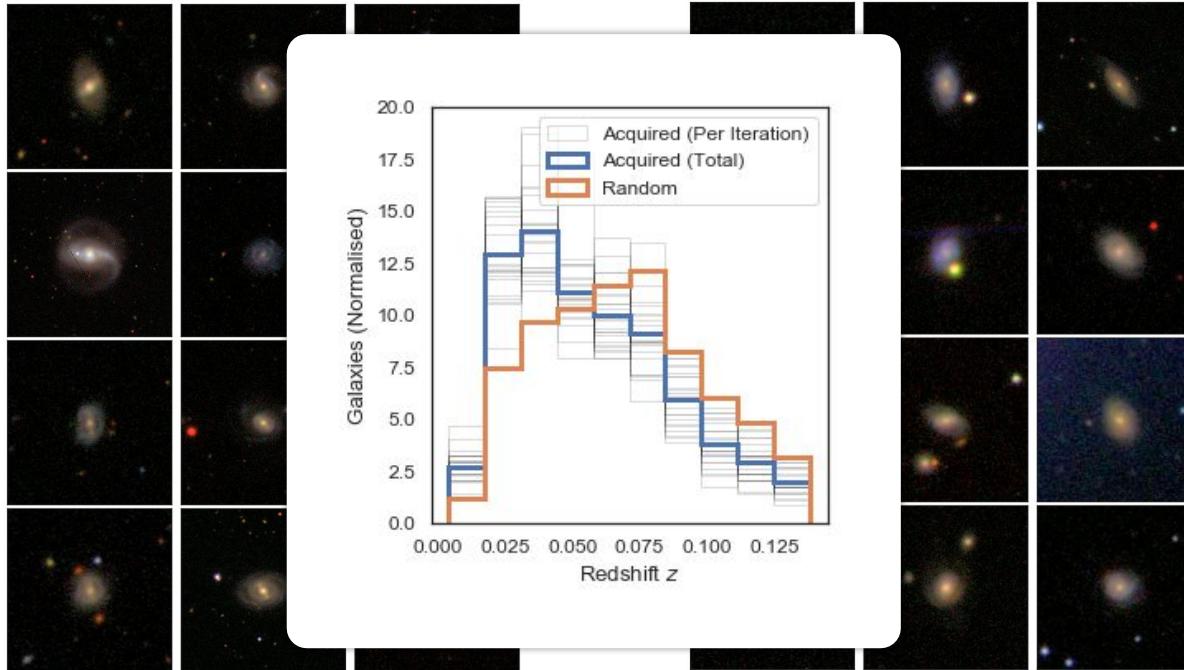


High mutual information



Low mutual information

Selected Galaxies for “Bar?”



Performance

~ 99% accurate on every question
for galaxies where the volunteers
are confident

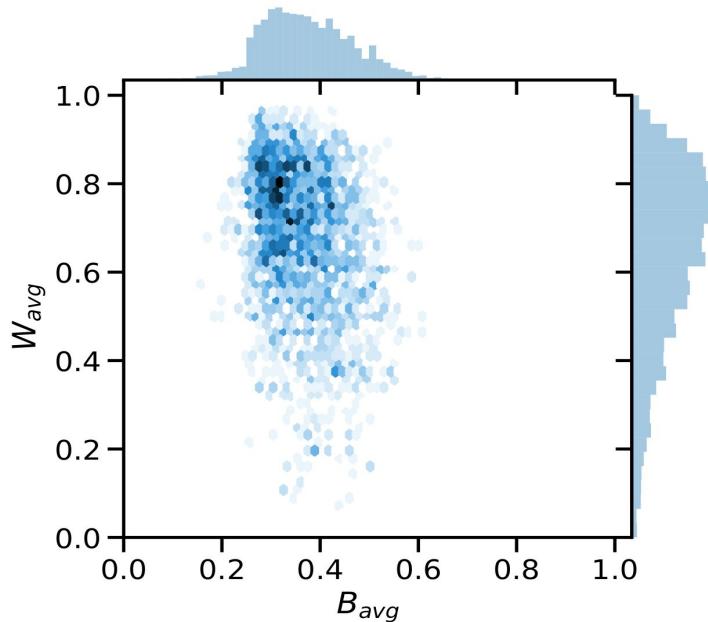
Question	Count	Accuracy	Precision	Recall	F1
Smooth Or Featured	3495	0.9997	0.9997	0.9997	0.9997
Disk Edge On	3480	0.9980	0.9980	0.9980	0.9980
Has Spiral Arms	2024	0.9921	0.9933	0.9921	0.9924
Bar	543	0.9945	0.9964	0.9945	0.9951
Bulge Size	237	1.0000	1.0000	1.0000	1.0000
How Rounded	3774	0.9968	0.9968	0.9968	0.9968
Edge On Bulge	258	0.9961	0.9961	0.9961	0.9961
Spiral Winding	213	0.9906	1.0000	0.9906	0.9953
Spiral Arm Count	659	0.9863	0.9891	0.9863	0.9871
Merging	3108	0.9987	0.9987	0.9987	0.9987

Classification metrics on confident galaxies

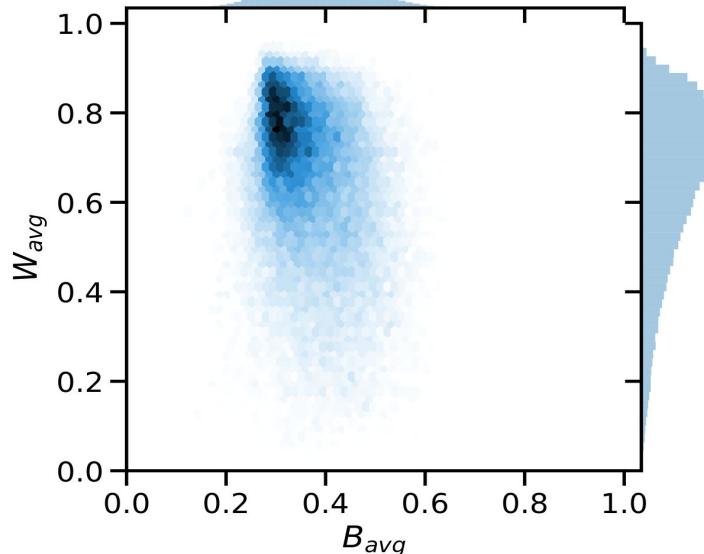
Live Demo

https://bit.ly/decals_viz

Volunteers (N=5378)



Automated (N=43672)



Winding angle vs. bulge size, measured by volunteers or deep learning

<https://zenodo.org/record/4196267>

1. Fast Radio Bursts: Shortcuts
2. Galaxy Zoo: Uncertainty and Information
3. Now what?

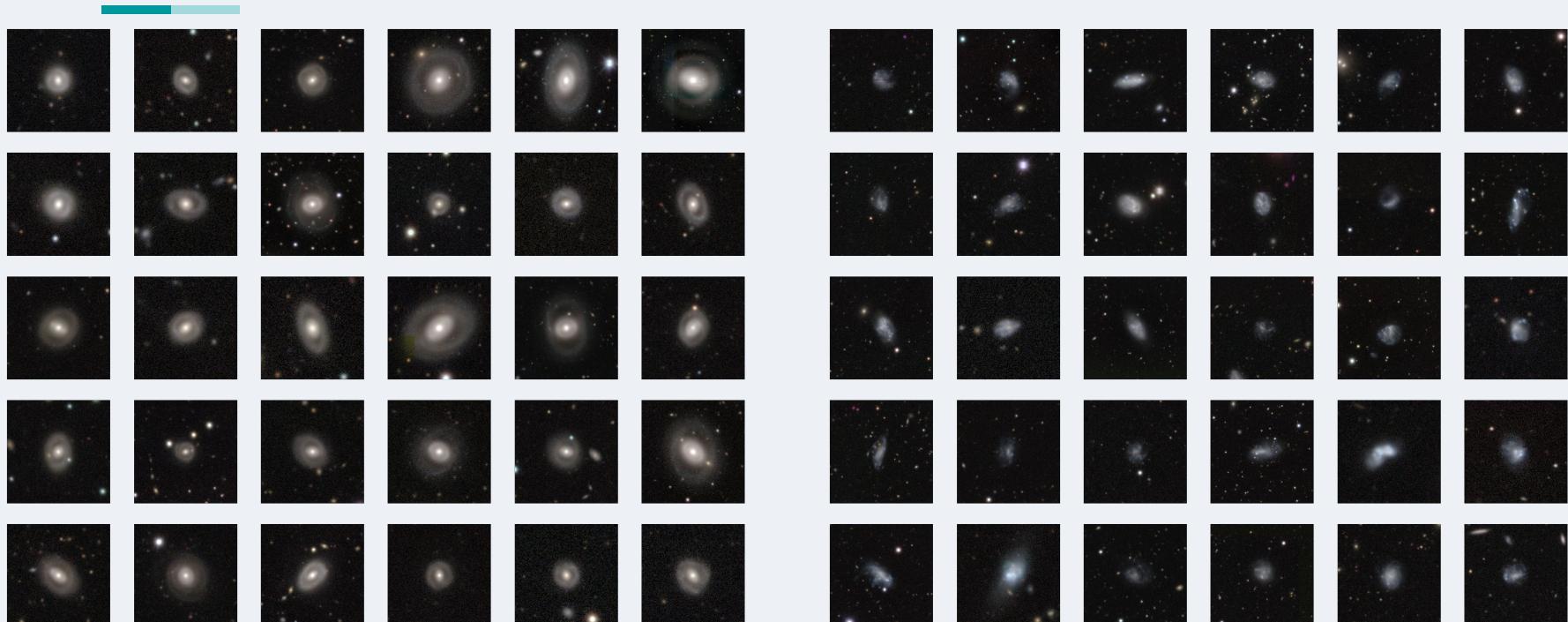
Live Demo #2!

https://bit.ly/decals_similarity

Transfer Learning

Automatic selection cuts: $\text{featured} > 0.6$, $\text{face-on} > 0.7$, $\text{has spiral arms} < 0.5$.

Training set: 212 positive rings



Max prob. “ring”, validation set

Min prob. “ring”, validation set

Zoobot

Navigation

Components

[Finetuning Guide](#)

[Training from Scratch](#)

[data_utils](#)

[estimators](#)

[training](#)

[predictions](#)

[schemas](#)

[label_metadata](#)

Quick search

Go

```
# Hiring 4 Python?
while is_open(job):
    try:
        # Hire easier!
        promote(RTD)
    finally:
        print('HIRED')
```

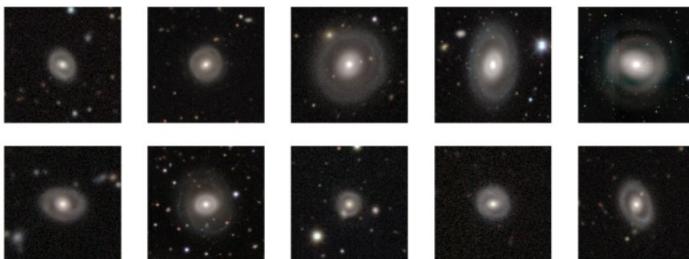
Hiring Python devs? Read the
Docs can help!

Sponsored · Ads served ethically

Zoobot Documentation

Guides

Zoobot makes it easy to train a state-of-the-art deep learning classifier to solve your galaxy morphology problem. For example, you can train a classifier to find ring galaxies in under [200 lines of code](#).



Ring galaxies found using Zoobot and 212 labelled examples.

You do not need to be a machine learning expert to use Zoobot. Zoobot includes [components](#) for common tasks like loading images, managing training, and making predictions. You simply need to assemble these together.

You will only need a small amount of labelled images; a few hundred is a good starting point. This is because Zoobot includes a classifier already trained to answer Galaxy Zoo questions for DECaLS galaxies. Retraining (finetuning) this model requires much less time and labels than starting from scratch.

If you do want to start from scratch, to reproduce or improve upon the pretrained classifier, [Zoobot can do that as well](#).

No experience required

zoobot.readthedocs.io

github.com/mwalsmsley/zoobot

@mike_walmsley_ 

CHIME:
bit.ly/burstsfromspace

Active Learning:
arxiv: 1905.07424

DECaLS:
arxiv: 2102.08414
zenodo: 4196267

Similarity:
bit.ly/decals_similarity

Transfer learning:
[github.com/mwalmsley/
zooobot](https://github.com/mwalmsley/zooobot)