Anomaly detection using local measures of uncertainty in latent representations

Fiona Porter & Anna Scaife
Jodrell Bank Centre for Astrophysics, University of Manchester

ML-IAP/CCA 2023, 1st December 2023

Motivation

- Next-generation telescope surveys like the Square Kilometre Array will significantly increase the number of sources we can detect, but the data rates will be too high for humans to label them

- Machine learning is the obvious solution for rapid and accurate class labels

- But - some of the sources we detect might be entirely new populations revealed by improved resolution and sensitivity. How do we make sure we don't accidentally miss them?

SKA South Africa site.
Photo: Mike Peel, https://www.mikepeel.net/

Motivation

- Hard to teach a model to recognise an “unknown unknown”, and can’t rely on serendipitous discovery - we need reliable **anomaly detection**
- A number of methods exist already for image data (ASTRONOMALY, Lochner & Bassett 2020; Self-Organising Maps, e.g. Ralph et al. 2019, Gupta et al. 2022)
- But - these are unsupervised methods: they don’t provide classes used by astronomers
- Can we do anomaly detection as a byproduct of our classification?

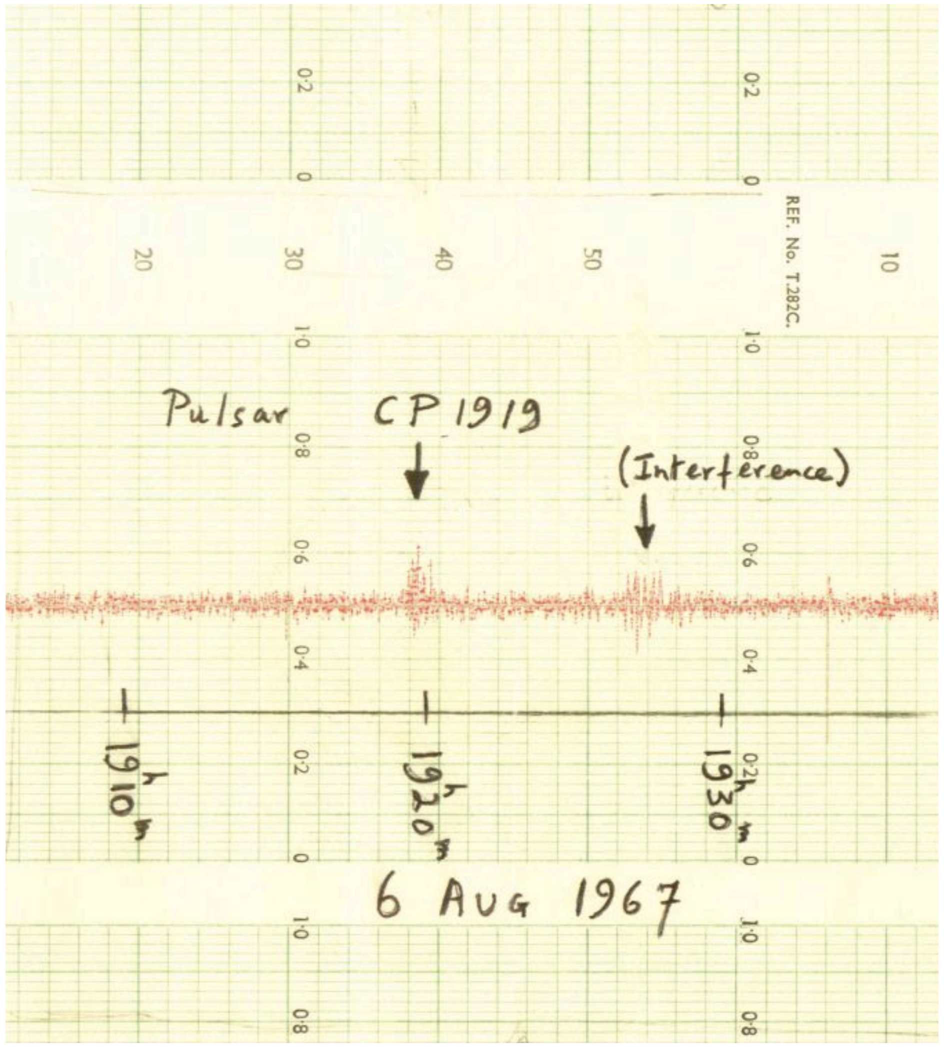


Chart of first pulsar discovered
Image courtesy of the Cavendish Laboratory

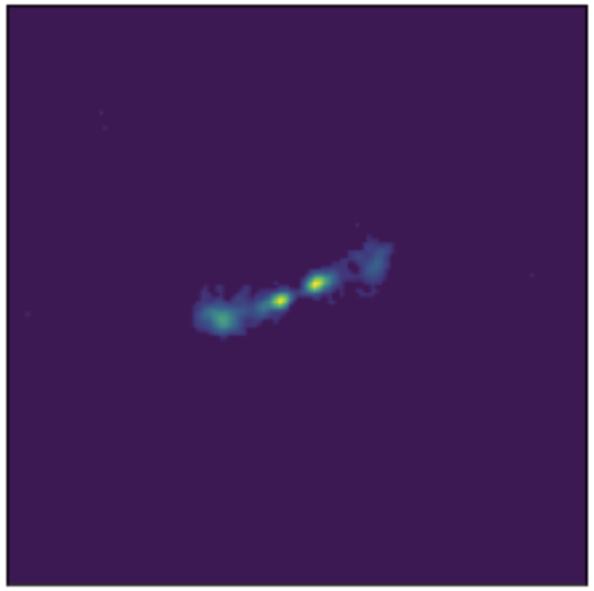
Lochner & Bassett (2020): https://arxiv.org/abs/2010.11202
Ralph et al. (2019): https://arxiv.org/abs/1906.02864
Gupta et al. (2022): https://arxiv.org/abs/2208.13997

Model and data

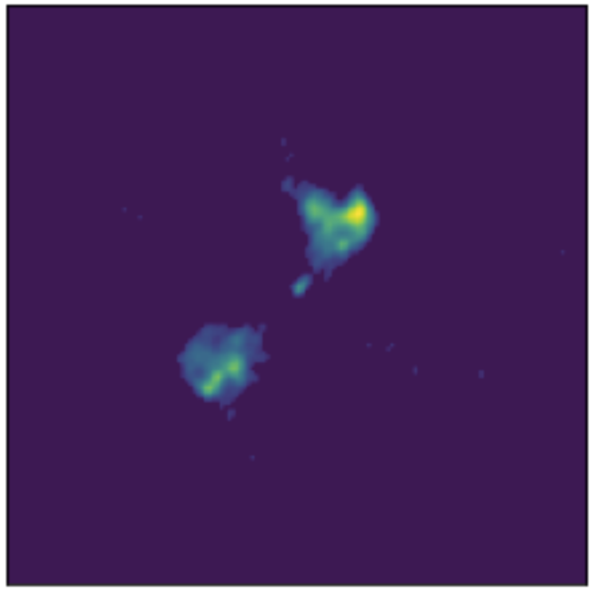
- Model: variant on LeNet-5 using spectral normalisation and Monte Carlo dropout to measure uncertainty
- Images from the MiraBest dataset: Fanaroff-Riley (FR) galaxies, classified FRI, FRII and hybrid
- Train a binary classifier on 1256 FRIs and FRIIs; use the 108 hybrid sources as our “anomalies” which might be mistaken for binary FR sources

MiraBest can be found on zenodo: 10.5281/zenodo.5588282

Porter & Scaife (2023): https://arxiv.org/abs/2305.11108



FRI: Diffuse jets, brightest near galactic host’s core. Surrounding IGM expected to be dense.



FRII: strong jets with prominent lobes, brightest near lobes. Surrounding IGM expected to be thin.

Anomaly detection: entropy

- Entropy is an information theory metric that measures total model uncertainty in a prediction
- Requires that a model's predictions can vary when classifying the same source multiple times (hence MC dropout)

$$H = -\sum_{i=1}^{N} \left(\frac{1}{T} \sum_{t} p_i \right) \log \left(\frac{1}{T} \sum_{t} p_i \right)$$

Notation:

N: no. of classes in model

T: no. of dropout configurations used

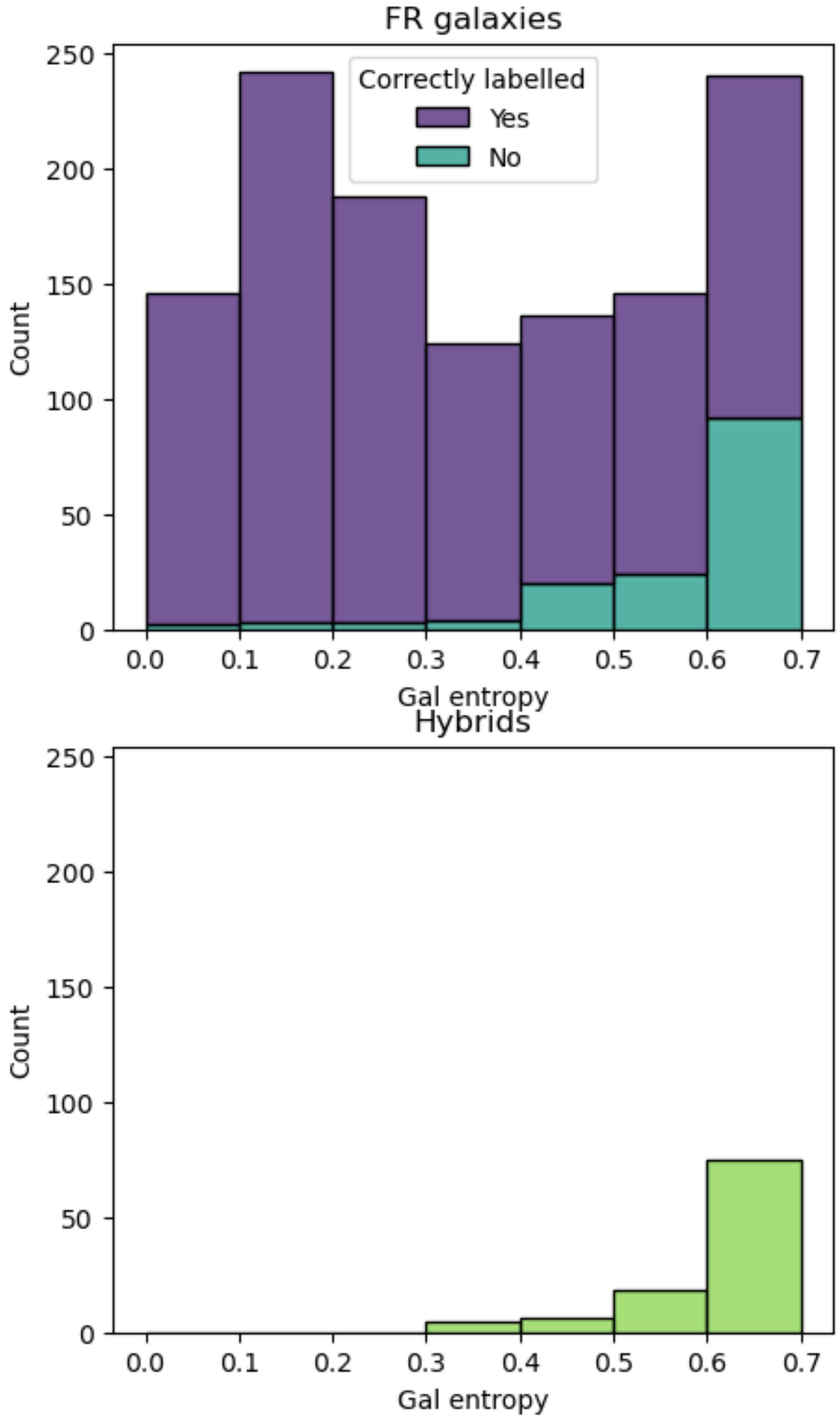
p_i: softmax probability of image being the ith class

- Minimised when model always predicts a class with 100% probability (H = 0); maximised when all classes are equally probable (H = log(N)).

Gal (2017): PhD thesis, Uncertainty in Deep Learning

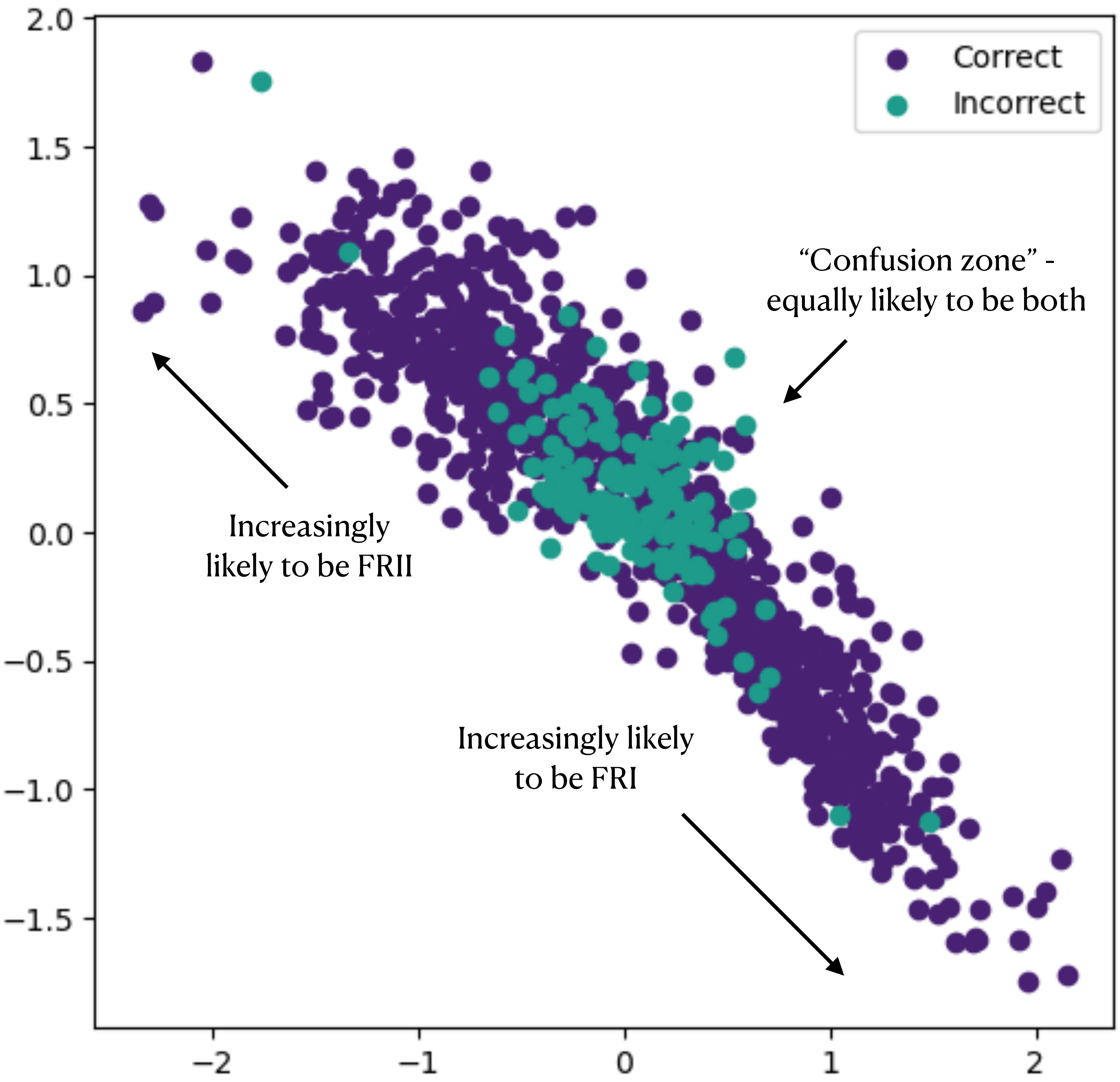
Anomaly detection: entropy

- Result: not very useful for anomaly detection specifically
- Hybrids do have high entropy, but they can't be separated from binary FR galaxies
- Examining the entropy > 0.6 bin would give you:
  - 75 hybrid sources (~24%)
  - 92 mislabelled FR galaxies (~29%)
  - 148 correctly-labelled FR galaxies (~47%)
- Finding all hybrids would flag 646 FRs!



Anomaly detection: latent space

- Extract latent representations from the final layer before normalisations (logits)
- Logical ordering in latent space; maximum confusion in the centre, which is where most (but not all) misclassifications are found
- Do hybrids "live" somewhere different from FRs?



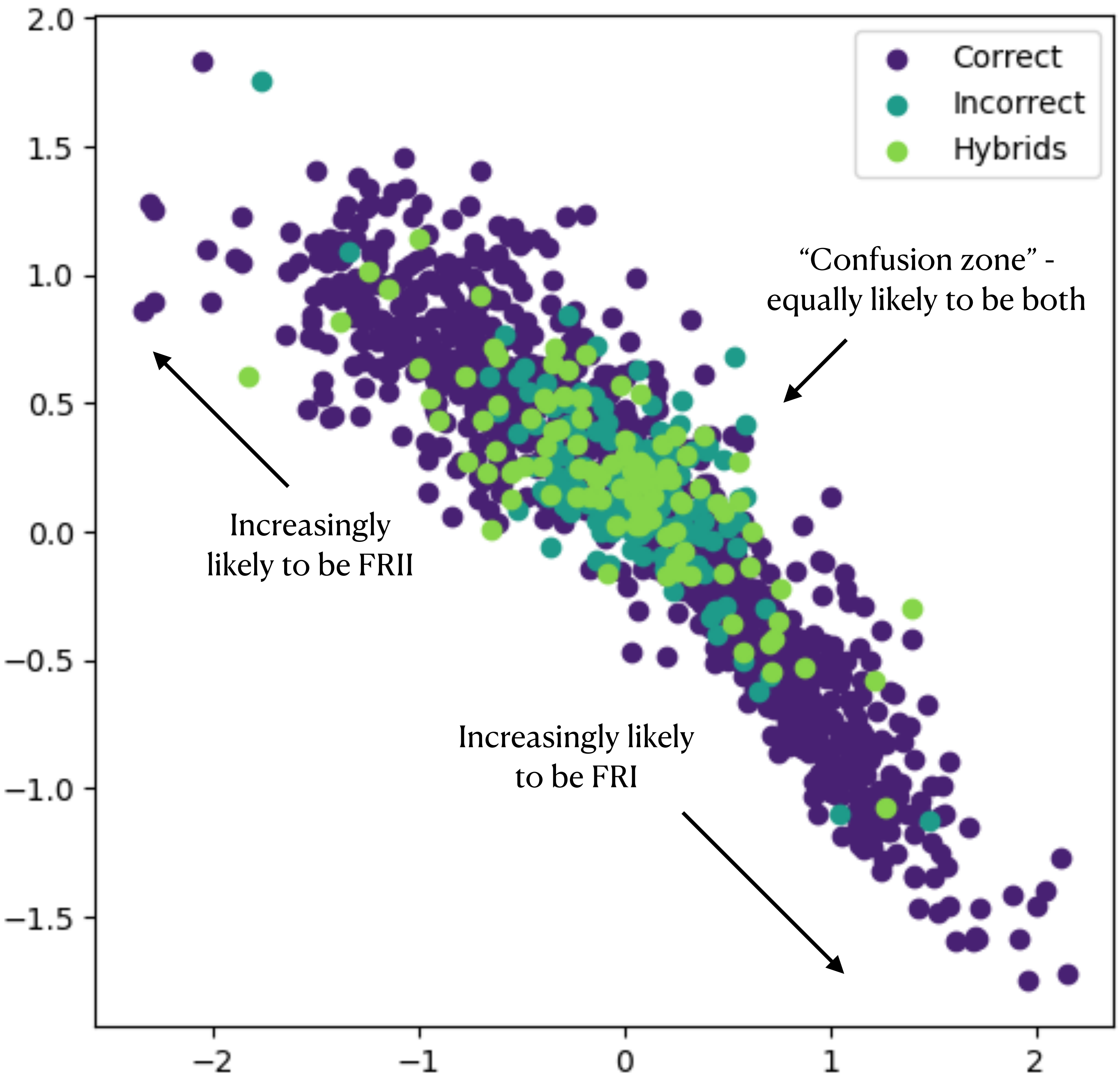
"Confusion zone" - equally likely to be both

Increasingly likely to be FRII

Increasingly likely to be FRI

Anomaly detection: latent space

- Unfortunately… not really, no
- Distribution almost entirely overlaps binary FRs, and isn't limited to the "confusion zone"
- Model also doesn't always confuse hybrids with same class; thinks some are very likely FRIs, others very likely FRIIs



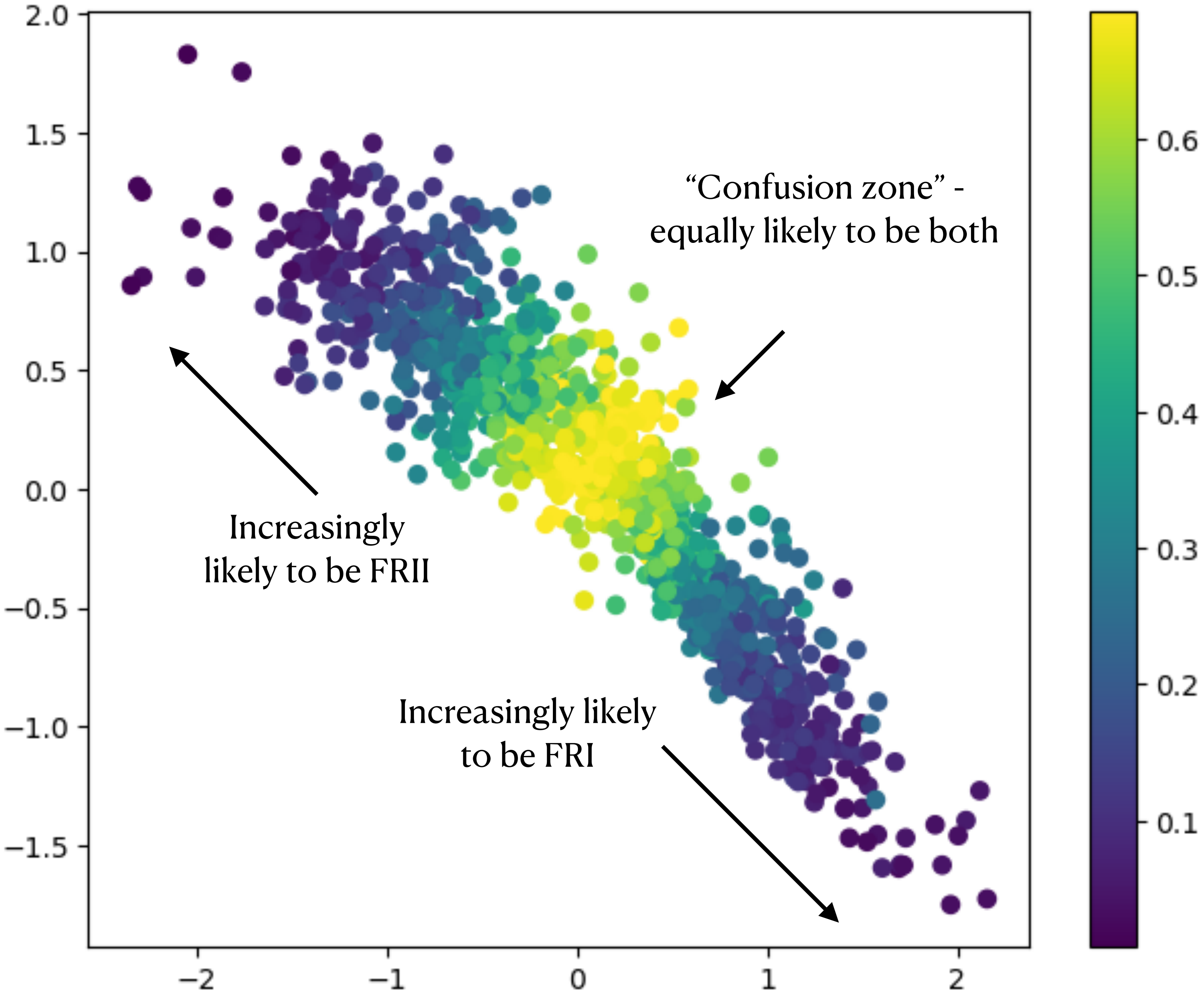
"Confusion zone" - equally likely to be both

Increasingly likely to be FRII

Increasingly likely to be FRI

Anomaly detection: both?

- Could we combine the metrics somehow?
- Entropy for FR galaxies generally decreases with distance from the “confusion zone”, but it’s not strictly linear
- Are hybrids higher entropy than FRs for their region in latent space?



“Confusion zone” - equally likely to be both

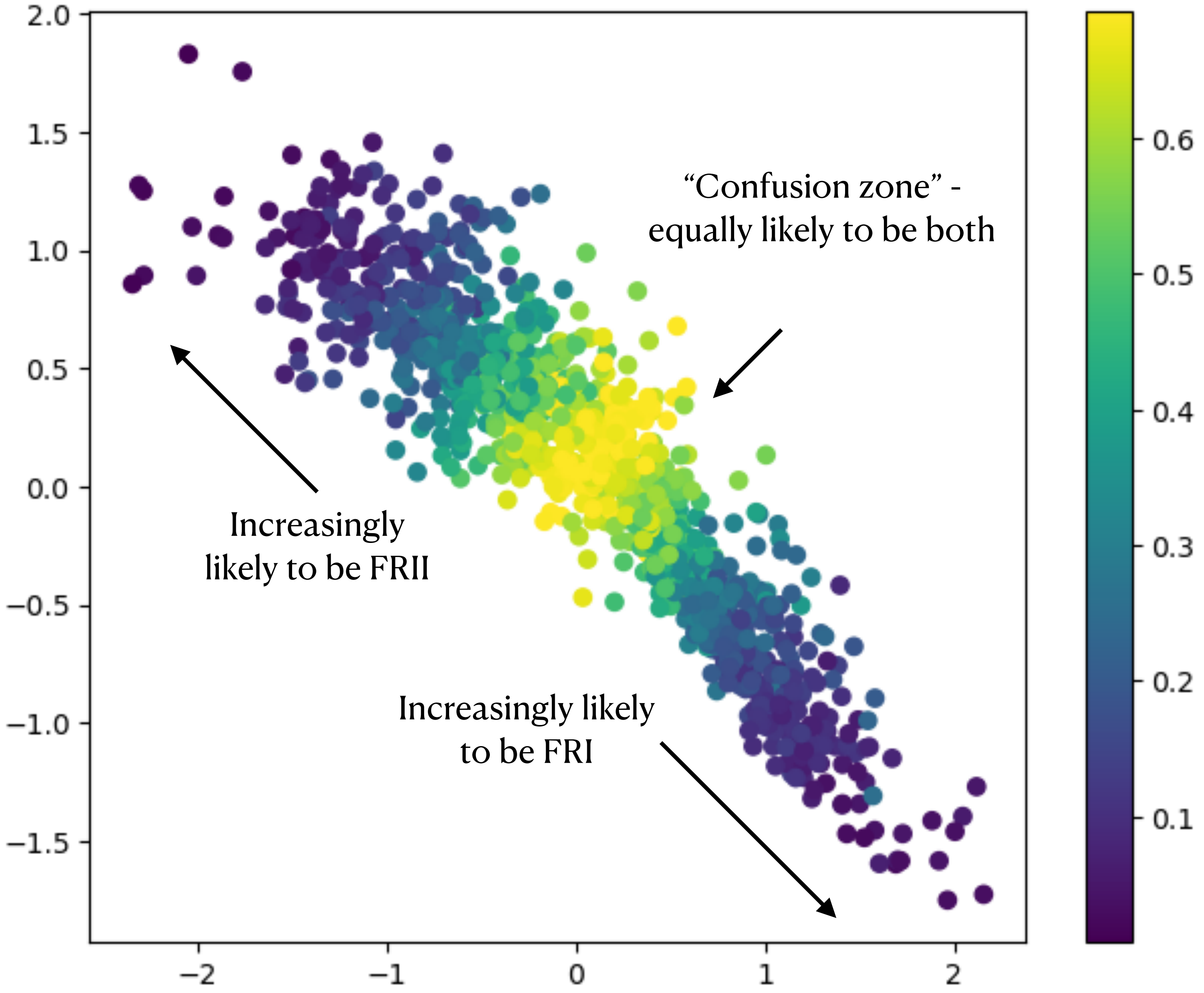
Increasingly likely to be FRII

Increasingly likely to be FRI

Anomaly detection: both?

- To assess how "anomalous" a hybrid's entropy is, we calculate "local entropy" - the entropy of the ten nearest correctly-labelled FR galaxies from the training set

- We declare a source "anomalous" if it's more than 3σ from the local entropy



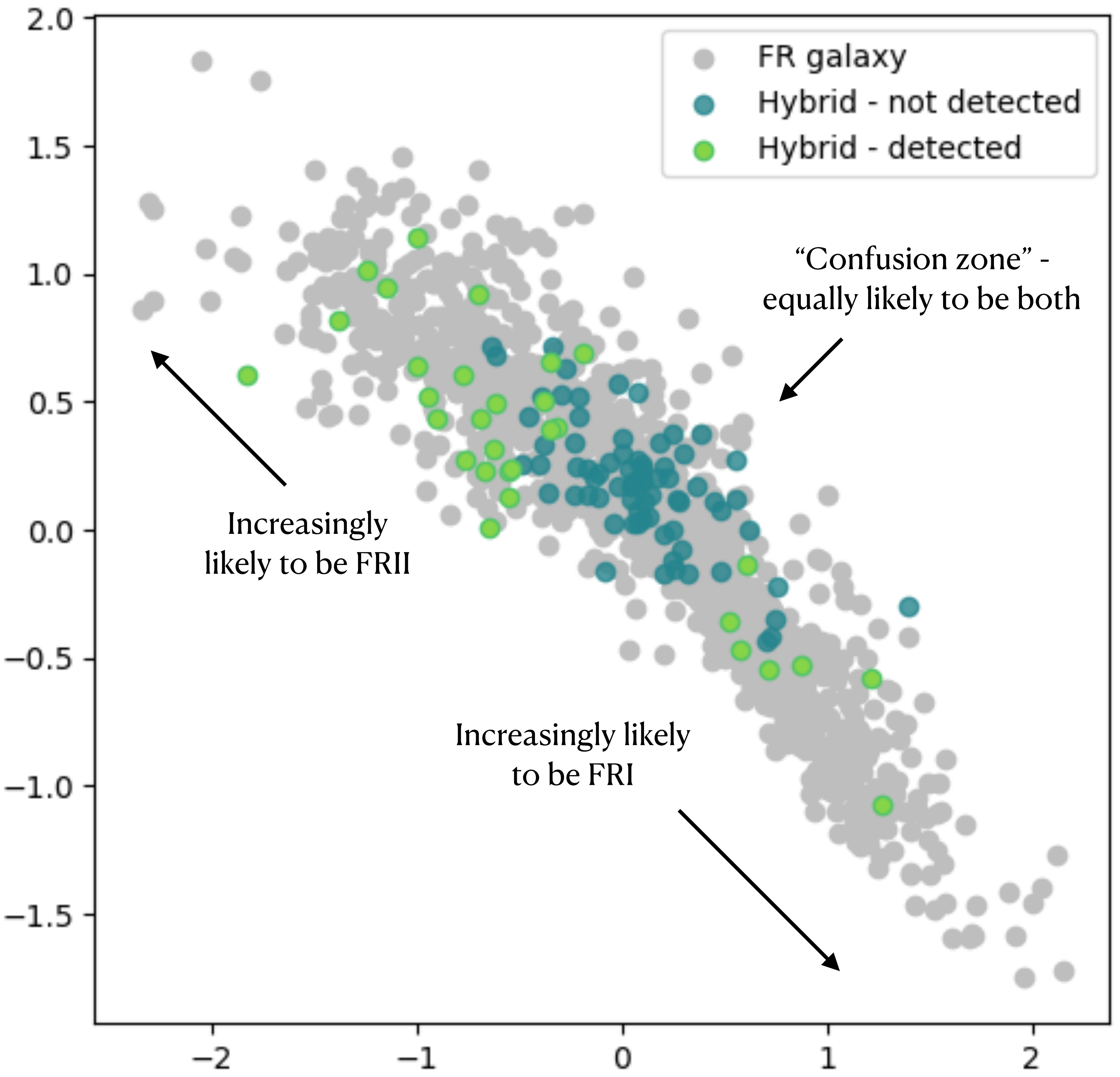
"Confusion zone" - equally likely to be both

Increasingly likely to be FRII

Increasingly likely to be FRI

Anomaly detection: both?

- Initial results: this might be useful!
- This flags 31/108 of the hybrids in our dataset, and does well at catching sources outside of the "confusion zone"
- Critically - **every source over 3σ** (including the test set FRs) **is a hybrid** - no false positive flags for anomalies

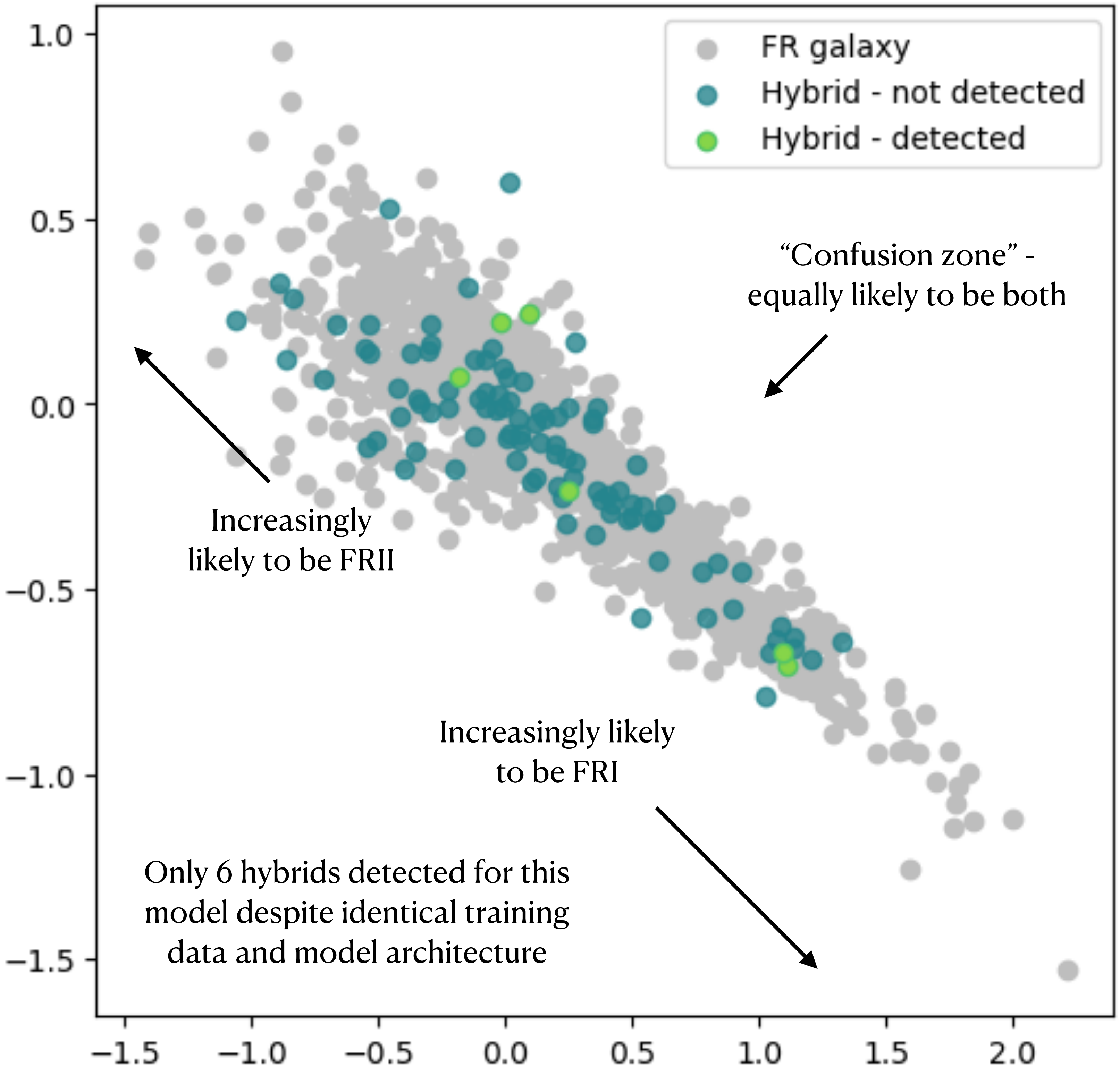


Except...

- From these initial results, decided to be rigorous and make sure it works reliably

- Bad news: **it doesn't always work this well** - some sort of model dependence even when using the same architecture and dataset

- Still trying to figure out exactly what allows it to work



"Confusion zone" - equally likely to be both

Increasingly likely to be FRII

Increasingly likely to be FRI

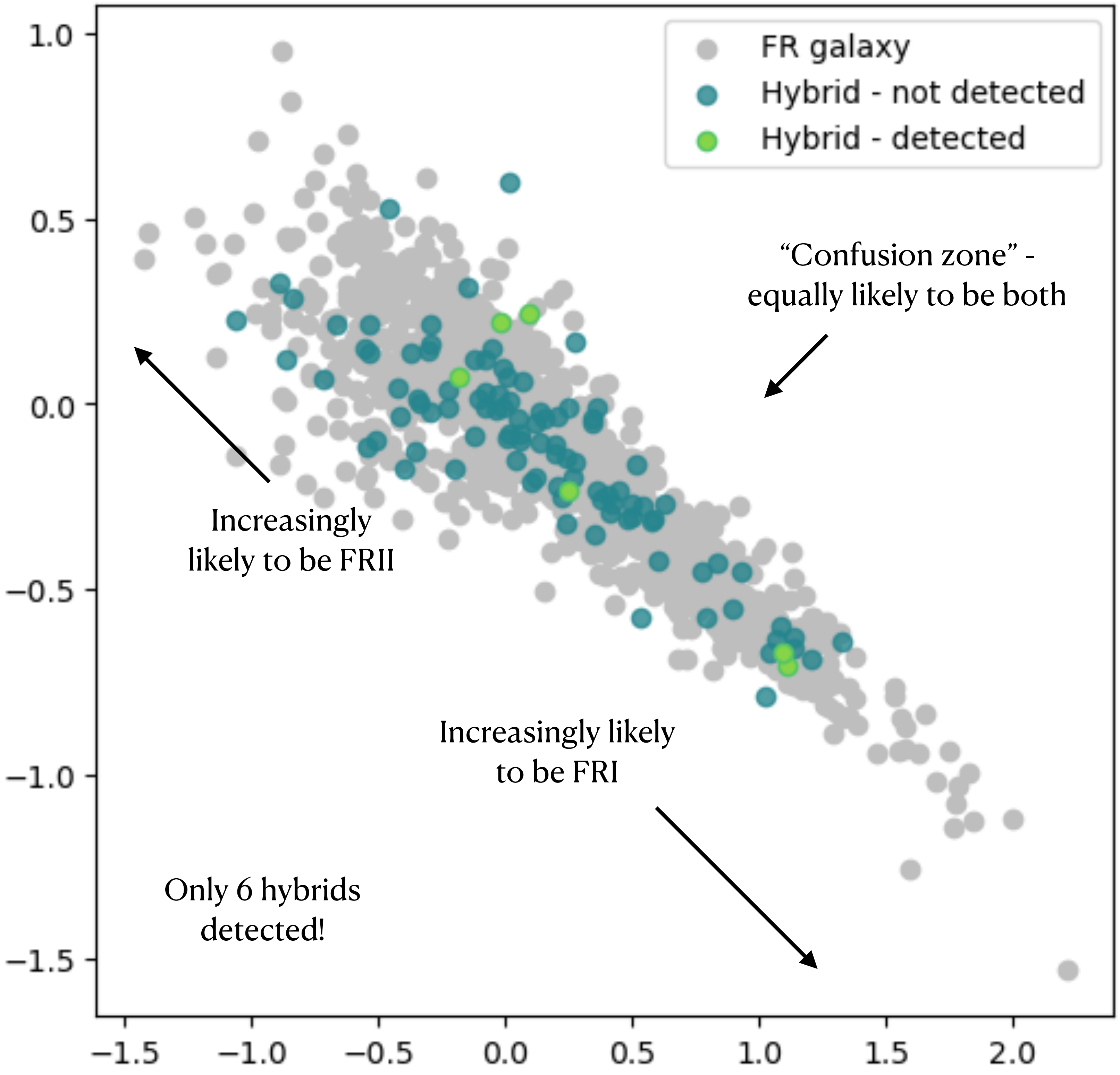
Only 6 hybrids detected for this model despite identical training data and model architecture

Even so…

- Even if it only flags a few “anomalies”, all sources > 3σ are still hybrids - not flagging any in-distribution sources as anomalous

- Requires no additional model training - just add MC dropout to a trained classifier

- Further work to try to make it more reliable



“Confusion zone” - equally likely to be both

Increasingly likely to be FRII

Increasingly likely to be FRI

Only 6 hybrids detected!

Conclusions

- Anomaly detection will be vital to detecting atypical, astronomically interesting sources in upcoming large-scale surveys
- Combining uncertainty measures that we get “for free” from our models offers a way to find (some) anomalies without dedicated anomaly detection pipelines
- More work to do to make it more reliable:
  - Checking uncertainty calibration
  - Other uncertainty measures
  - Alternative methods of finding local entropy in sparse regions

MiraBest dataset paper on arXiv

GitHub: fmporter

Email: fiona.porter@manchester.ac.uk

Bluesky: alicemayhap