

# Anomaly Detection in Astronomy

Michelle Lochner

Senior Lecturer

University of the Western Cape/  
South African Radio Astronomy  
Observatory



UNIVERSITY of the  
WESTERN CAPE



SARAO  
South African Radio  
Astronomy Observatory

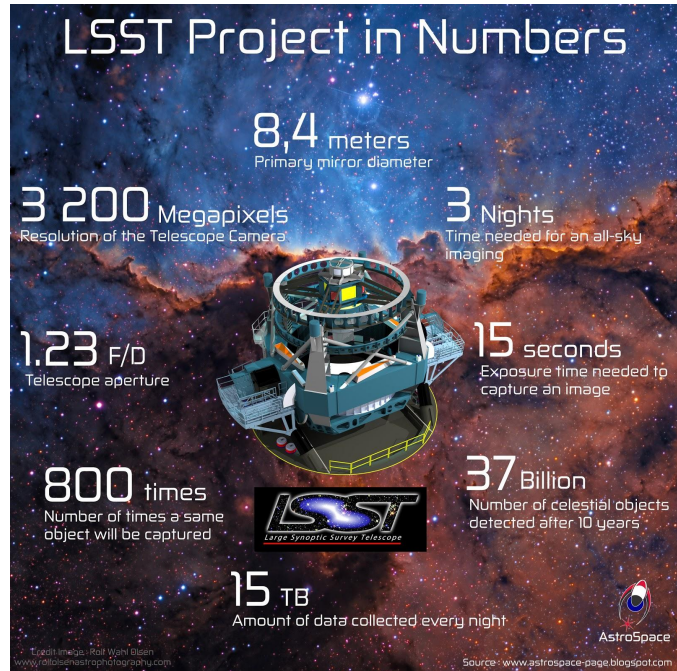


science & innovation

Department:  
Science and Innovation  
REPUBLIC OF SOUTH AFRICA

# The Vera C. Rubin Observatory

## Legacy Survey of Space and Time (LSST)



Rubin Obs/NSF/AURA & Bruno C. Quint

Day 000



Vera C. Rubin Observatory



# The Square Kilometre Array



## SKA1-mid

the SKA's mid-frequency instrument



Location:  
South Africa

Frequency range:  
**350 MHz**  
to  
**15.3 GHz**  
with a goal of 24 GHz



**197 dishes**  
(including 64 MeerKAT dishes)



Maximum baseline:  
**150km**

## SKA1-low

the SKA's low-frequency instrument



Location: Australia

Frequency range:  
**50 MHz**  
to  
**350 MHz**



**~131,000**  
antennas spread between  
512 stations



Maximum baseline:  
**~65km**

Square Kilometre Array Observatory

# MeerKAT



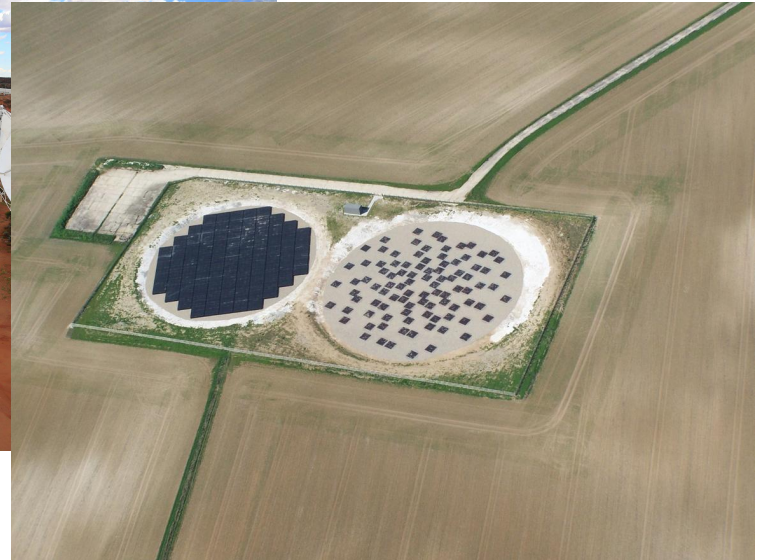
SARAO

# ASKAP



CSIRO

# LOFAR



LOFAR

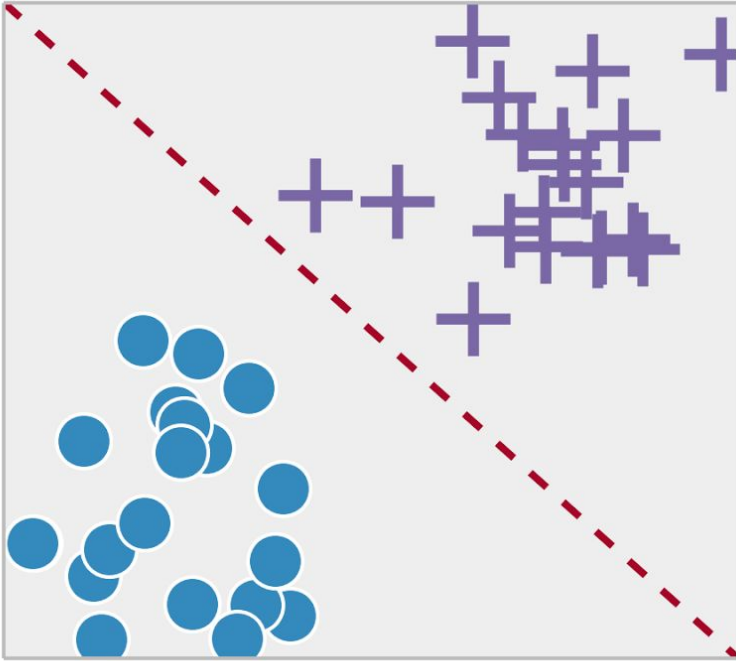


# We're facing a data explosion



# Machine Learning

# Supervised Machine Learning

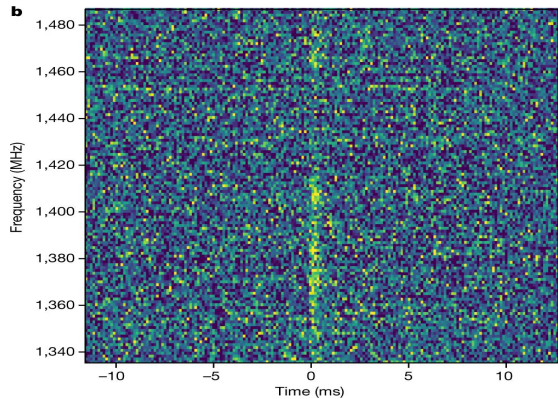
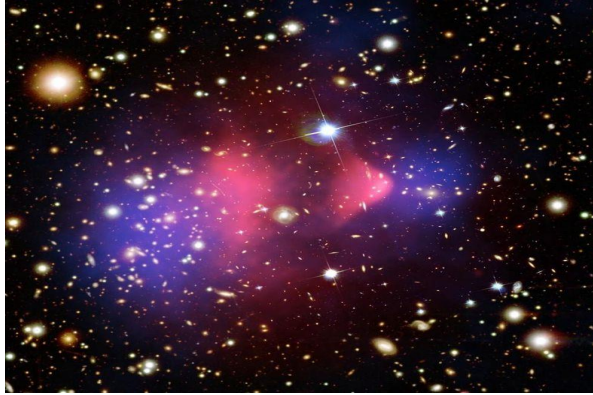


Automatically learns a **model** to map inputs to outputs, using a **training set**.

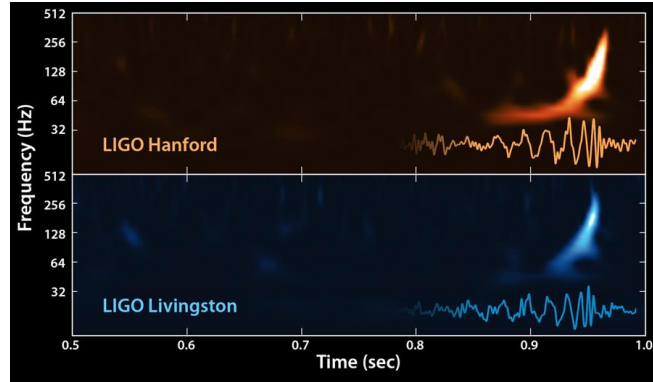


# Anomaly Detection

# Known Unknowns - rare events

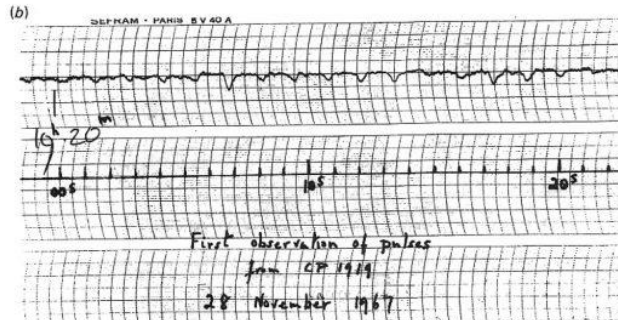
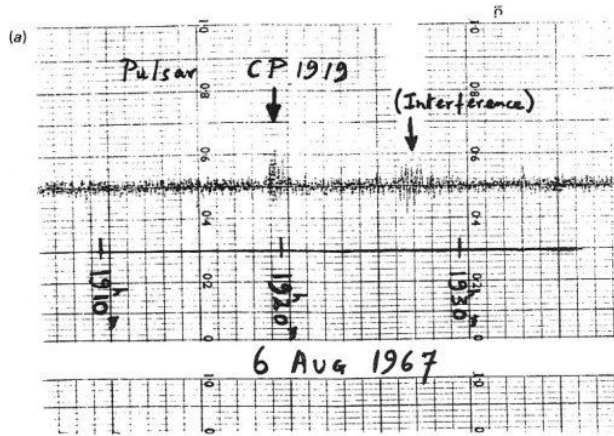


Ravi et al. (2019)



Hubble/ NASA/ ESA

# Unknown Unknowns - new anomalies

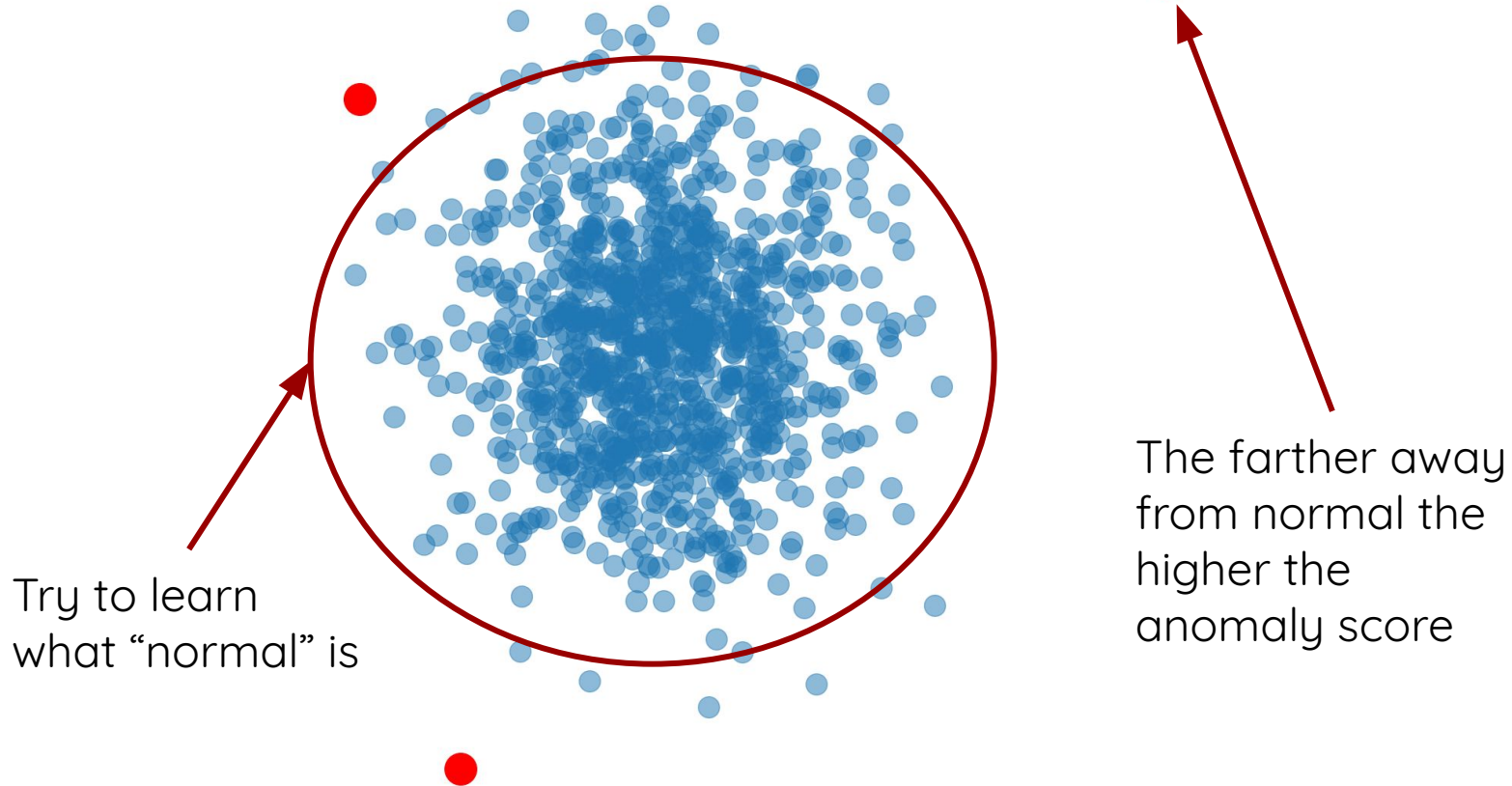




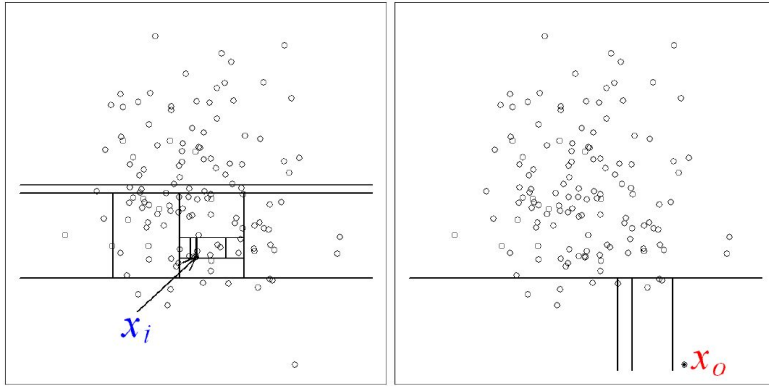
How do we discover new phenomena...

...among 10 million possibilities?

# Anomaly Detection

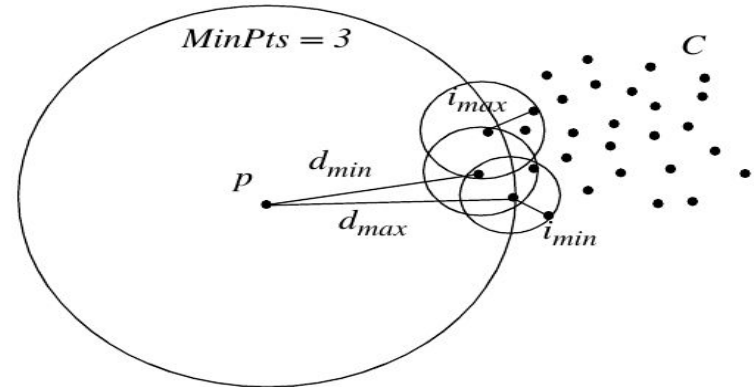


# Anomaly Detection Algorithms



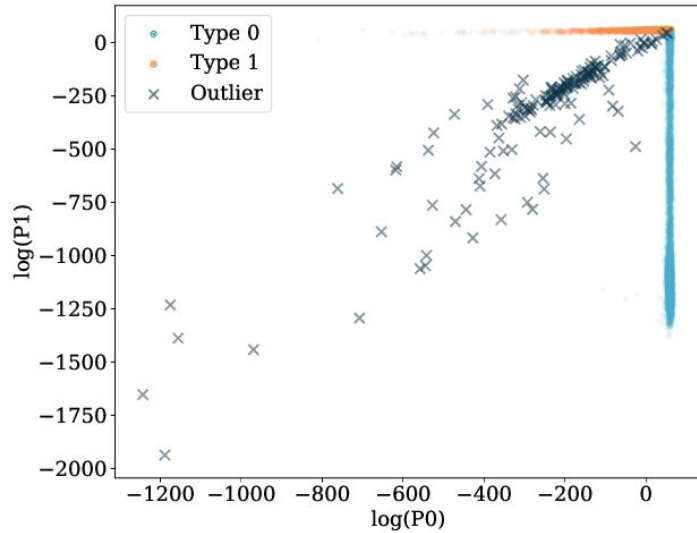
Isolation Forest (Liu,  
Ting & Zhou; 2008)

Local Outlier Factor  
(Breunig et al; 2008)



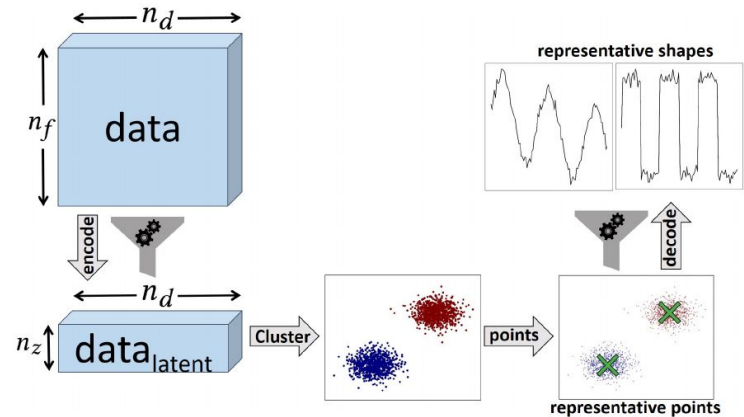


# Anomaly Detection Algorithms



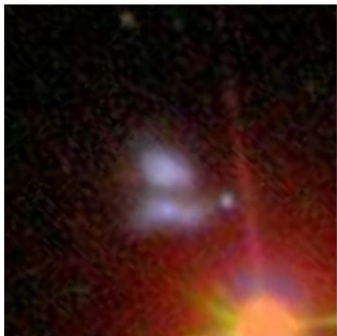
**BADAC** (Roberts, Bassett & Lochner - 1902.08627)

**DRAMA** (Vafaei Sadr, Bassett & Kunz - 1909.04060)

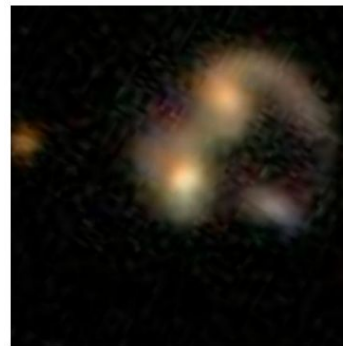


# Anomaly Detection Isn't Enough

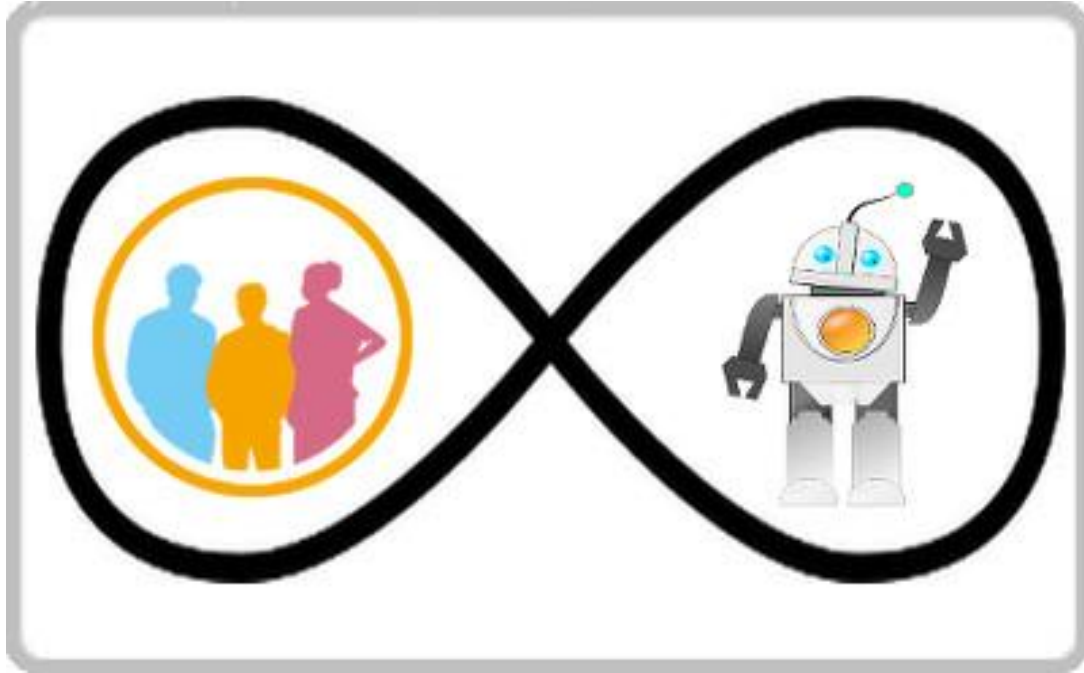
Artefacts



Real sources

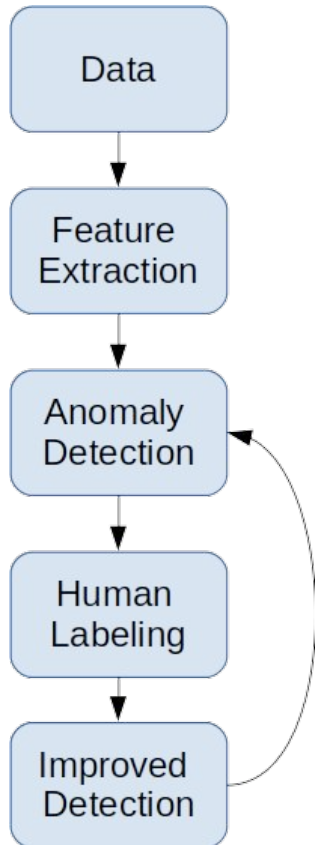


# Active Learning





# Astronomy



The screenshot shows the Astronomy web application interface. The browser address bar displays "127.0.0.1:5000". The application has two tabs: "ANOMALY SCORING" (active) and "VISUALISATION".

The main content area displays a central image of a galaxy. To the right, there are two panels:

- Metadata:**

```
filename: /home/michelle/BigData/Anomaly/  
Class6.1: 1  
score: 4.7  
human_label: 5
```
- Features:**

```
Residual_90: 1.21  
Residual_80: 7.24  
Residual_70: 5.46  
Residual_60: 11.6  
Residual_50: 11.5  
Residual_0: 3.59
```

Below the image, there is a navigation bar with a "3" index and "Index" label. Below that, a question "HOW INTERESTING IS THIS OBJECT?" is followed by a row of buttons labeled 0, 1, 2, 3, 4, and 5. The button "5" is highlighted in blue. Below the buttons, there is a dropdown menu for "Raw anomaly score" and a toggle switch for "Show unlabelled objects first". At the bottom right, there are two buttons: "DELETE LABELS" and "RETRAIN".

Lochner and Bassett (2020) - [2010.11202](https://arxiv.org/abs/2010.11202)  
<https://github.com/MichelleLochner/astronomy>

# Galaxy Zoo - Random Examples

The screenshot displays the 'Astronomy' application interface. The browser address bar shows '127.0.0.1:5000'. The application has two tabs: 'ANOMALY SCORING' (active) and 'VISUALISATION'. The main content area features a large image of a galaxy. To the right of the image are two panels: 'Metadata' and 'Features'. Below the image is a slider for 'Index' with a value of 0. At the bottom, there is a section titled 'HOW INTERESTING IS THIS OBJECT?' with a scale from 0 to 5. A 'Random' button is present, along with a checkbox for 'Show unlabelled objects first', and 'DELETE LABELS' and 'RETRAIN' buttons.

**Metadata**

- filename: /home/michelle/BigData/Anomaly/
- Class6.1: 0.279
- score: 0.0847
- human\_label: -1

**Features**

- Residual\_90: -0.364
- Residual\_80: -0.277
- Residual\_70: 0.299
- Residual\_60: -0.326
- Residual\_50: -0.529
- Residual\_0: -0.428

Index: 0

HOW INTERESTING IS THIS OBJECT?

0 1 2 3 4 5

Random

Show unlabelled objects first

DELETE LABELS RETRAIN

Scoring method to sort by

# Galaxy Zoo - No Active Learning

The screenshot displays the Galaxy Zoo web interface for anomaly scoring. The browser window shows the URL '127.0.0.1:5000'. The interface has two tabs: 'ANOMALY SCORING' (active) and 'VISUALISATION'. The main content area features a central image of a galaxy with a bright core and spiral arms. To the right of the image are two panels: 'Metadata' and 'Features'. The 'Metadata' panel contains the following information:

```
filename: /home/michelle/BigData/Anomaly/  
Class6.1: 0.916  
score: 5  
human_label: 5
```

The 'Features' panel contains the following information:

```
Residual_90: 2.58  
Residual_80: 23.2  
Residual_70: 12.9  
Residual_60: 8.1  
Residual_50: 5.38  
Residual_0: 3.19
```

Below the image is a rating scale from 0 to 5. The scale is currently set to 5. Below the scale is a dropdown menu for 'Raw anomaly score' and a checkbox for 'Show unlabelled objects first'. At the bottom right are two buttons: 'DELETE LABELS' and 'RETRAIN'.



# Galaxy Zoo - Active Learning

The screenshot displays the Galaxy Zoo Active Learning web interface. At the top, a browser window shows the URL 127.0.0.1:5000. The interface has two tabs: "ANOMALY SCORING" (active) and "VISUALISATION".

The central part of the interface features a large image of a galaxy with a bright orange core and blueish-purple spiral arms. Below the image is a progress indicator showing "0" out of "Index" items, with left and right navigation arrows.

Below the progress indicator is a question: "HOW INTERESTING IS THIS OBJECT?". There are five buttons labeled 0, 1, 2, 3, 4, and 5. The button labeled 5 is highlighted in blue, indicating the current score.

At the bottom left, there is a dropdown menu for "Active learning score" with a downward arrow. Below it, the text "Scoring method to sort by" is visible.

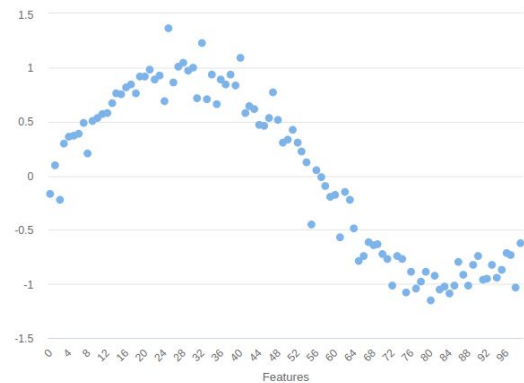
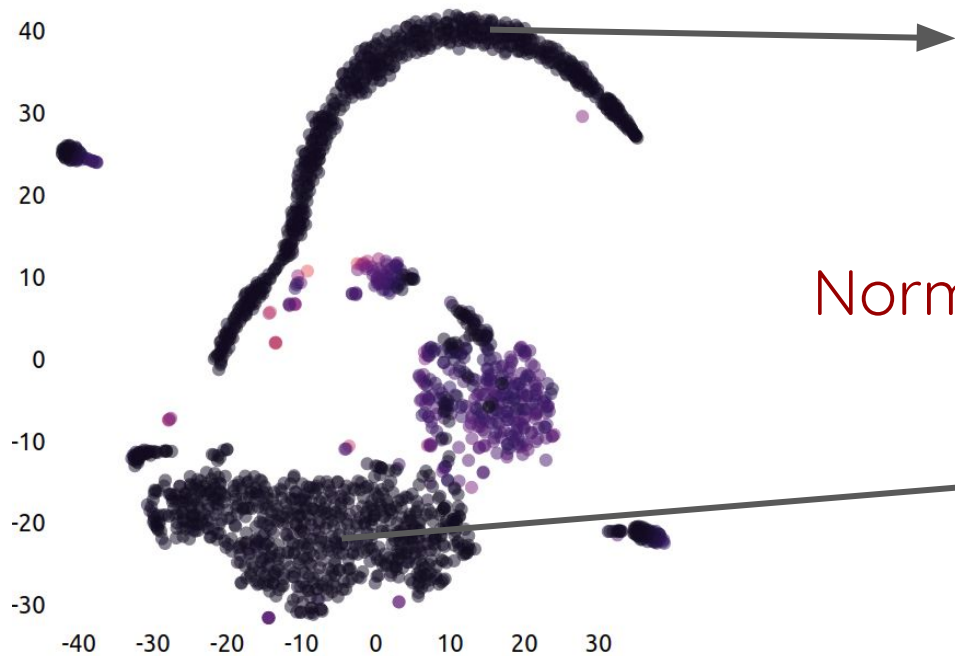
At the bottom center, there is a toggle switch labeled "Show unlabelled objects first" which is currently turned off.

At the bottom right, there are two blue buttons: "DELETE LABELS" and "RETRAIN".

On the right side of the interface, there are two panels:

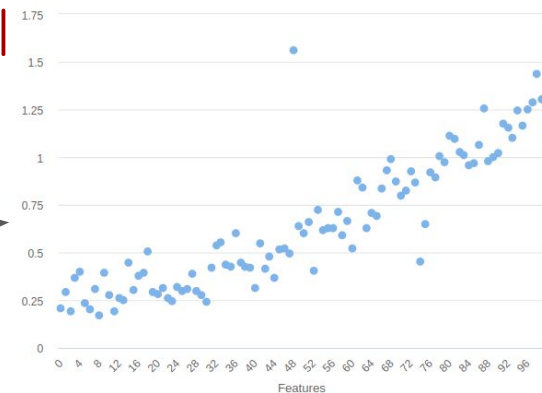
- Metadata:**
  - filename: /home/michelle/BigData/Anoma
  - Class6.1: 0.916
  - score: 5
  - human\_label: 5
  - predicted\_user\_score: 4.61
- Features:**
  - Residual\_90: 2.58
  - Residual\_80: 23.2
  - Residual\_70: 12.9
  - Residual\_60: 8.1
  - Residual\_50: 5.38
  - Residual\_0: 3.19

# Synthetic data



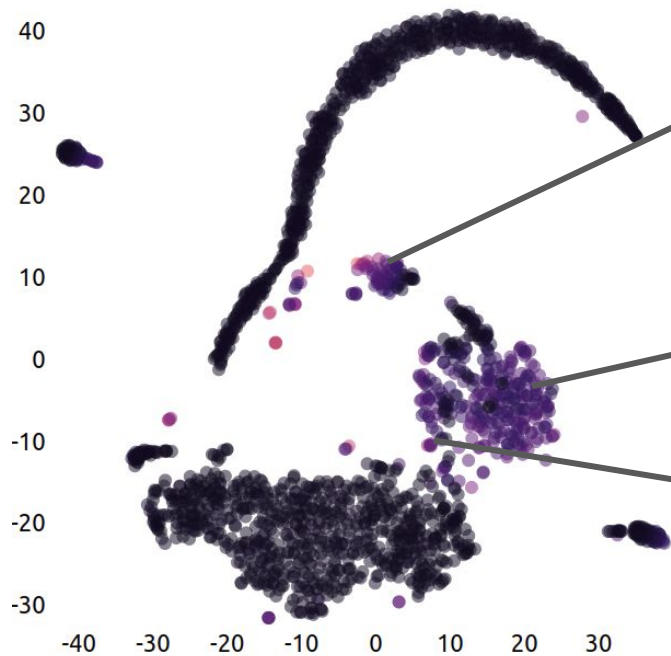
Anomaly Score: 0.270122

Normal

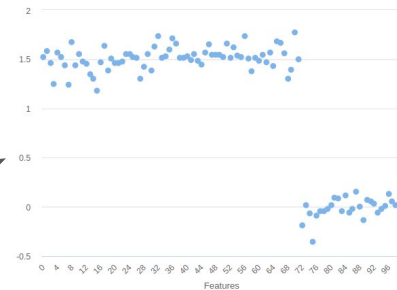


Anomaly Score: 0.241527

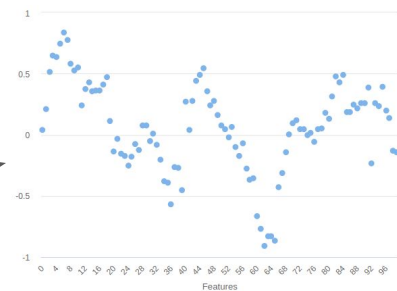
# Synthetic data



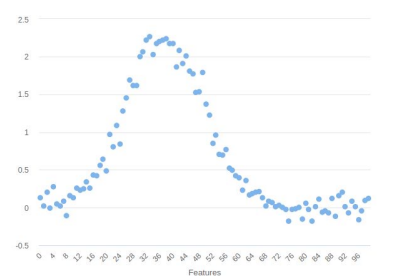
Anomalous



Anomaly Score: 3.112012

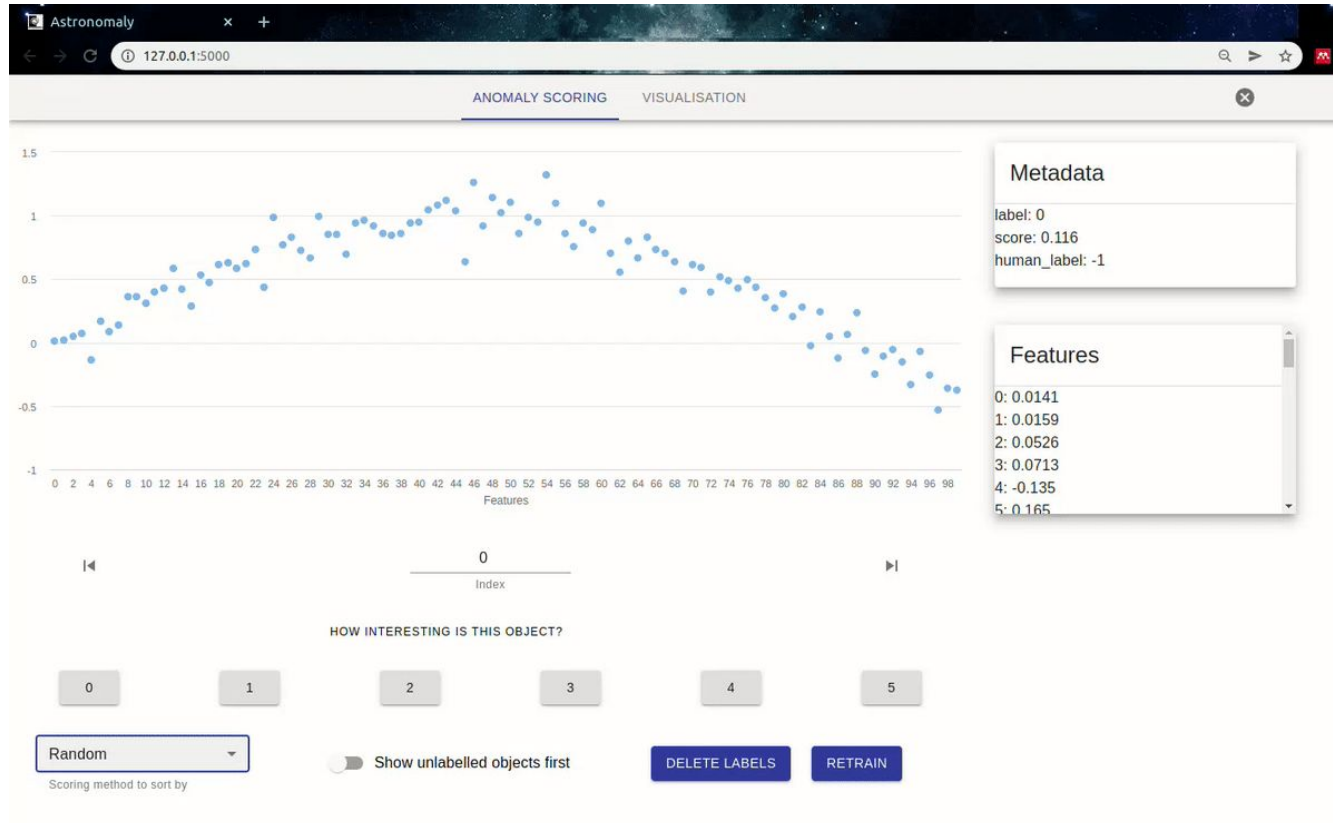


Anomaly Score: 1.303772

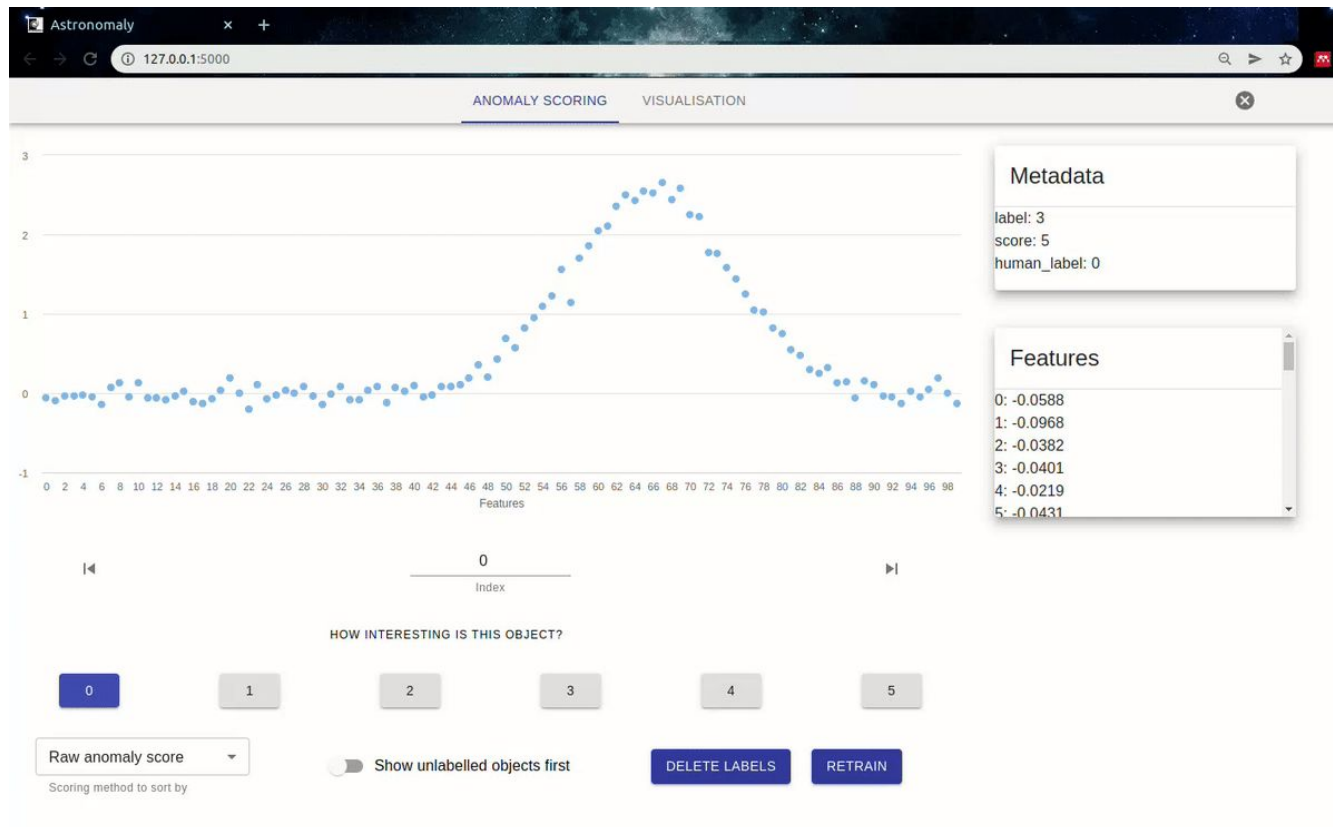


Anomaly Score: 3.92992

# Synthetic data - Random

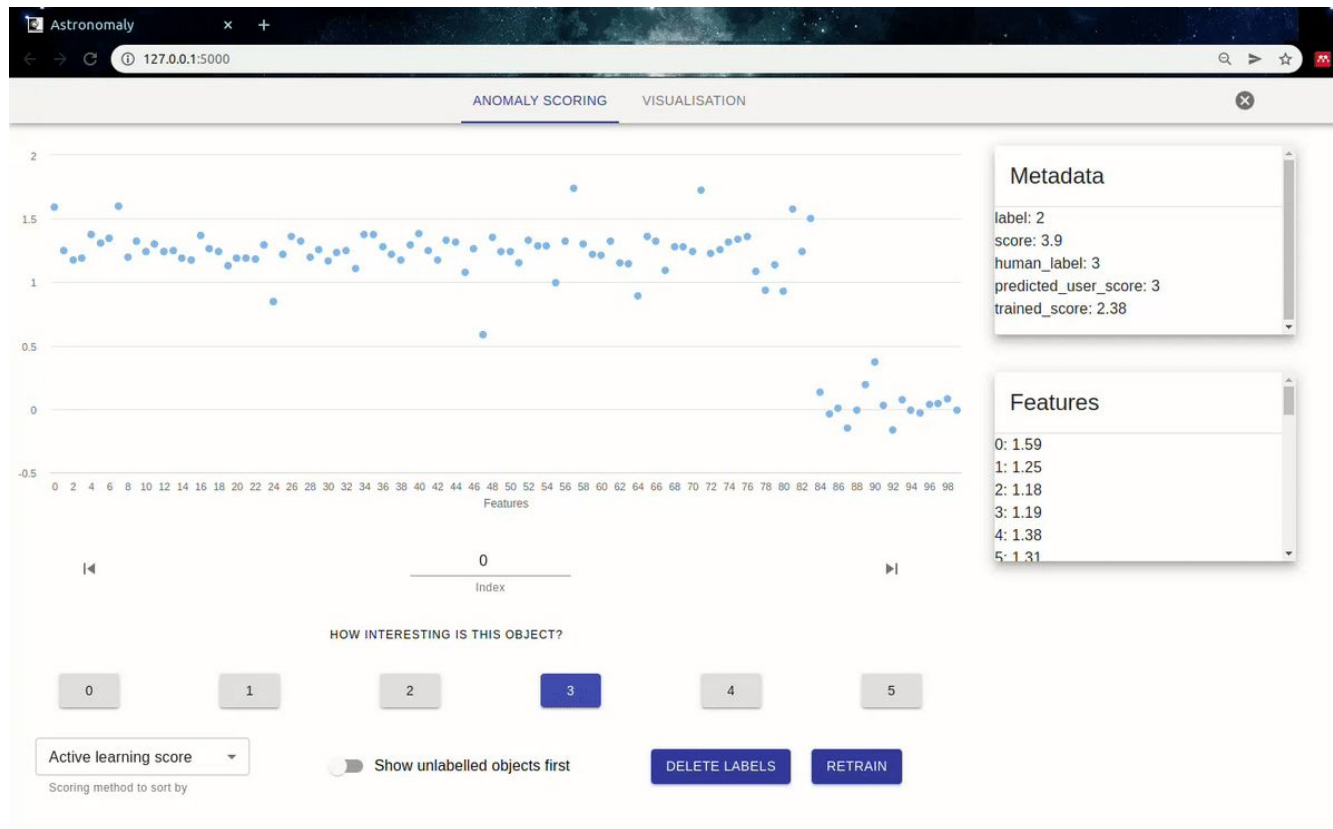


# Synthetic data - No active Learning



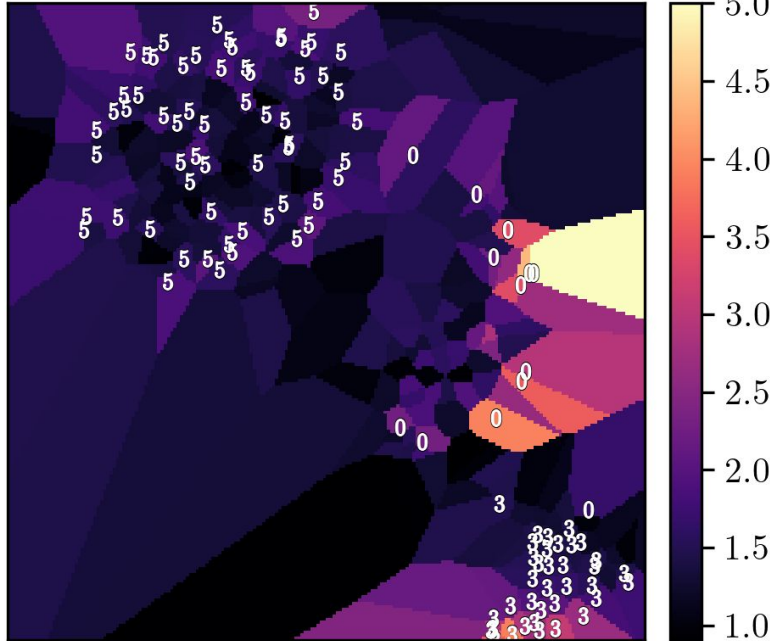


# Synthetic data - Active Learning

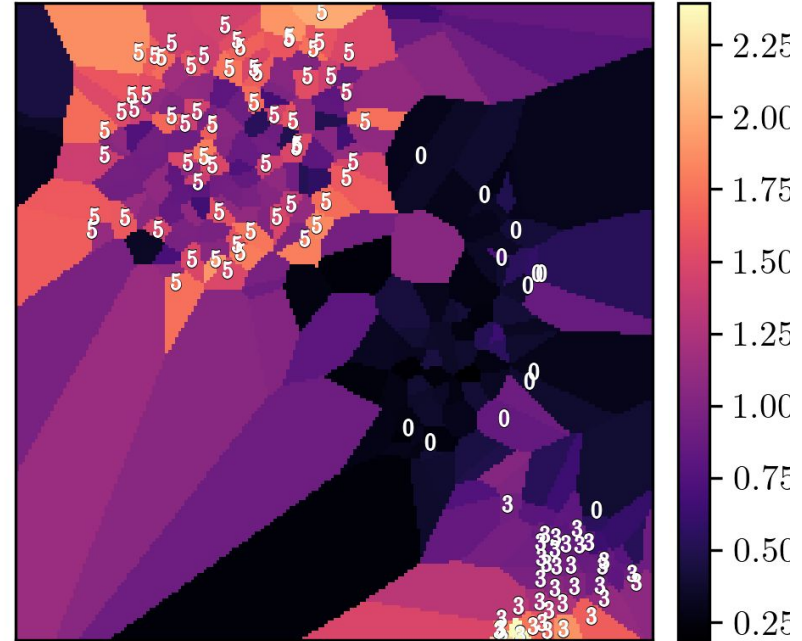


# Visualisation with Synthetic Data

Raw anomaly score

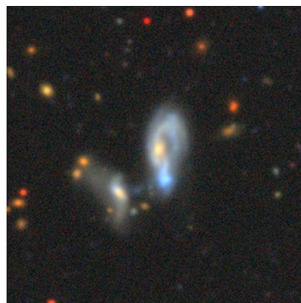
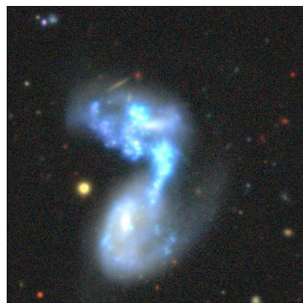
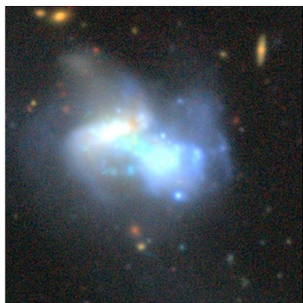
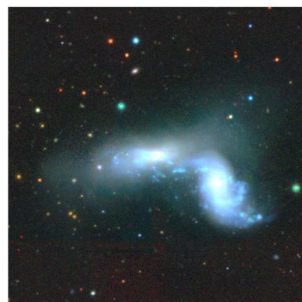
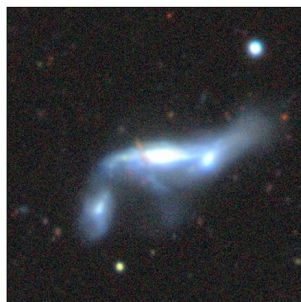
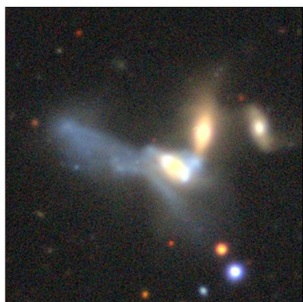


Trained anomaly score



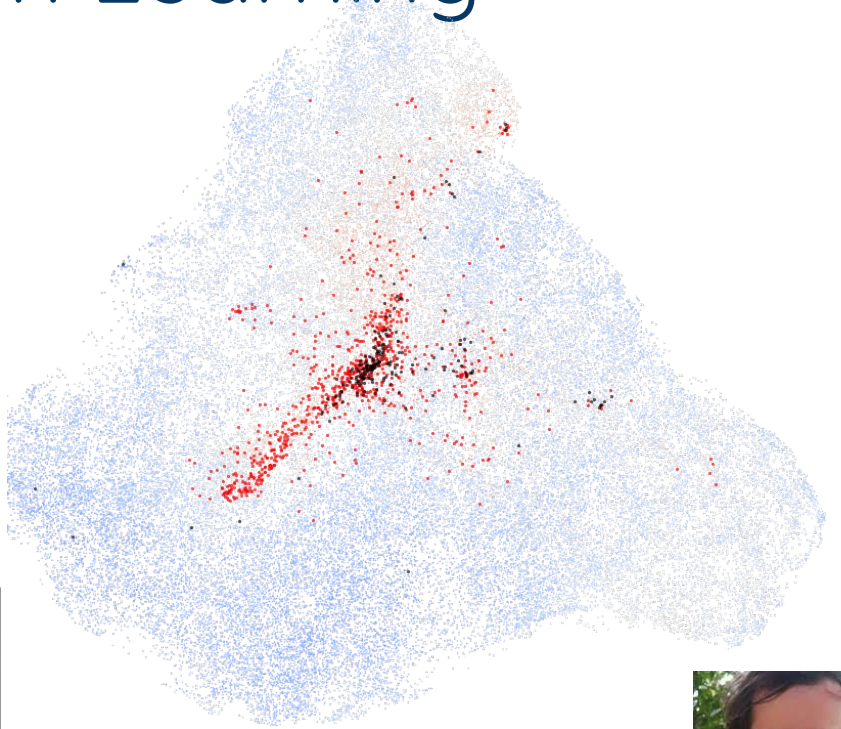
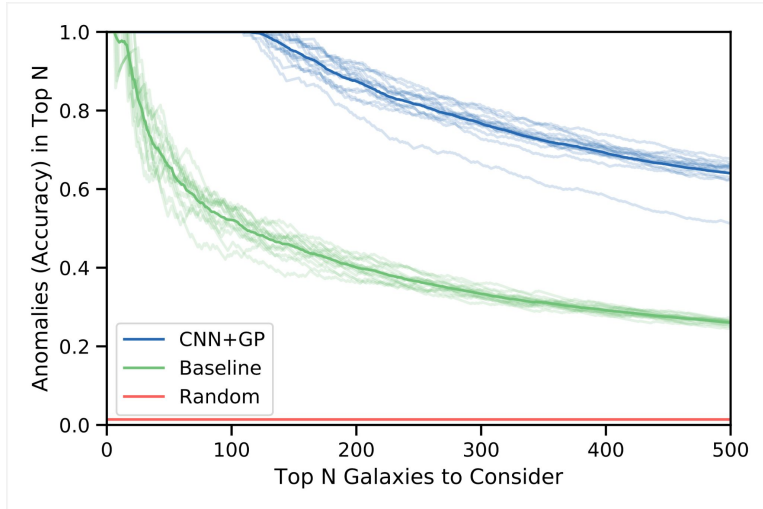
# Anomalies in DECAALS

Verlon Etsebeth (MSc student)





# Deep Representation Learning

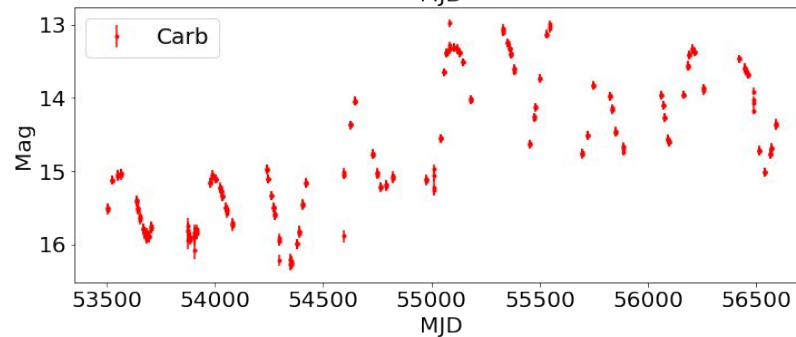
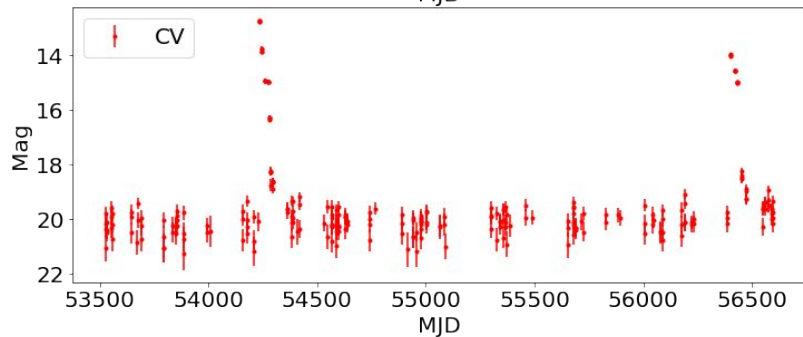
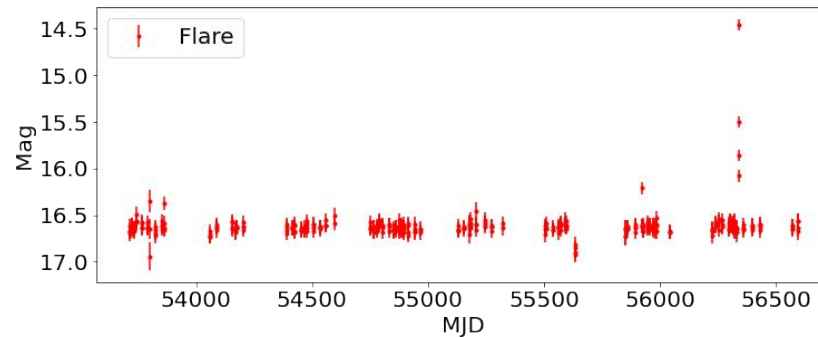
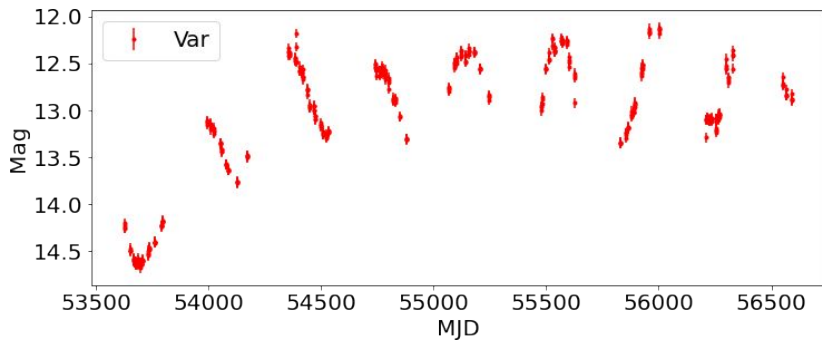


Walmsley et al. (2021) [2110.12735](#)

# Anomalous Transients

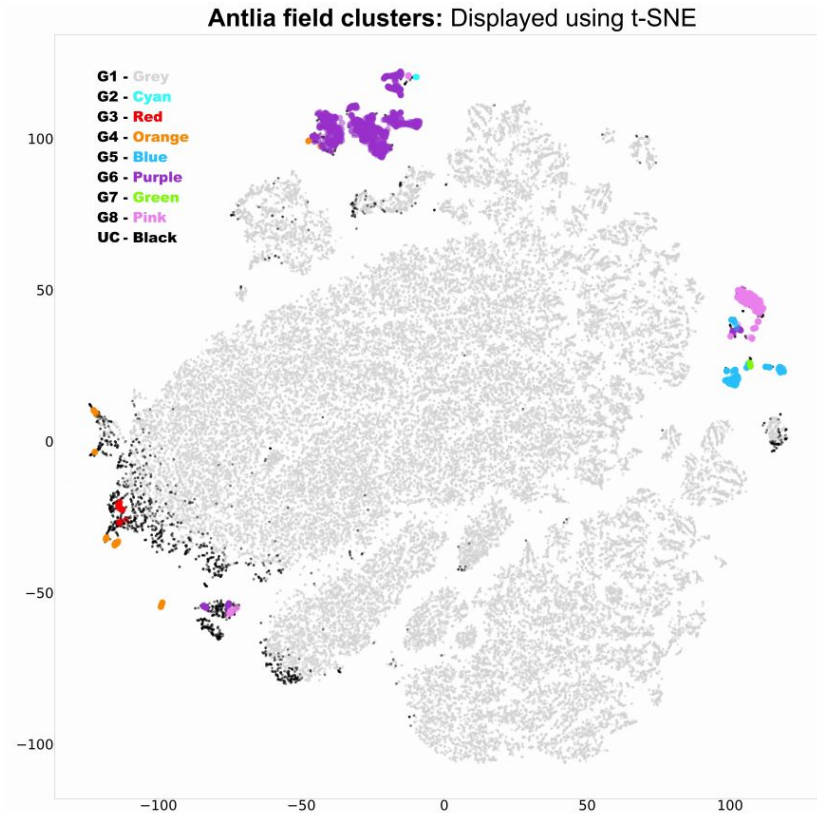


Malema Ramonyai (MSc student)

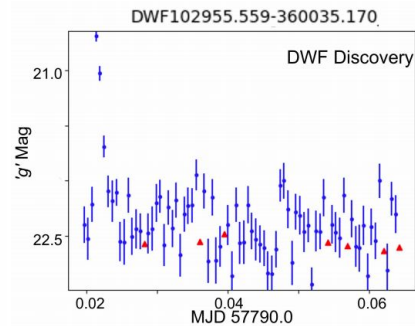
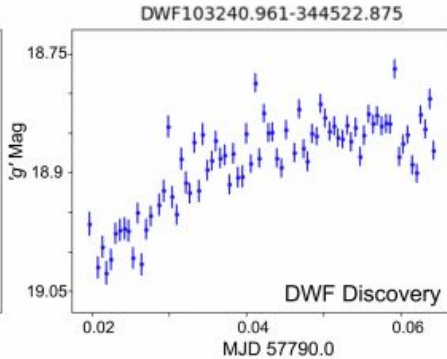
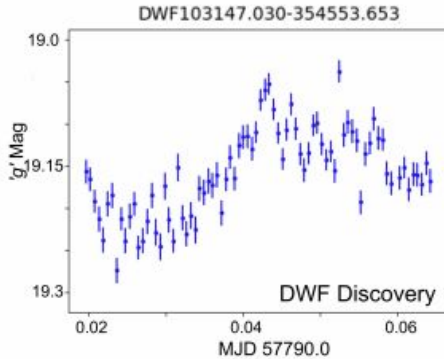
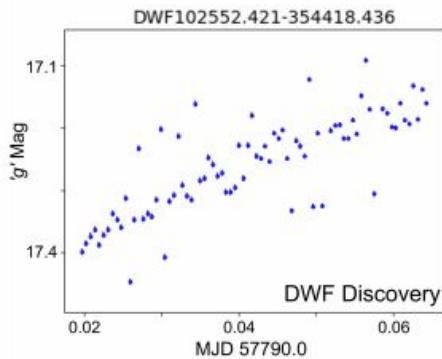
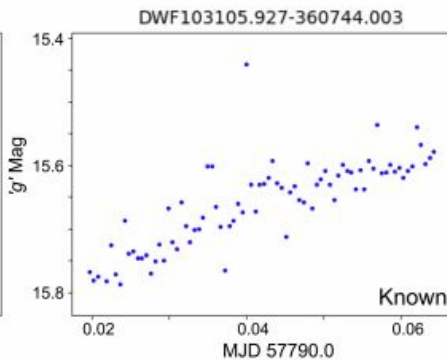
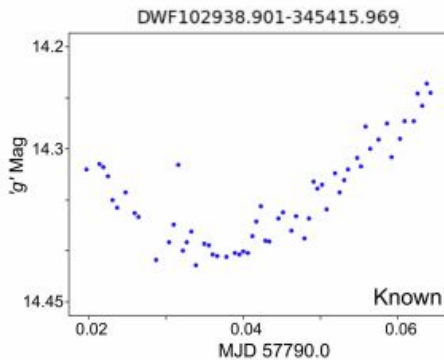
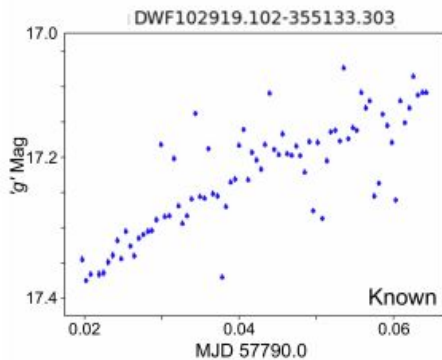




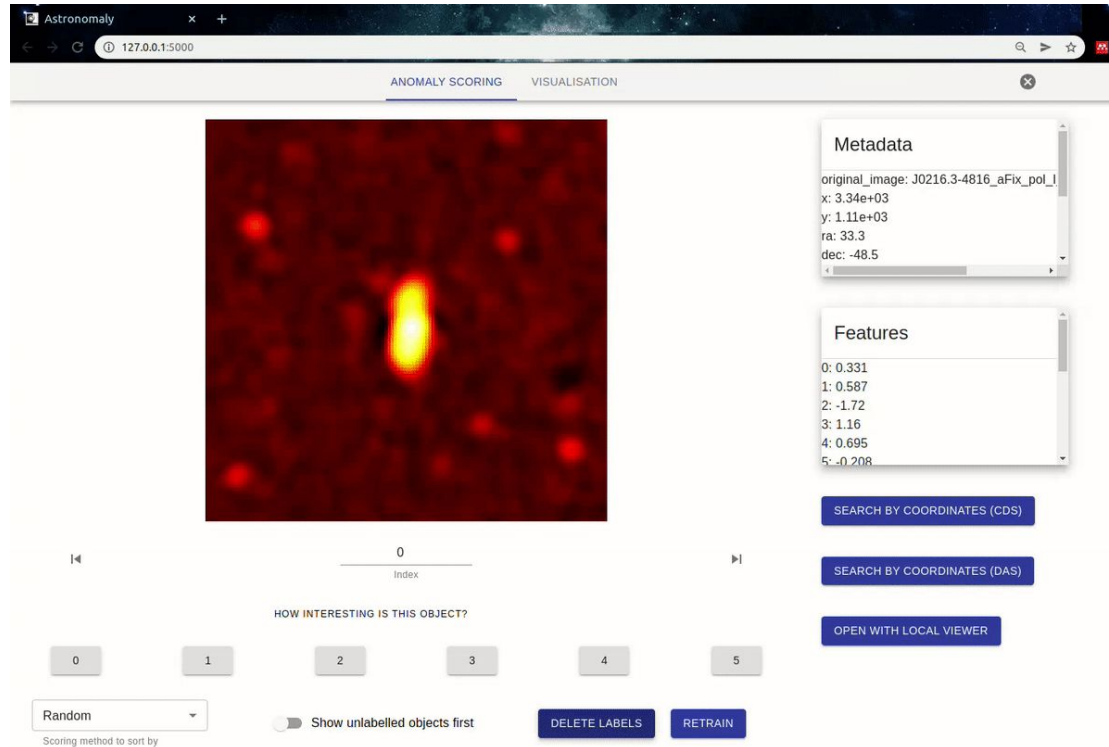
# Astronomy Applied to DWF



# Astrometry Applied to DWF



# MeerKAT Data - Random Examples



The screenshot displays the 'Astronomy' web application interface. The browser address bar shows '127.0.0.1:5000'. The application has two tabs: 'ANOMALY SCORING' (active) and 'VISUALISATION'. The main content area features a large red astronomical image with a central bright yellow and white object. Below the image is a slider labeled 'Index' with a value of 0. A section titled 'HOW INTERESTING IS THIS OBJECT?' contains a row of buttons labeled 0 through 5. Below this is a dropdown menu set to 'Random' and a toggle switch for 'Show unlabelled objects first'. At the bottom are 'DELETE LABELS' and 'RETRAIN' buttons. On the right side, there are two panels: 'Metadata' and 'Features'. The 'Metadata' panel lists: original\_image: J0216.3-4816\_aFix\_pol\_I, x: 3.34e+03, y: 1.11e+03, ra: 33.3, dec: -48.5. The 'Features' panel lists: 0: 0.331, 1: 0.587, 2: -1.72, 3: 1.16, 4: 0.695, 5: -0.208. Below these panels are three buttons: 'SEARCH BY COORDINATES (CDS)', 'SEARCH BY COORDINATES (DAS)', and 'OPEN WITH LOCAL VIEWER'.

Metadata

```
original_image: J0216.3-4816_aFix_pol_I
x: 3.34e+03
y: 1.11e+03
ra: 33.3
dec: -48.5
```

Features

```
0: 0.331
1: 0.587
2: -1.72
3: 1.16
4: 0.695
5: -0.208
```

SEARCH BY COORDINATES (CDS)

SEARCH BY COORDINATES (DAS)

OPEN WITH LOCAL VIEWER

Index: 0

HOW INTERESTING IS THIS OBJECT?

0 1 2 3 4 5

Random

Show unlabelled objects first

DELETE LABELS RETRAIN

# MeerKAT Data - Anomalies

The screenshot displays the 'Astronomy' web application interface. The browser address bar shows the URL '127.0.0.1:5000'. The application has two tabs: 'ANOMALY SCORING' (active) and 'VISUALISATION'. The main content area features a large red-tinted astronomical image of a nebula. Below the image is a slider labeled 'Index' with a value of 0. A section titled 'HOW INTERESTING IS THIS OBJECT?' contains a row of buttons labeled 0 through 5, with button 5 highlighted. Below this is a dropdown menu for 'Active learning score' and a toggle switch for 'Show unlabelled objects first'. At the bottom are buttons for 'DELETE LABELS' and 'RETRAIN'. On the right side, there are two scrollable panels: 'Metadata' and 'Features'. The 'Metadata' panel lists: original\_image: Abell\_85\_aFix\_pol\_I\_Farc, x: 2.05e+03, y: 1.65e+03, ra: 10.4, dec: -9.37. The 'Features' panel lists: Residual\_90: 31.8, Residual\_80: 6.63, Residual\_70: 3.43, Residual\_60: 6.55, Residual\_50: 5.81, Residual\_0: 0.874. Below these panels are three buttons: 'SEARCH BY COORDINATES (CDS)', 'SEARCH BY COORDINATES (DAS)', and 'OPEN WITH LOCAL VIEWER'.

**Metadata**

- original\_image: Abell\_85\_aFix\_pol\_I\_Farc
- x: 2.05e+03
- y: 1.65e+03
- ra: 10.4
- dec: -9.37

**Features**

- Residual\_90: 31.8
- Residual\_80: 6.63
- Residual\_70: 3.43
- Residual\_60: 6.55
- Residual\_50: 5.81
- Residual\_0: 0.874

SEARCH BY COORDINATES (CDS)

SEARCH BY COORDINATES (DAS)

OPEN WITH LOCAL VIEWER

Index: 0

HOW INTERESTING IS THIS OBJECT?

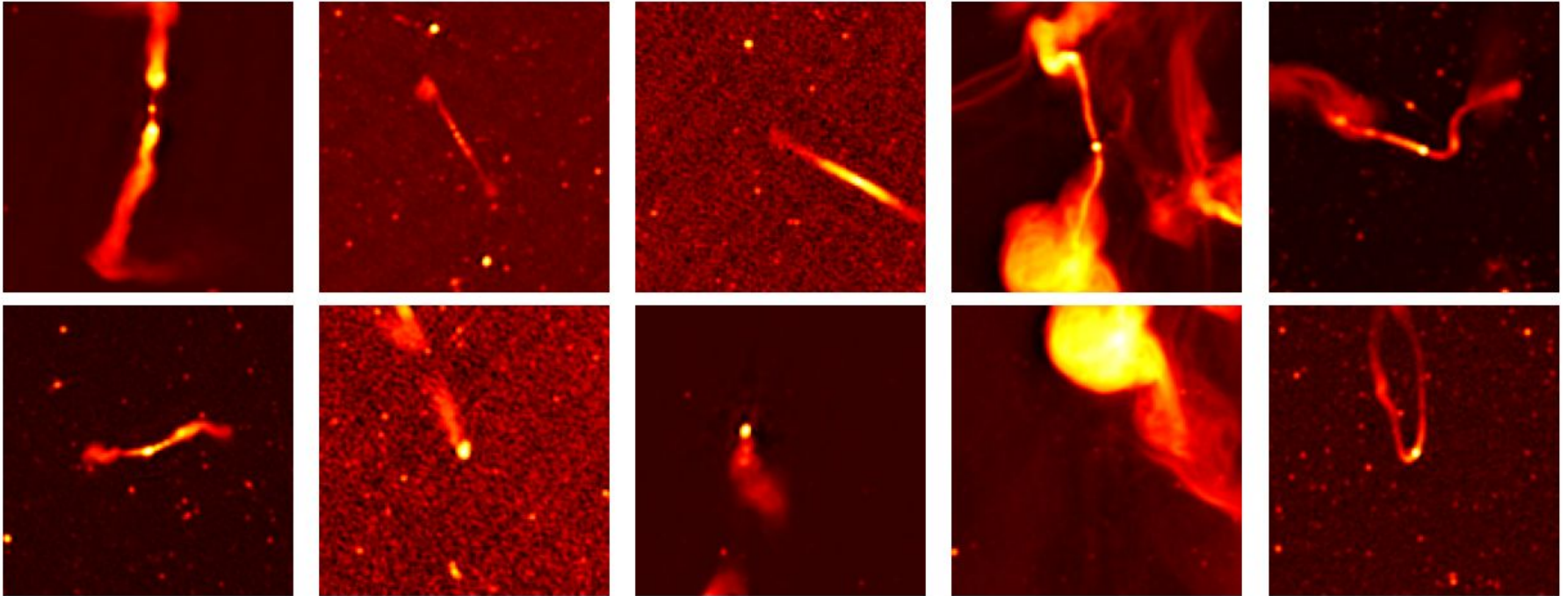
0 1 2 3 4 5

Active learning score

Show unlabelled objects first

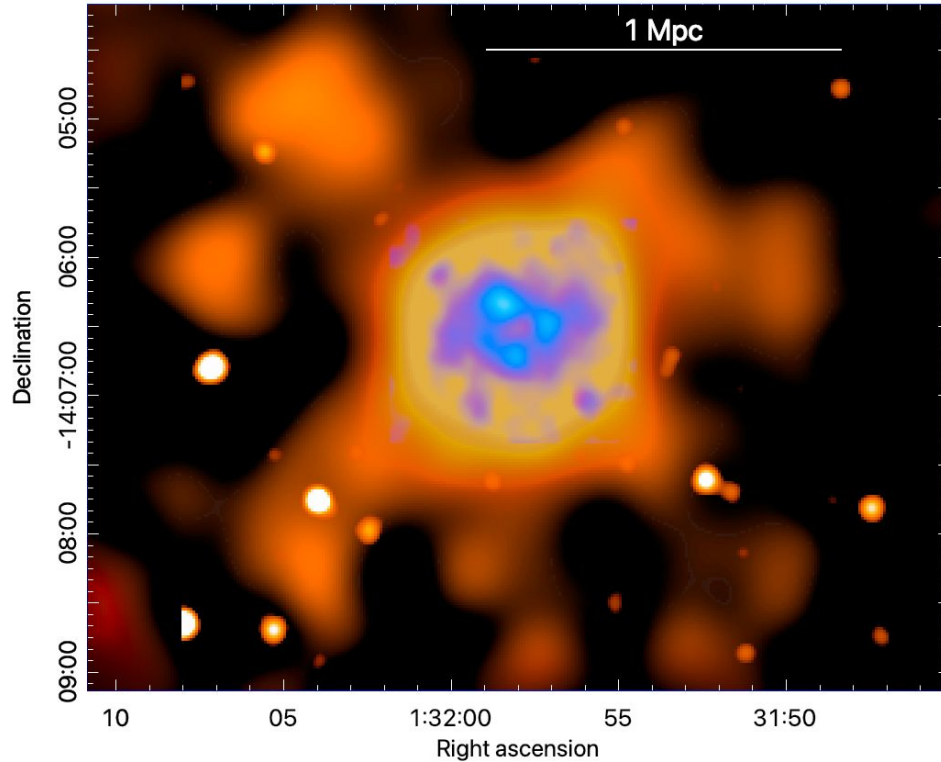
DELETE LABELS RETRAIN

# MeerKAT Galaxy Cluster Legacy Survey





# A genuine machine learning discovery



SANGEETA UJJWAL



**Country:** India  
**Research Field:** Physics  
**Institute:** University of Delhi  
**Position:** Postdoc

SATYA GONTCHO A GONTCHO



**Country:** United States of America  
**Research Field:** Astrophysics and Cosmology  
**Institute:** Lawrence Berkeley National Laboratory  
**Position:** Project Scientist

SHAZRENE S. MOHAMED



**Country:** South Africa  
**Research Field:** Astronomy  
**Institute:** South African Astronomical Observatory and University of Cape Town  
**Position:** Faculty

SIPHEPHILE NCUBE



**Country:** South Africa  
**Research Field:** Physics  
**Institute:** University of the Witwatersrand  
**Position:** Postdoc

SIYI ZHOU



**Country:** Sweden  
**Research Field:** Theoretical Physics  
**Institute:** Stockholm University  
**Position:** Postdoc

SUDESHNA BORO SAIKIA



**Country:** Austria  
**Research Field:** Astrophysics and Cosmology  
**Institute:** University of Vienna  
**Position:** Post doctoral researcher

SWARNAMALA SIRSI



**Country:** India  
**Research Field:** Theoretical Physics  
**Institute:** University of Mysore  
**Position:** Associate Professor (retired)

VALERIA PETTORINO



**Country:** France  
**Research Field:** Astrophysics and Cosmology  
**Institute:** CEA  
**Position:** STAFF Scientist

# Conclusions

- **Machine learning** is critical in facing the data deluge
- We need automated **anomaly detection** to ensure scientific discoveries in datasets aren't missed
- Check out **Astronomy**:
  - <https://arxiv.org/abs/2010.11202>
  - <https://github.com/MichelleLochner/astronomy>
- And the Supernova Foundation:
  - <https://www.supernovafoundation.org/>