A Convolutional Autoencoder-Based Pipeline for Anomaly Detection and Classification of Periodic Variables

Ho-Sang Chan[1,2], Siu-Hei Cheung[1,2], Victoria Ashley Villar[3], Shirley Ho[2]

Machine Learning and the Physical Sciences workshop

Workshop at the 35th Conference on Neural Information Processing Systems
December 13, 2021

1. The Chinese University of Hong Kong
2. Center for Computational Astrophysics, Flatiron institute
3. The Pennsylvania State University

Introduction And Motivation

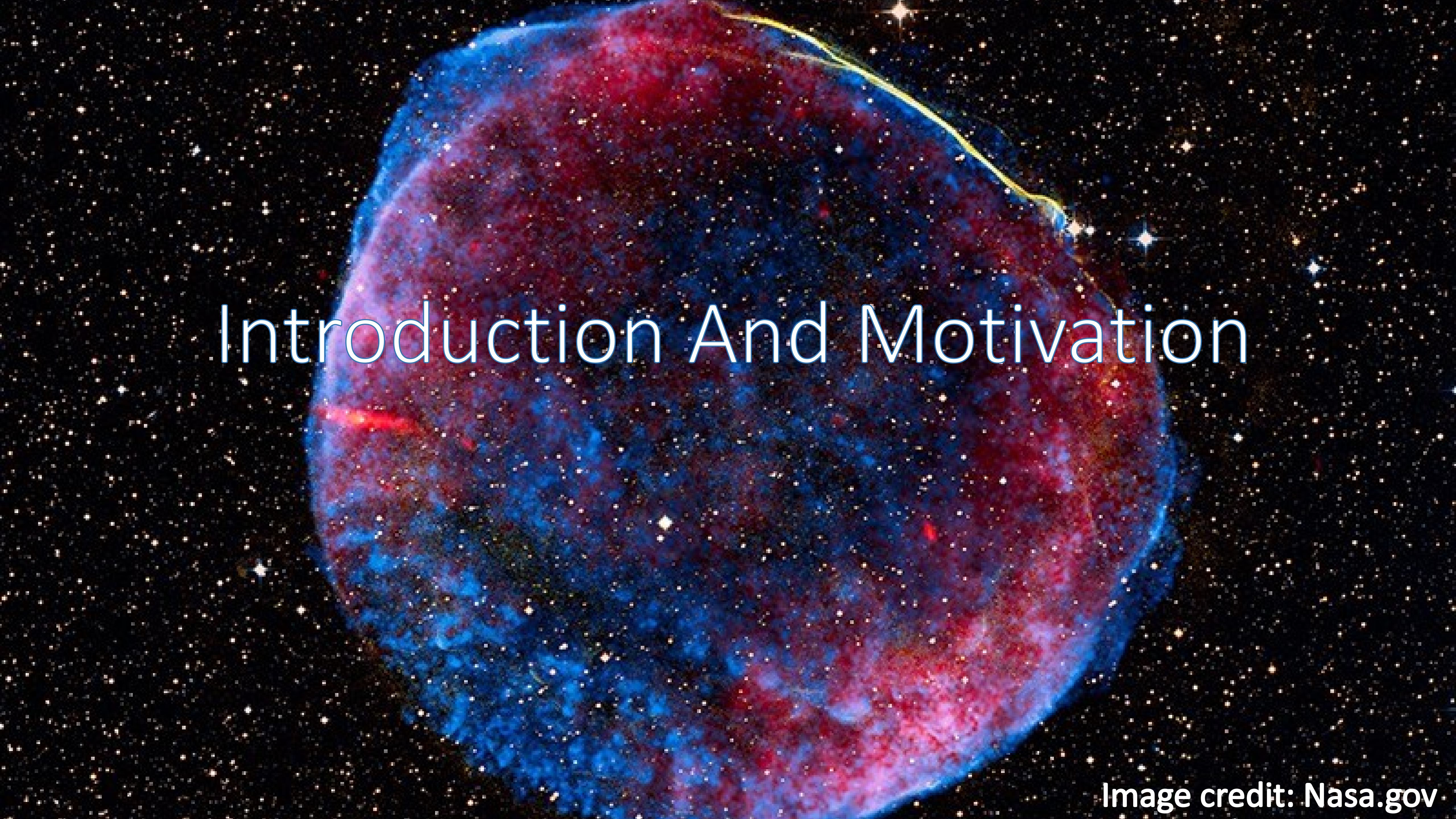


Image credit: Nasa.gov

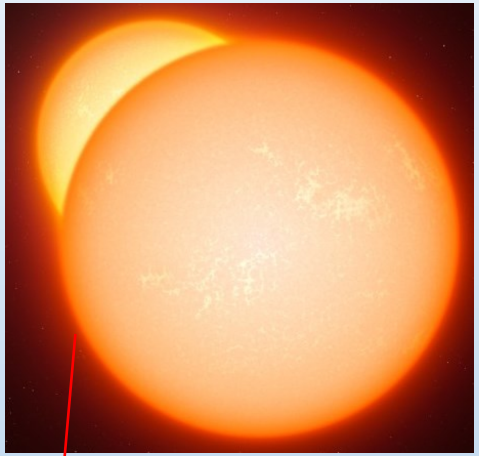
Searching For Anomalous Periodic Variable Stars

Image credit: astro.wisc.edu

Intrinsic

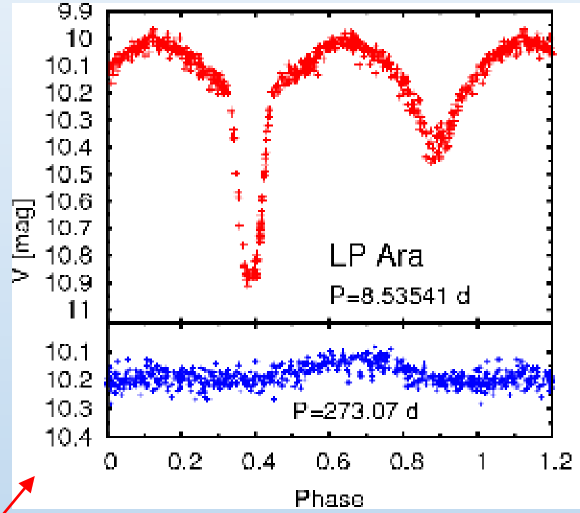
Or

Image credit: newatlas.com



Extrinsic

Image credit: Michalska et al. (2009)



- Periodic variable stars → brightness changes periodically
- Source: Pulsations, eclipsing, and more …
- Physics of sources encoded in their light curves
- Search for wild cats → New discoveries

Methodology

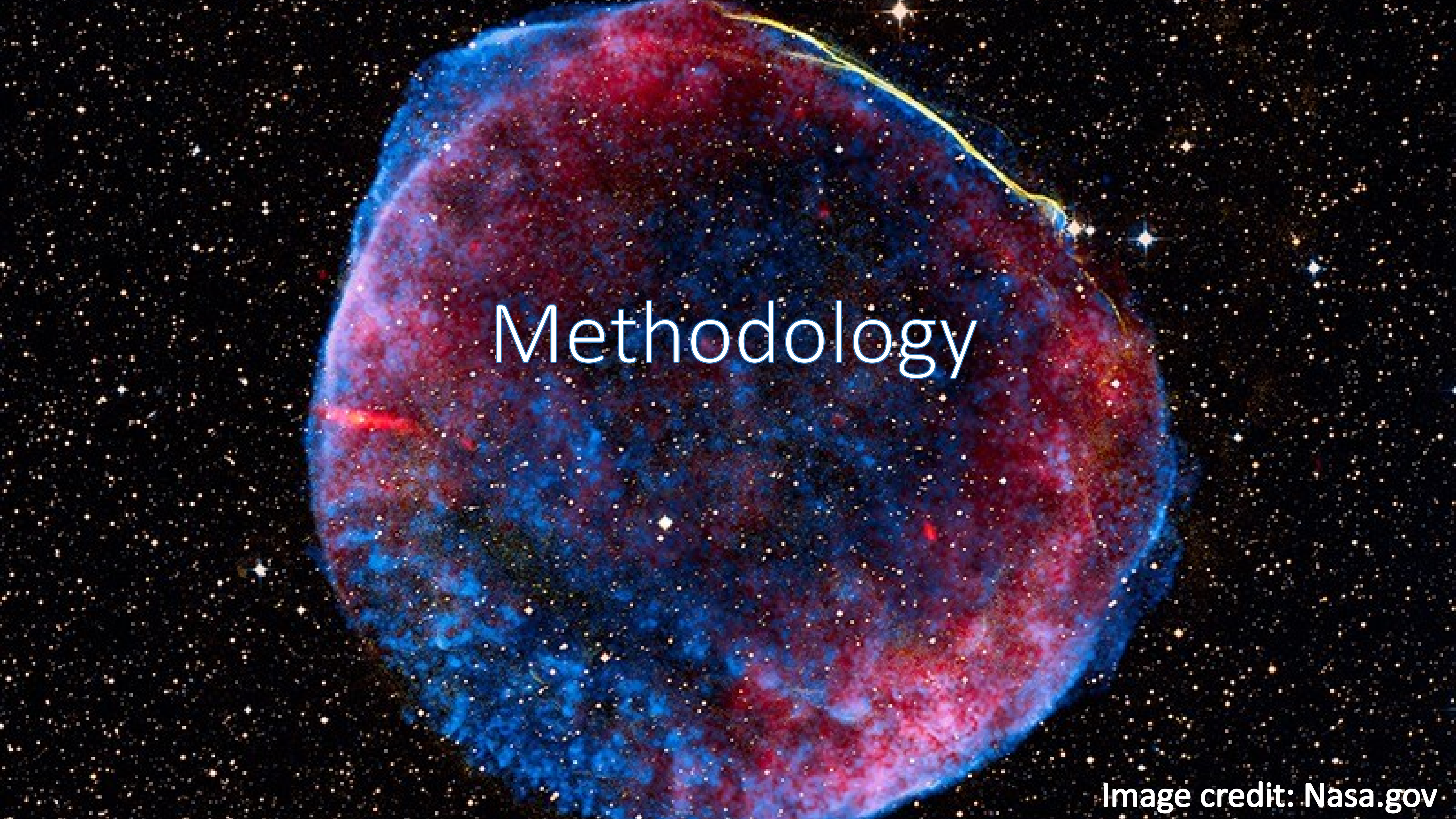


Image credit: Nasa.gov

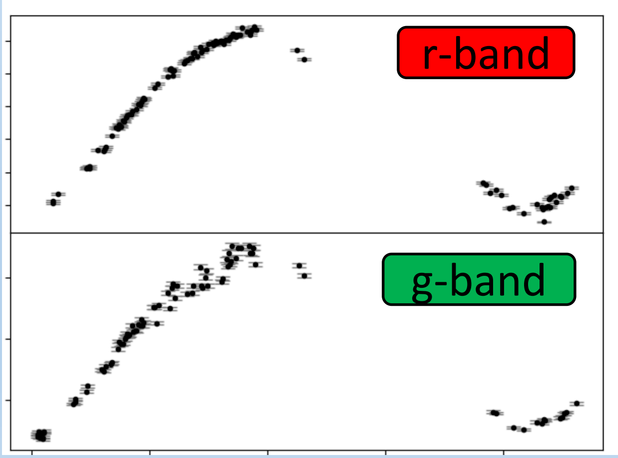
Data Pre-Processing

The Zwicky Transient Facility Catalog of Periodic Variable Stars

Xiaodian Chen[1], Shu Wang[1], Licai Deng[1], Richard de Grijs[2,3,4], Ming Yang[5], and Hao Tian[6]
[1] CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, People's Republic of China
chenxiaodian@nao.cas.cn
[2] Department of Physics and Astronomy, Macquarie University, Balaclava Road, Sydney, NSW 2109, Australia
[3] Research Centre for Astronomy, Astrophysics and Astrophotonics, Macquarie University, Balaclava Road, Sydney, NSW 2109, Australia
[4] International Space Science Institute–Beijing, 1 Nanertiao, Zhongguancun, Hai Dian District, Beijing 100190, People's Republic of China
[5] IAASARS, National Observatory of Athens, Vas. Pavlou & I. Metaxa, Penteli 15236, Greece
[6] CAS Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, People's Republic of China

Received 2019 November 29; revised 2020 June 10; accepted 2020 June 11; published 2020 July 17

Raw Detections



Phase-Folding

Multivariate-GPR

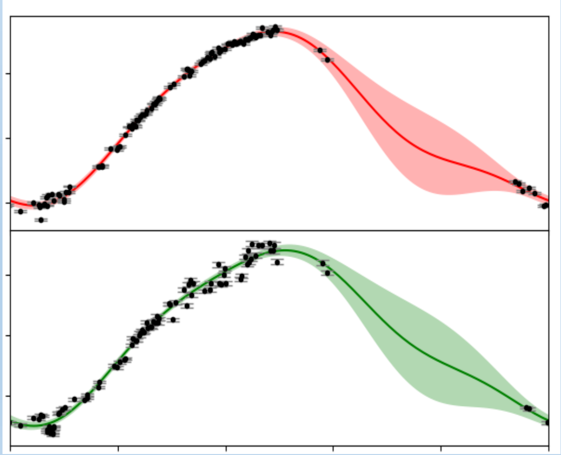
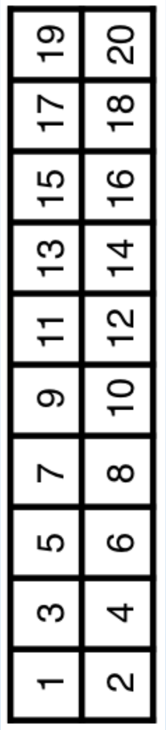


Image credit: wordpress.com

1	3	5	7	9	11	13	15	17	19
2	4	6	8	10	12	14	16	18	20

Stacked "Images"

Latent Features Extractions



Stacked "Images"

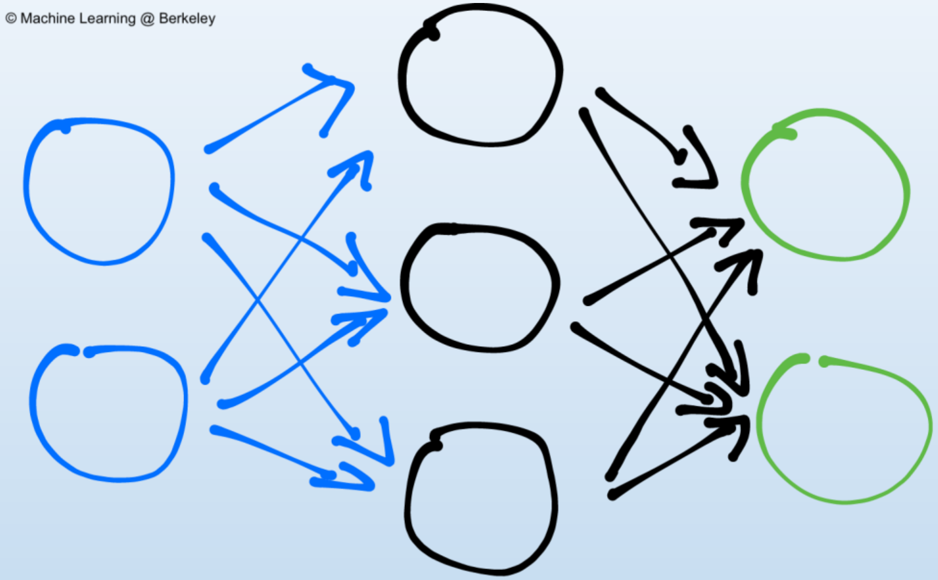
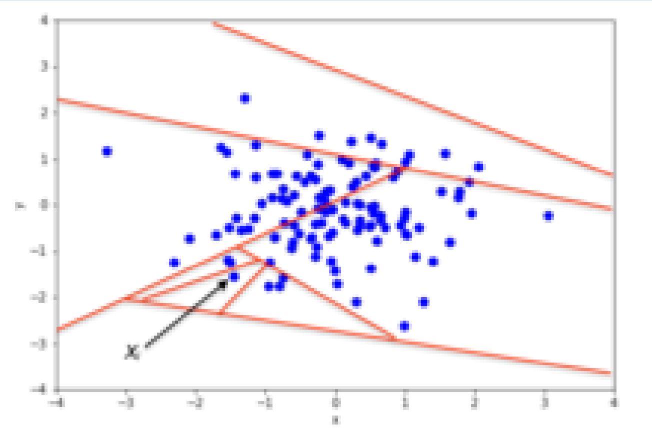


Image credit: UC Berkeley

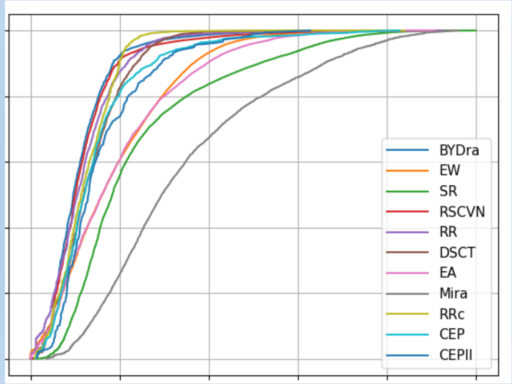
Convolutional AutoEncoder

Latent Features

Image credit: Wikipedia.com



Isolation Forest



List of Anomalies

Results

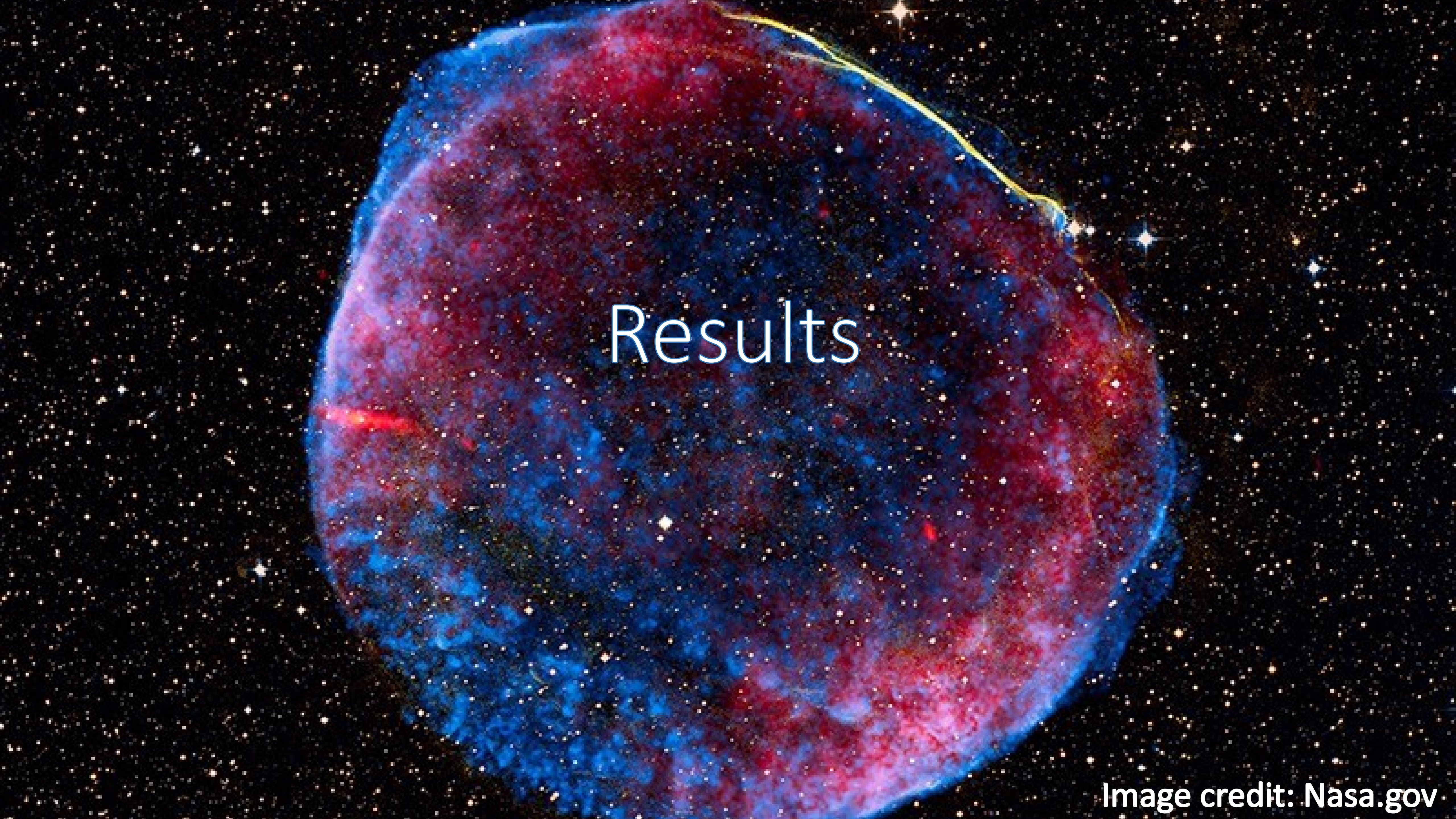
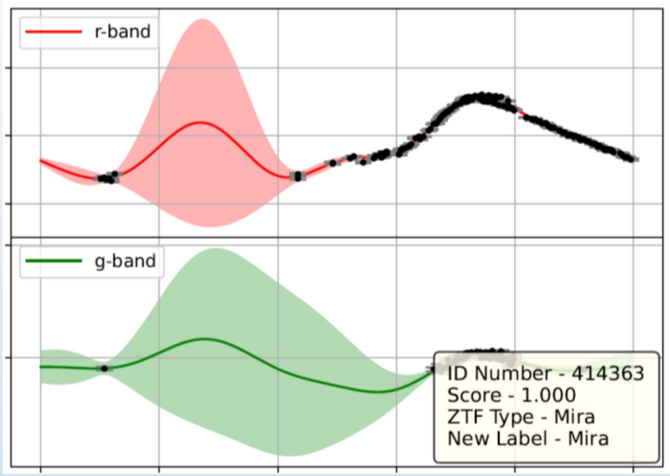


Image credit: Nasa.gov

The Anomalies

- Anomalous periodic variables are
  - Irregular oscillating
  - High variability

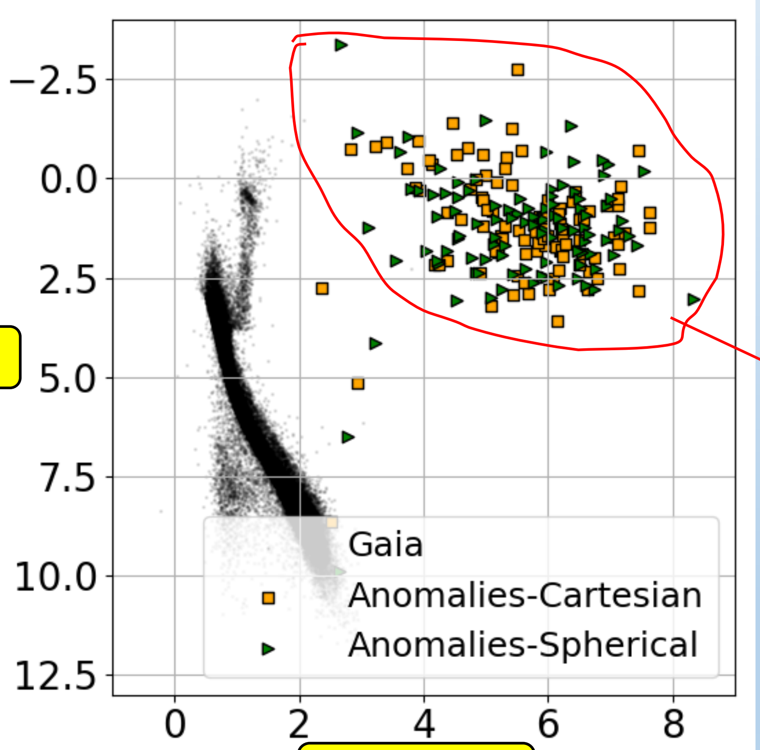
Illustrations



- Plotted HR-Diagram
- Anomalies are
  - Brighter
  - Cooler
- Corresponds to evolved stars In their late phase of evolution

Brighter

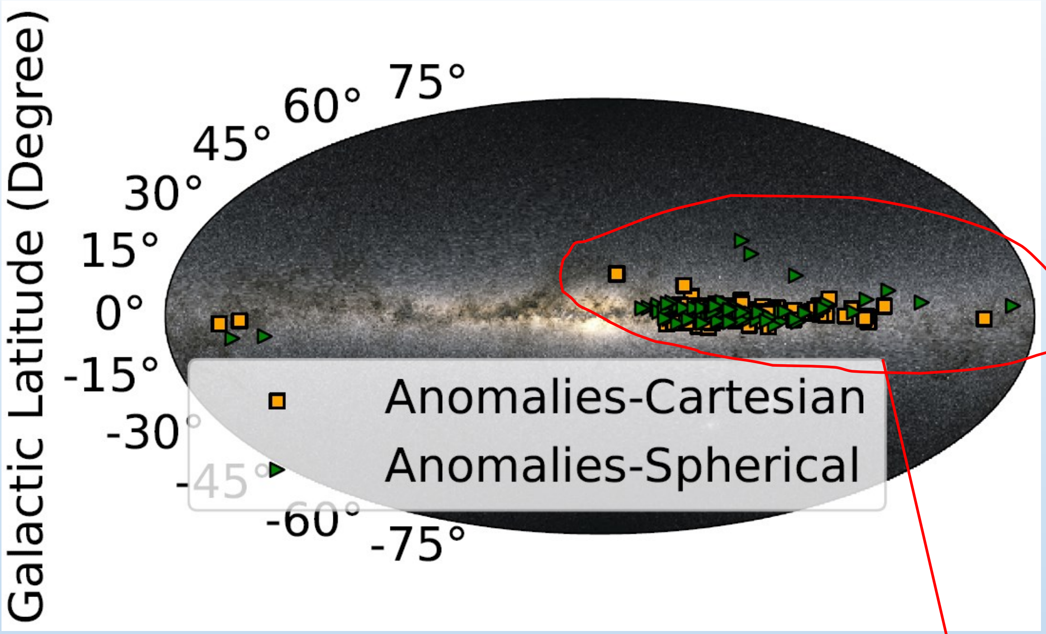
Brightness



Hotness

Cooler

The Anomalies



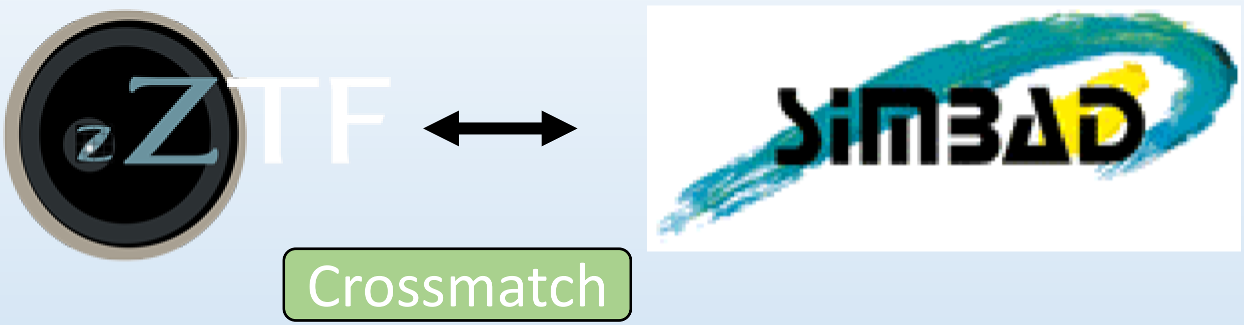
- Located in the vicinity of the Galactic disk
- Younger (with respect to the Galactic age)

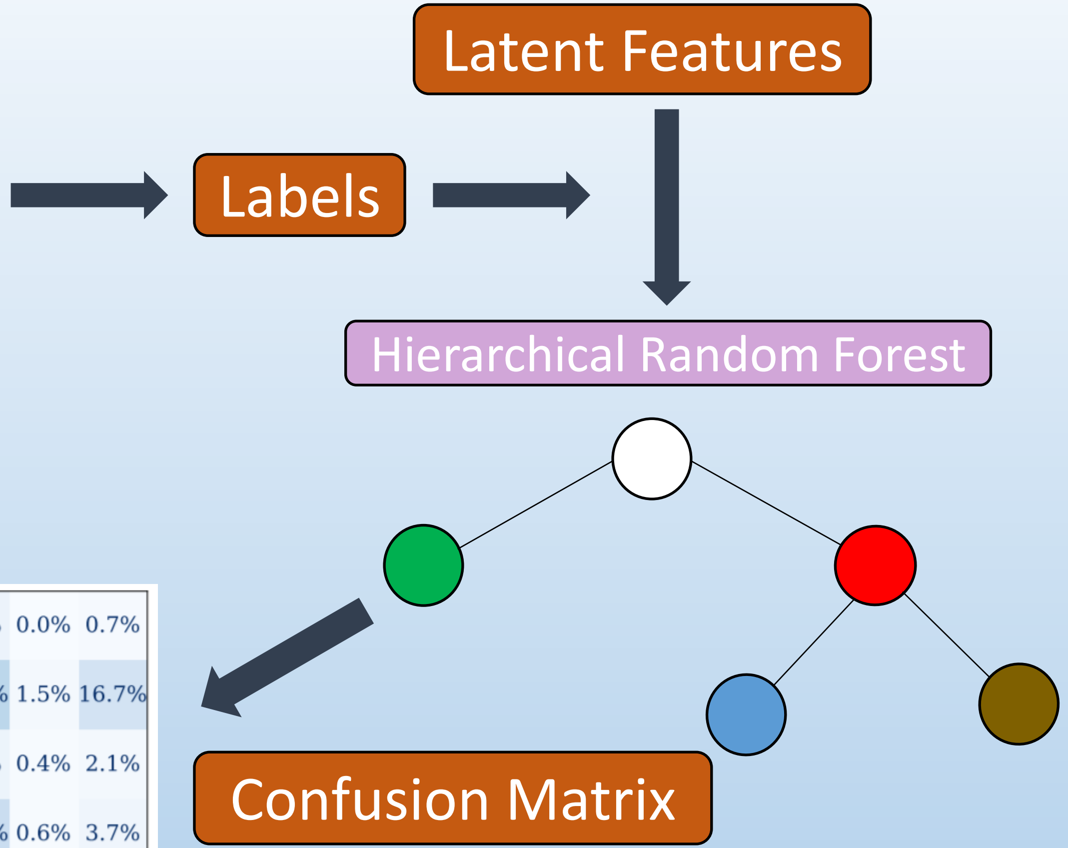
LSST

Image credit: symmetrymagazine.org

Detailed Spectroscopic Follow-Up Is Strongly Recommended!

Classifications Using The SIMBAD Labels





- Class labels from the SIMBAD catalogue
- Reliable but more expensive

True label \ Predicted label	C-Type	HB	RGB	S-Type	Voth	YSOL
C-Type	55.6%	9.6%	1.9%	5.0%	11.9%	16.1%
HB	0.0%	74.1%	0.0%	3.7%	11.1%	11.1%
RGB	6.7%	0.0%	80.0%	0.0%	6.7%	6.7%
S-Type	35.5%	3.2%	6.5%	38.7%	9.7%	6.5%
Voth	11.4%	5.9%	3.0%	3.0%	59.9%	16.9%
YSOL	11.4%	10.3%	1.6%	2.7%	37.5%	36.4%

True label \ Predicted label	AGNL	CEP	EB	LPV	Mira	Pec	Puloth	RR
AGNL	92.6%	0.0%	0.7%	0.7%	0.0%	5.1%	0.0%	0.7%
CEP	0.0%	48.5%	6.1%	1.5%	0.0%	25.8%	1.5%	16.7%
EB	0.1%	0.2%	92.0%	0.4%	0.1%	4.7%	0.4%	2.1%
LPV	1.1%	0.0%	16.7%	51.7%	5.1%	21.2%	0.6%	3.7%
Mira	0.5%	0.0%	19.9%	5.0%	46.3%	24.9%	0.5%	3.0%
Pec	1.2%	1.3%	16.7%	8.8%	2.2%	62.4%	1.2%	6.2%
Puloth	0.0%	2.0%	22.4%	1.0%	1.0%	10.2%	51.0%	12.2%
RR	0.0%	0.0%	4.4%	0.3%	0.2%	2.7%	0.1%	92.3%

- Good accuracy for SOME classes

Conclusion

I showed the application of machine learning in Astronomy for …

1. Detecting anomalous periodic variable stars
2. Building classification model for periodic variable stars

Thank You

Referee Report

- DGMS workshop – One reject one accept – Lucky

Interesting paper applying DGM for anomaly detection in astronomy

➦ Searching for the Weirdest Stars: A Convolutional Autoencoder-Based Pipeline For Detecting Anomalous Periodic Variable Stars

Review:
The paper introduces an interesting method to detect anomalous Periodic Variable Stars. It is definitely within the scope of this workshop, and it makes important contribution to the field of astronomy.

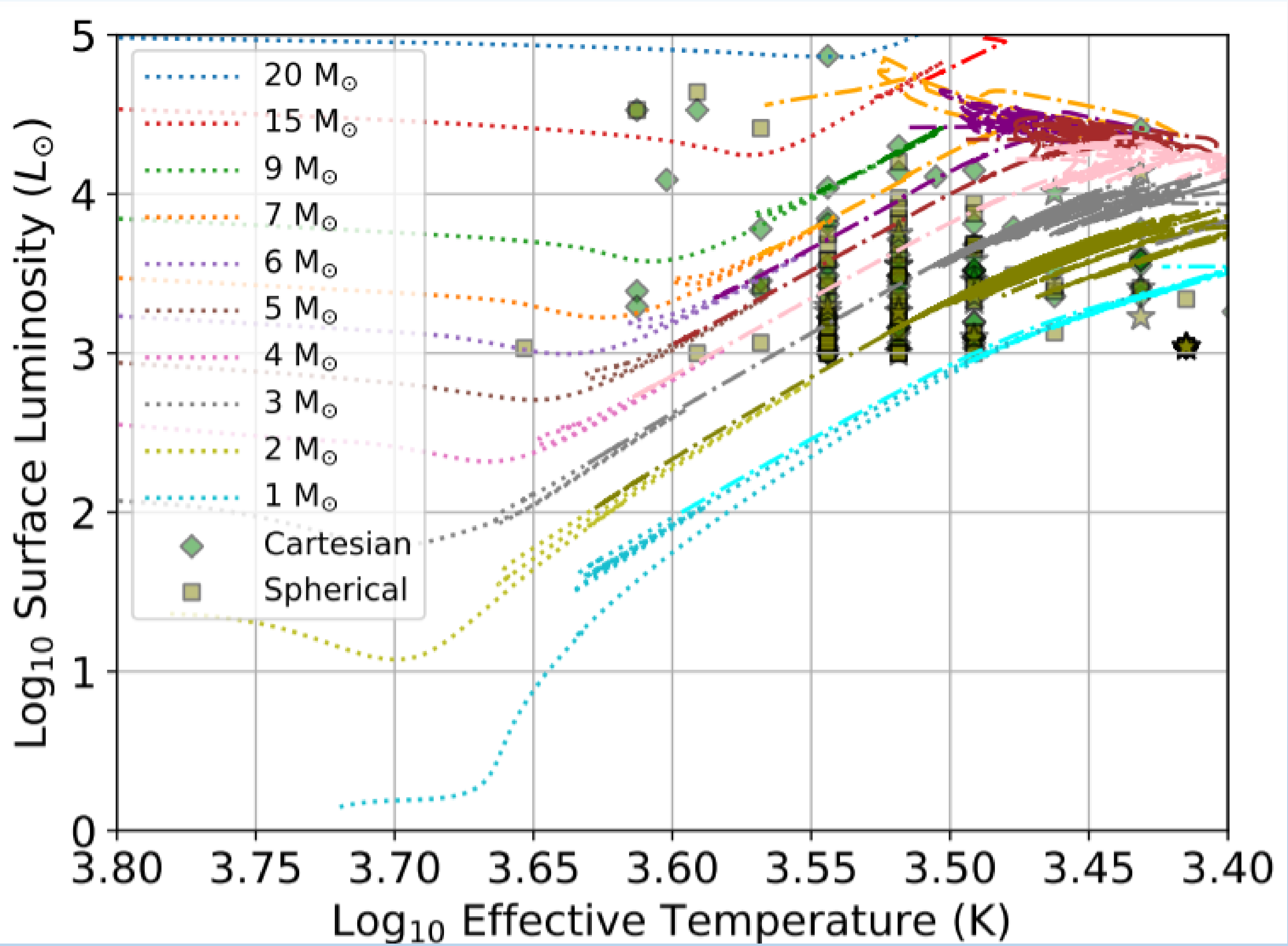
On suggestion is, the current way of detection anomalous data after training the VAE is isolation forest, but it does not introduce why this is a good detection method. In addition, the authors can look into other detection score introduced for VAE or deep neural network in general (such as mahalanobis distance in the feature space).

Rating: 7: Good paper, accept

Confidence: 4: The reviewer is confident but not absolutely certain that the evaluation is correct

- Discuss why we choose isolation forest is needed
- Slight paper editing
- Register for AAS and NeruIPS

HR-Diagram



- Added stellar evolution phases – dotted (RGB) and dash-dotted (AGB

Referee Reports

- Overall, they are satisfied with our work (lol …)
- They are not so happy about us missing important information in the methodology section
- E.g., How do we get the hyper-parameters? Are there any tuning? The specific values of the hyper-parameters?
- Created a google document for the cover letter

3D Dust Map

Table 1: Bayestar19 extinction coefficients (R)

g	r	i	z	y	J	H	K_s
3.518	2.617	1.971	1.549	1.263	0.7927	0.4690	0.3026

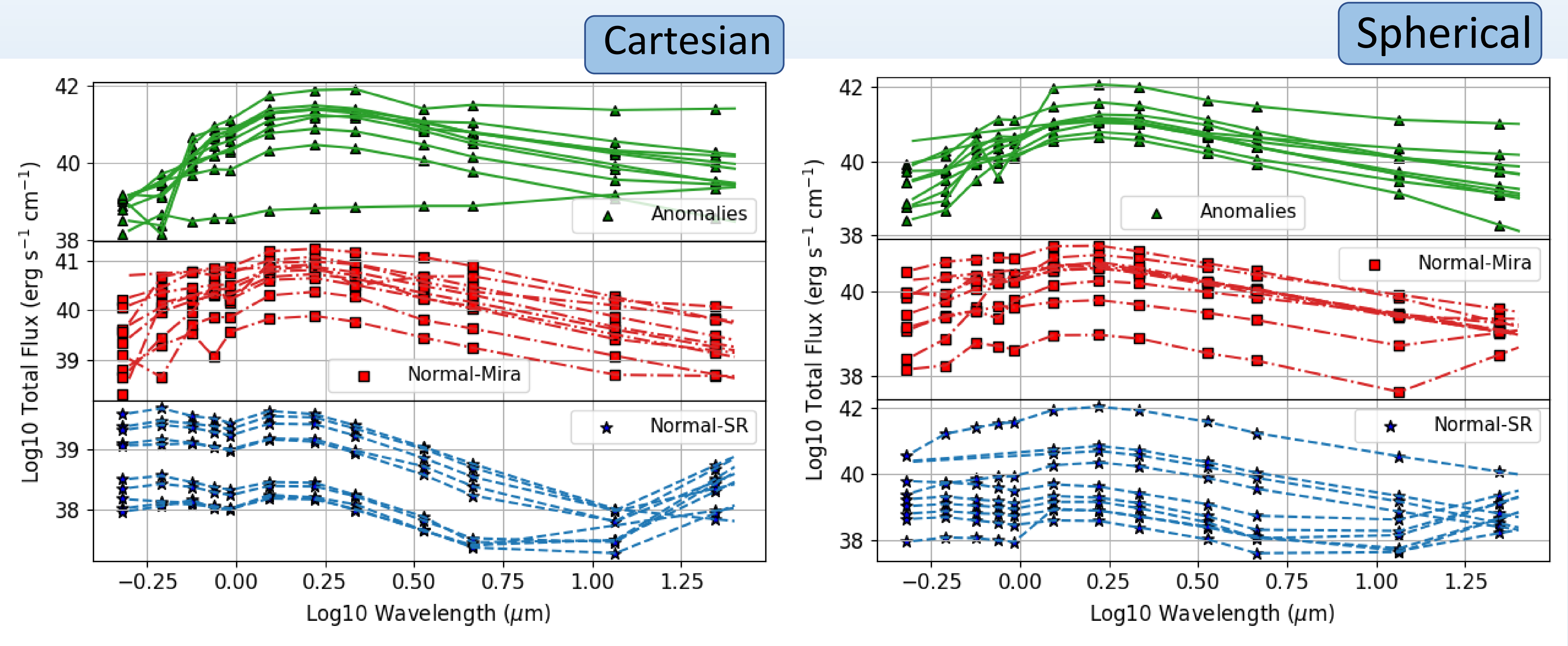
where σ_m and σ_ϖ are the uncertainty in the observed magnitudes and parallax, and $A(E)$ is the extinction in the observed passbands, assumed to be a linear function of reddening, E. We assume that for each star, extinction is given by

$$\vec{A}(E) = E\,\vec{R}\,, \quad (5)$$

where $\vec{R}$ is the "extinction vector," relating a scalar reddening to the extinction in each passband. We

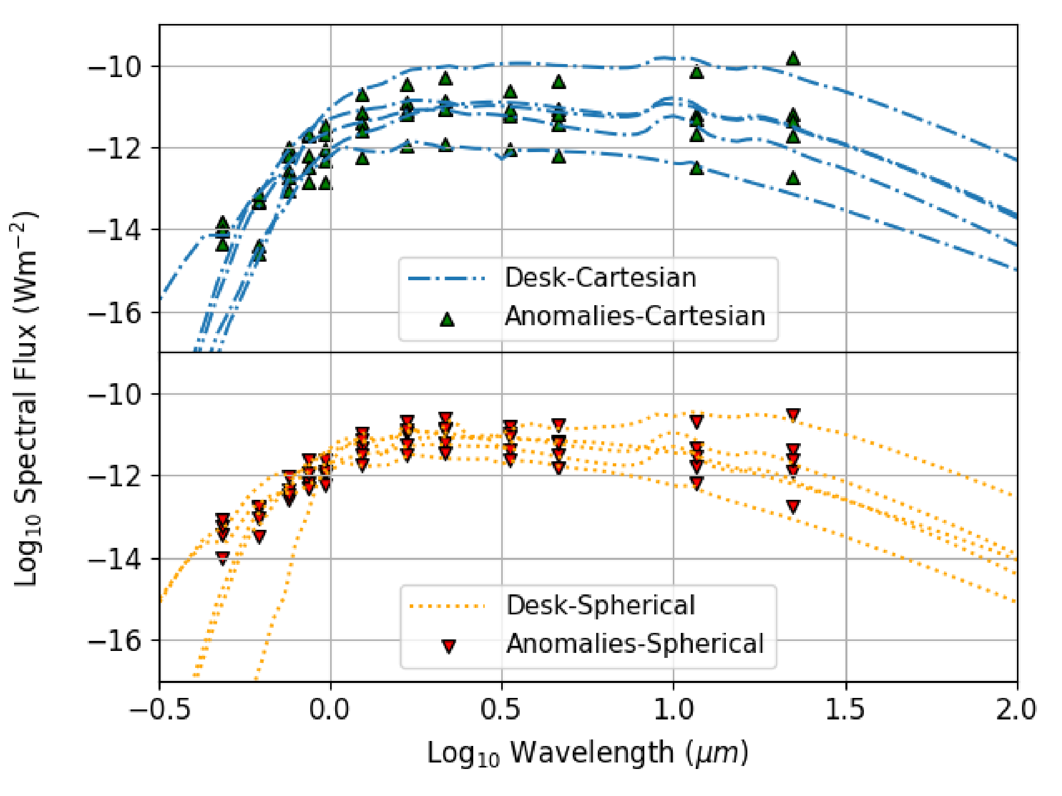
- 3D probabilistic dust map. Modify $m \rightarrow m - A(\lambda)$
- We take the mean extinction (computed overall direction I guess?)

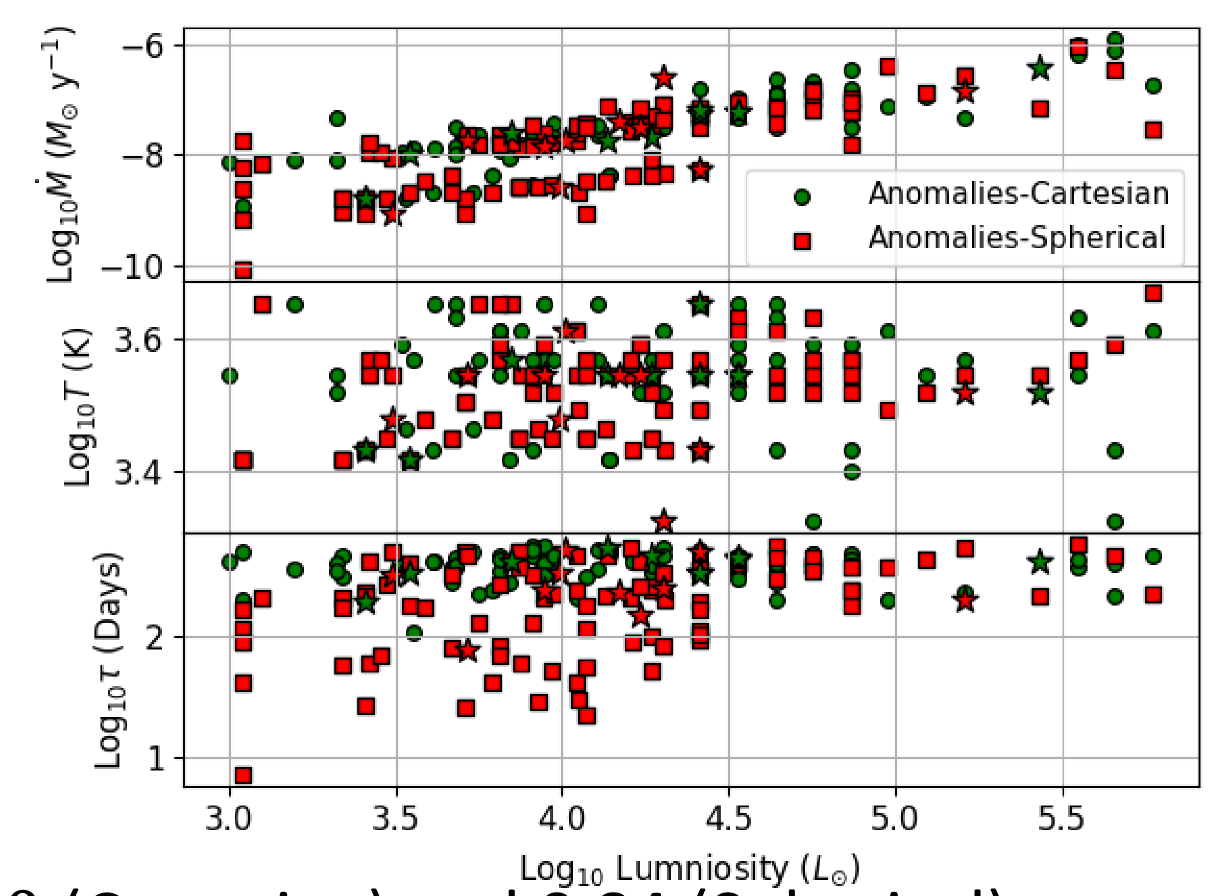
Spectral Energy Distribution



- Still very cool (Reddish)

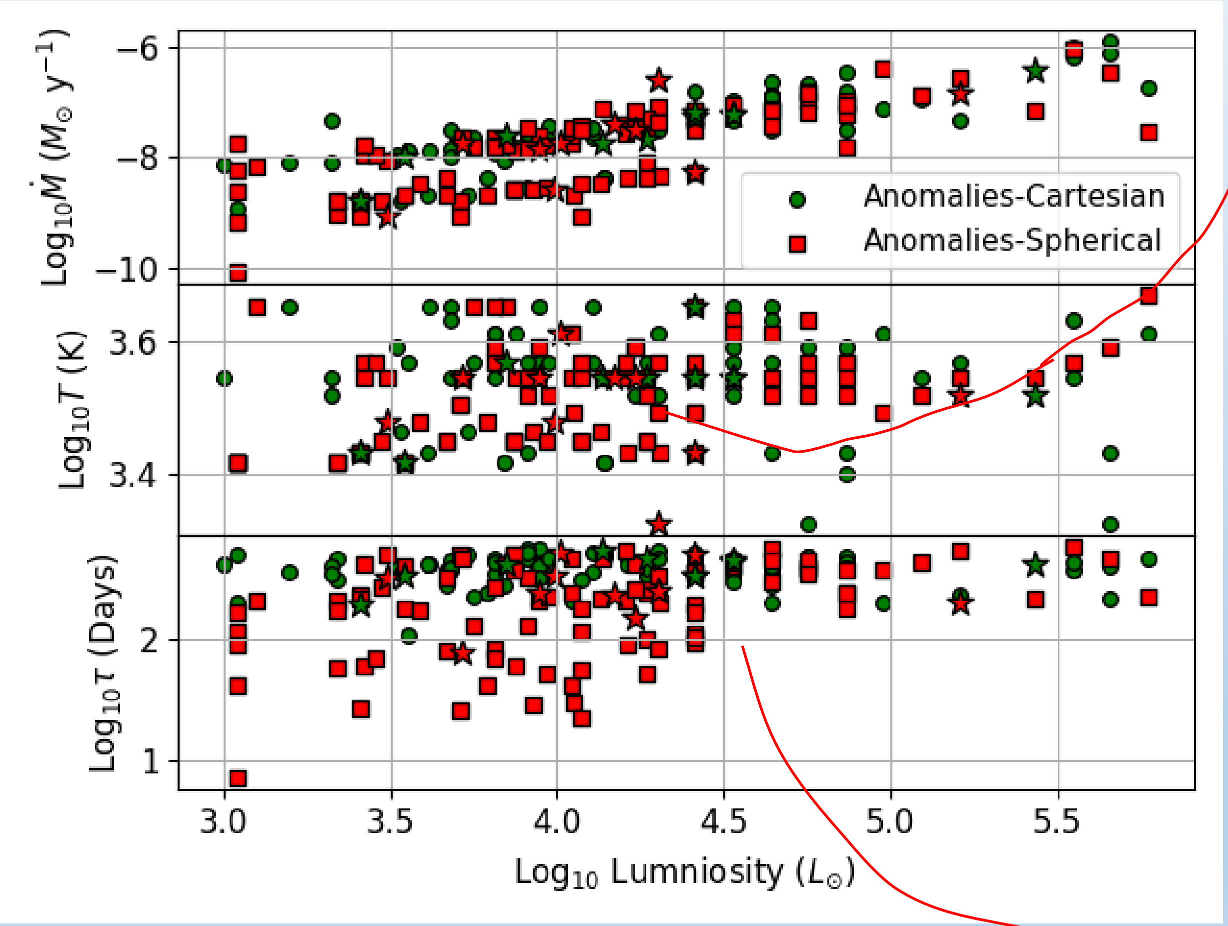
Fitted Spectra And LMT Relations

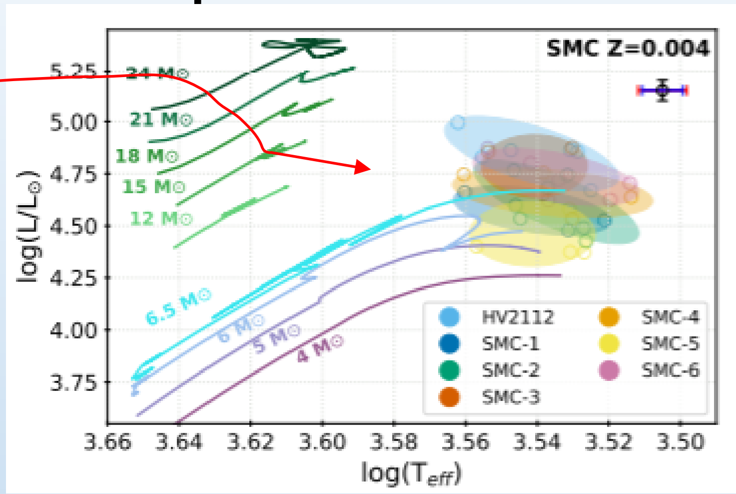


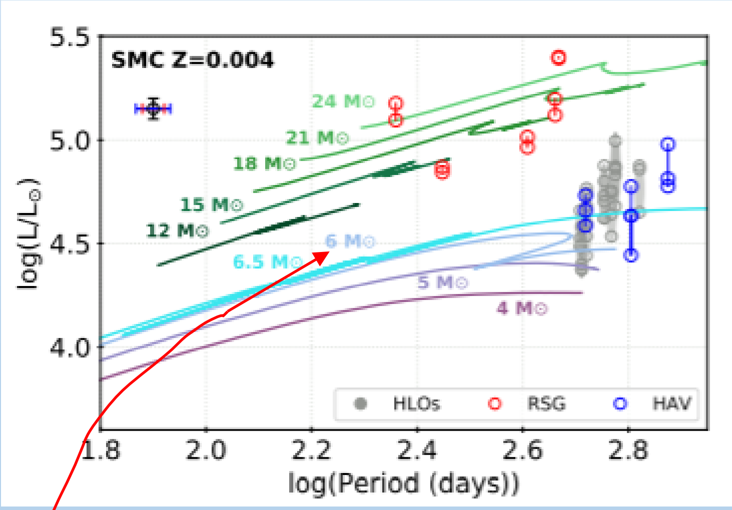


- Mean optical depth is now $\tau_{\nu} = 1.20$ (Cartesian) and 0.84 (Spherical)

HR-Diagram From The Canada Group



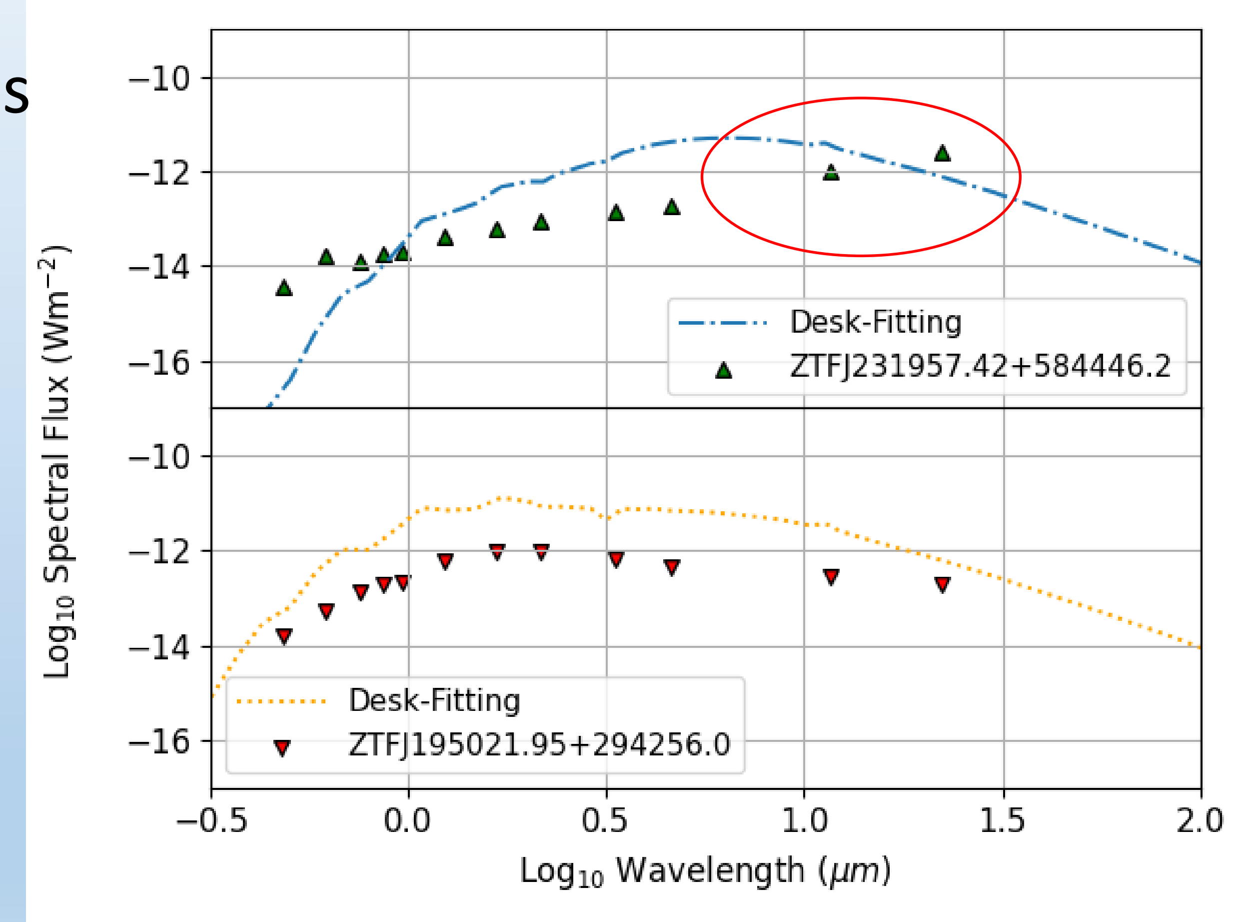




- A better presentation (?)

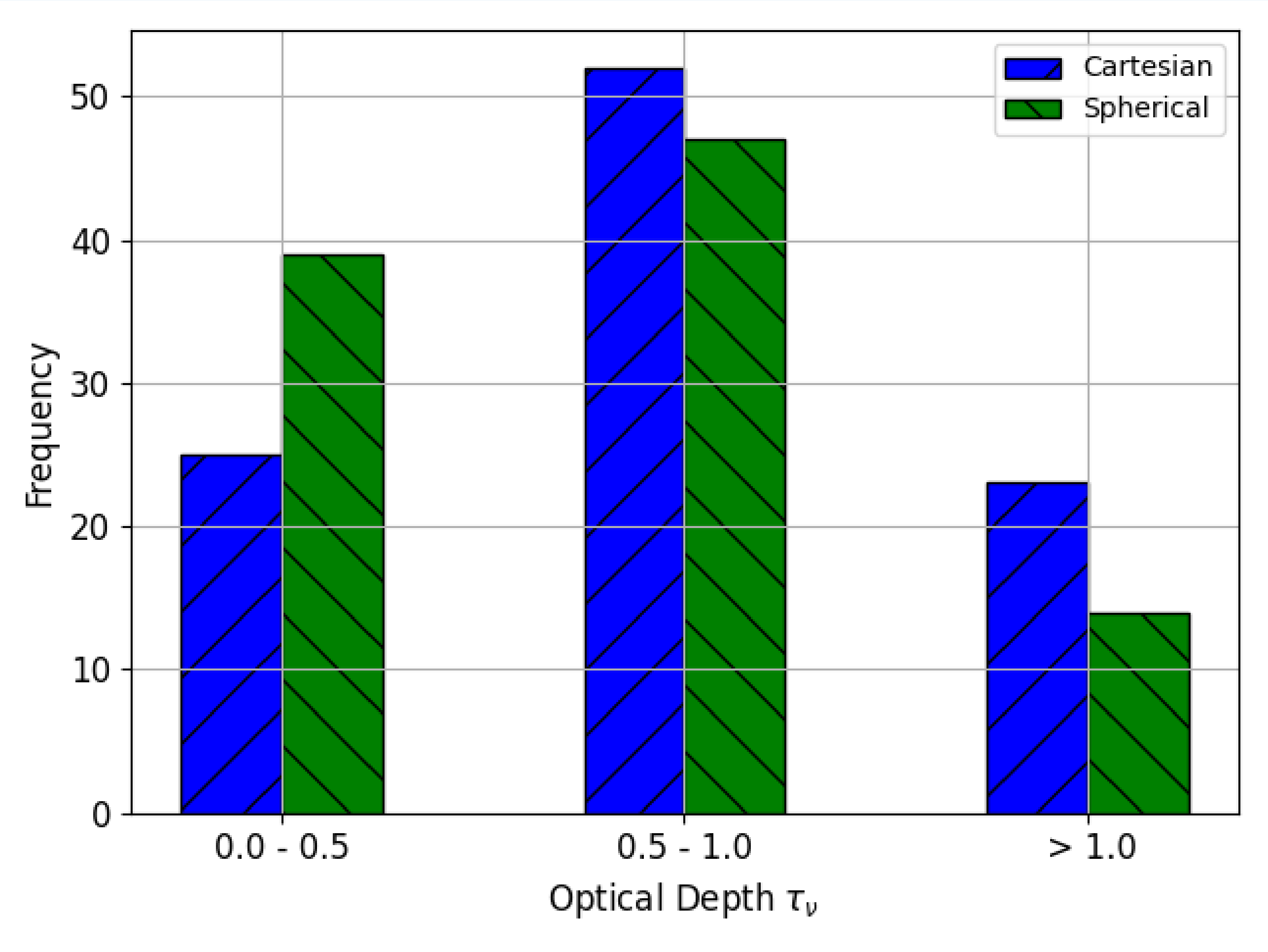
Non-Evolved Stars

- Cannot be well-fitted by the DESK
- Infrared Excess



- Possibly are YSO (?)

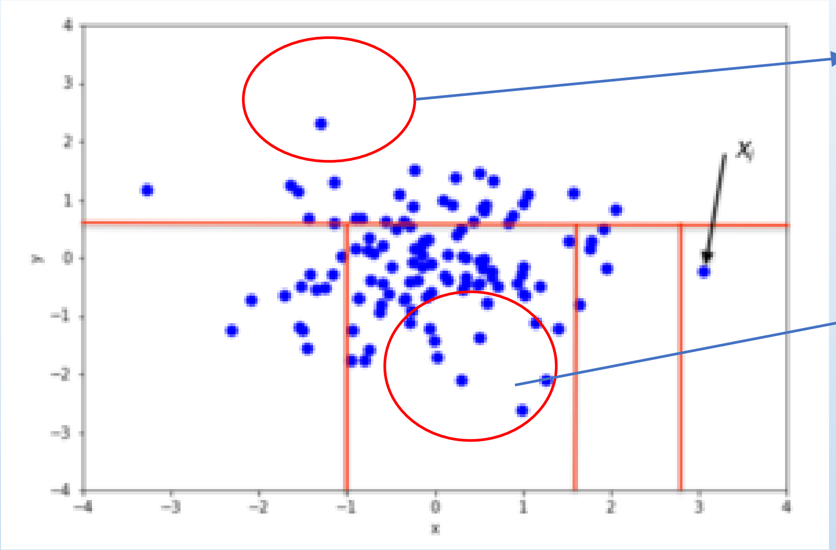
Optical Depth



- Mean optical depth is now $\tau_\nu = 1.20$ (Cartesian) and 0.84 (Spherical)

Data Pre-processing

Isolation Forest

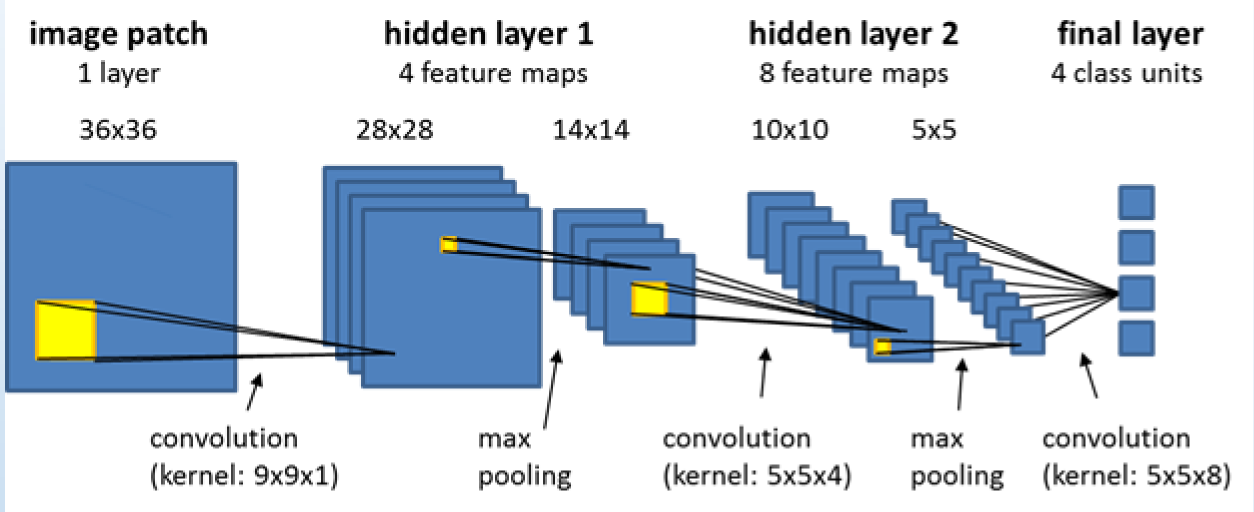


- One cut only. Should be a strong outlier
- Data that clustered with each other
- Need many more cuts

- Draw a lots of decision trees (cuts) in the feature spaces
- Cuts to separate data from another

Isolation Forest Would Be Applied To The Latent Space

Convolutional Neural Network

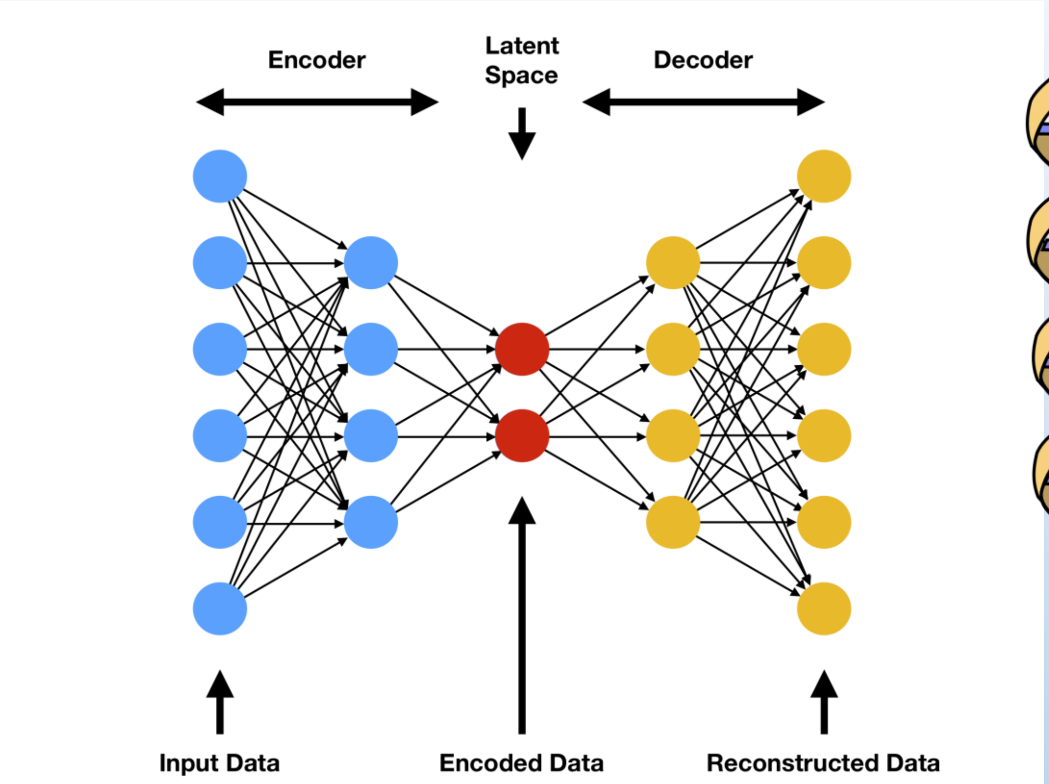


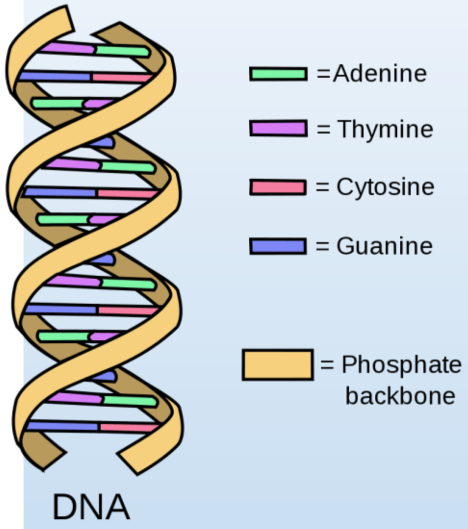
- Majorly used for image recognition
- Assume correlation between neighbourhood grid

- A blue grid (sea) probably adjacent to a blue grid (sea)

Could Be Useful For PHASE-FOLDED Periodic Light Curves

Auto-Encoder And Latent Variables





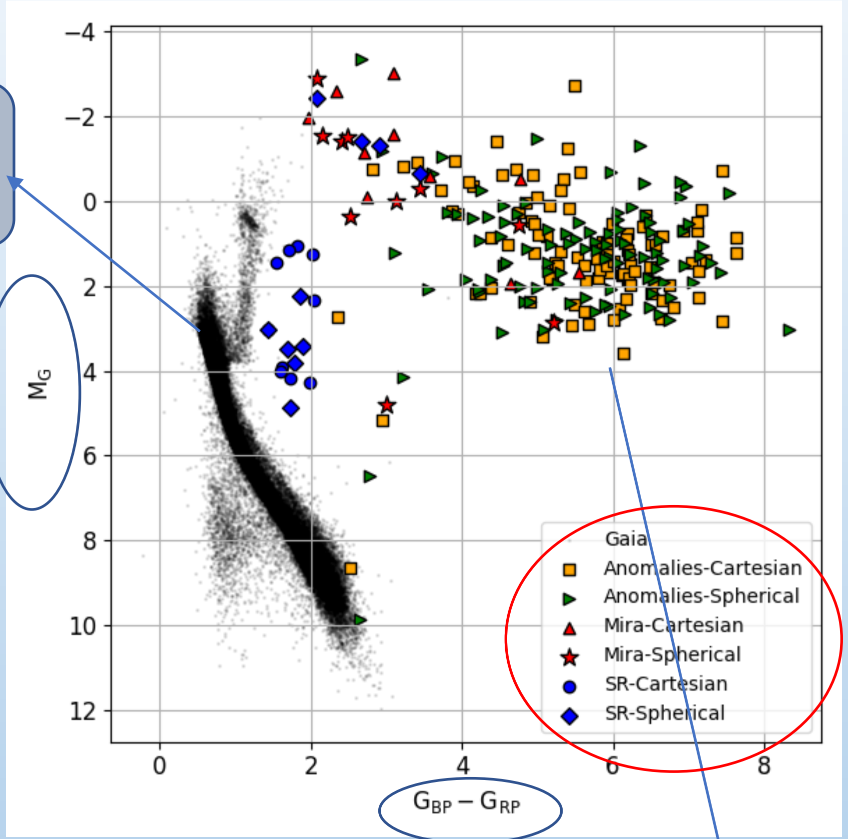
Governs

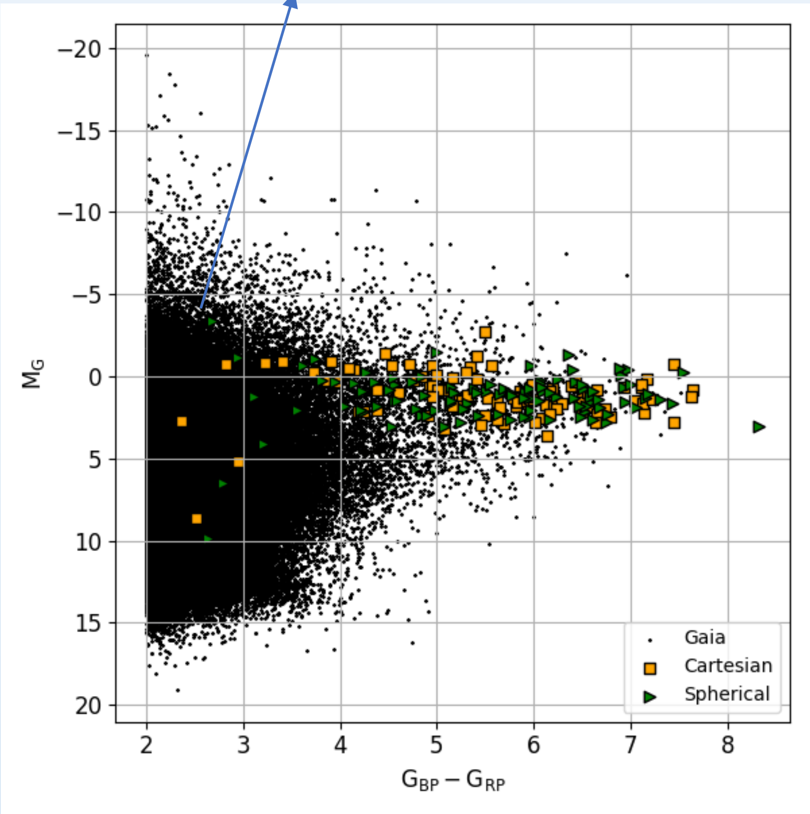
- A Convolutional Variational Auto-Encoder
- Input – Light curves in g- and r-band (appearances of human being)
- Latent variables – Resemble DNA for biological species

HR-Diagram

100000 Samples

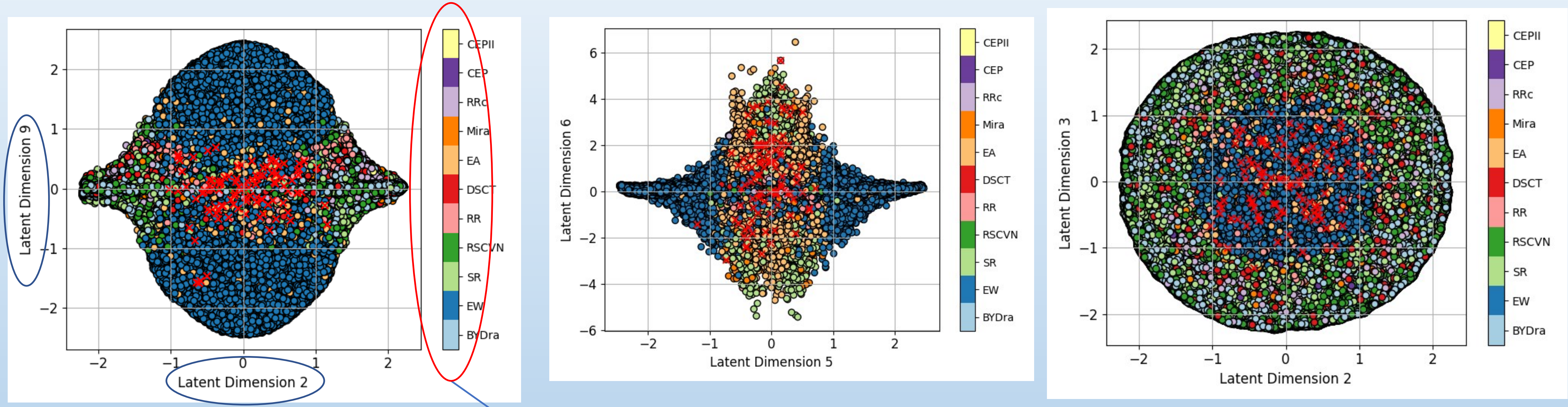
100000 Samples





- MG (Absolute Brightness) vs GBP – GRP (Hotness)
- Show that they are cool and bright (Also relative to normal ones)
- Why are they so “cold” … Dust?

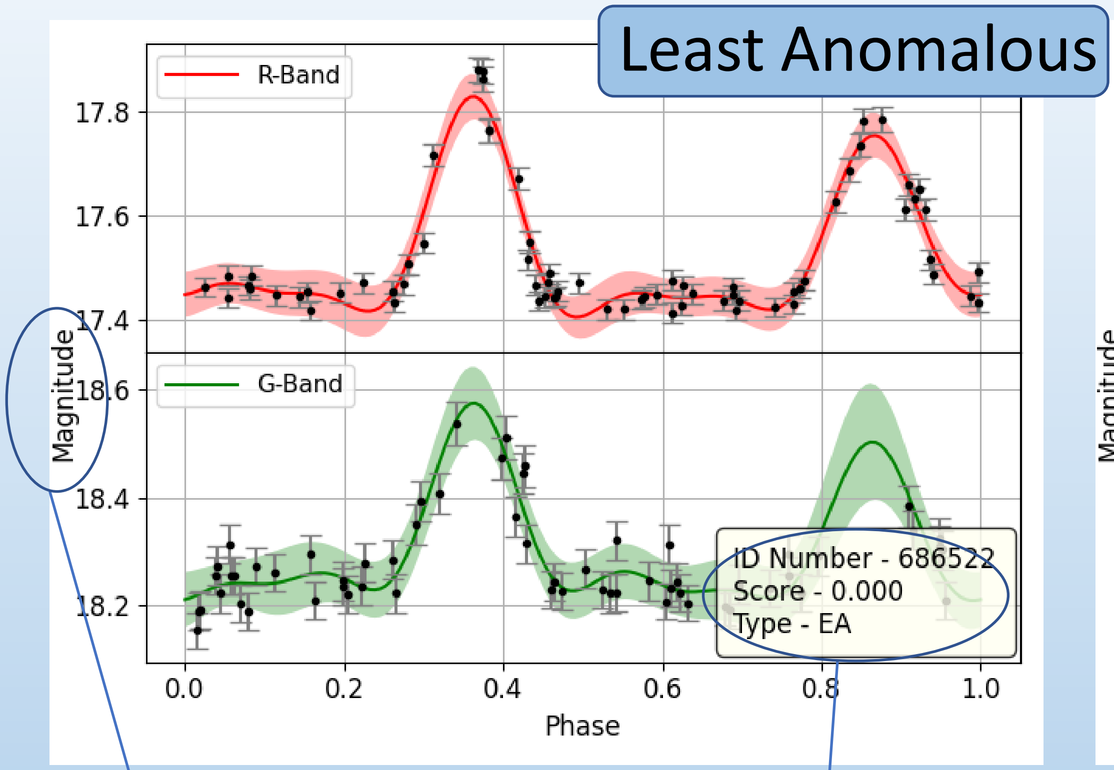
Latent Feature Space

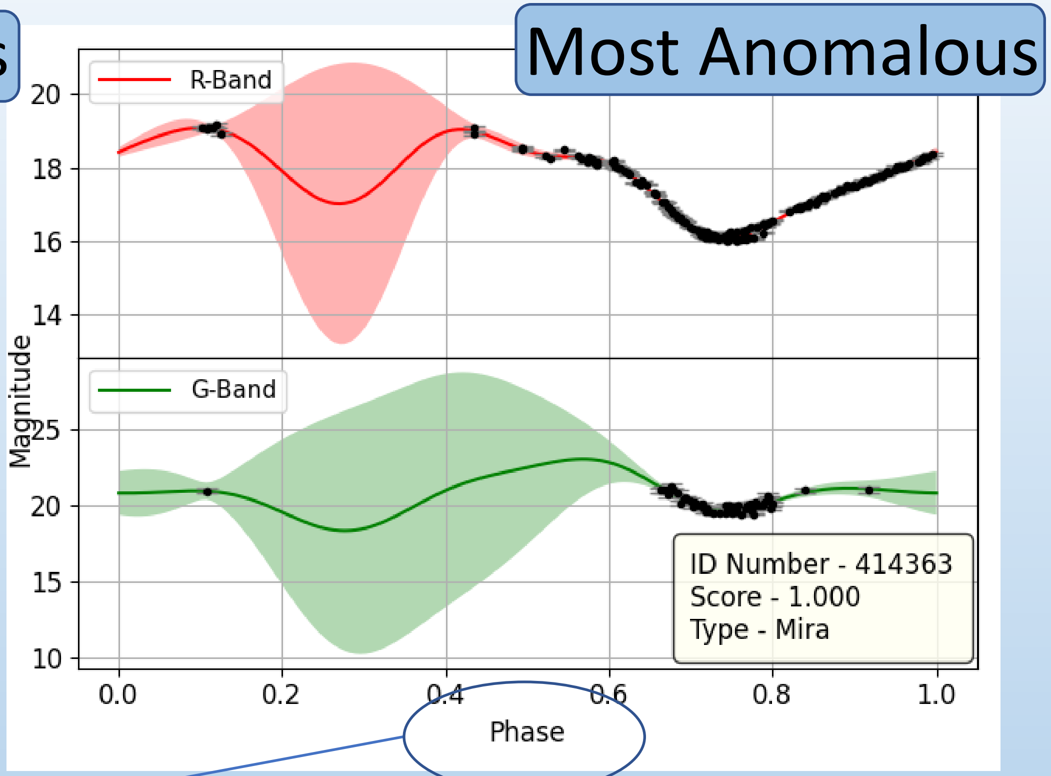


- 2D-Latent Plots
- Robust in separating different class of variable stars
- Exhibit Spherical Structure
- Maybe worth to transform into “Spherical Coordinates”

Random Forest

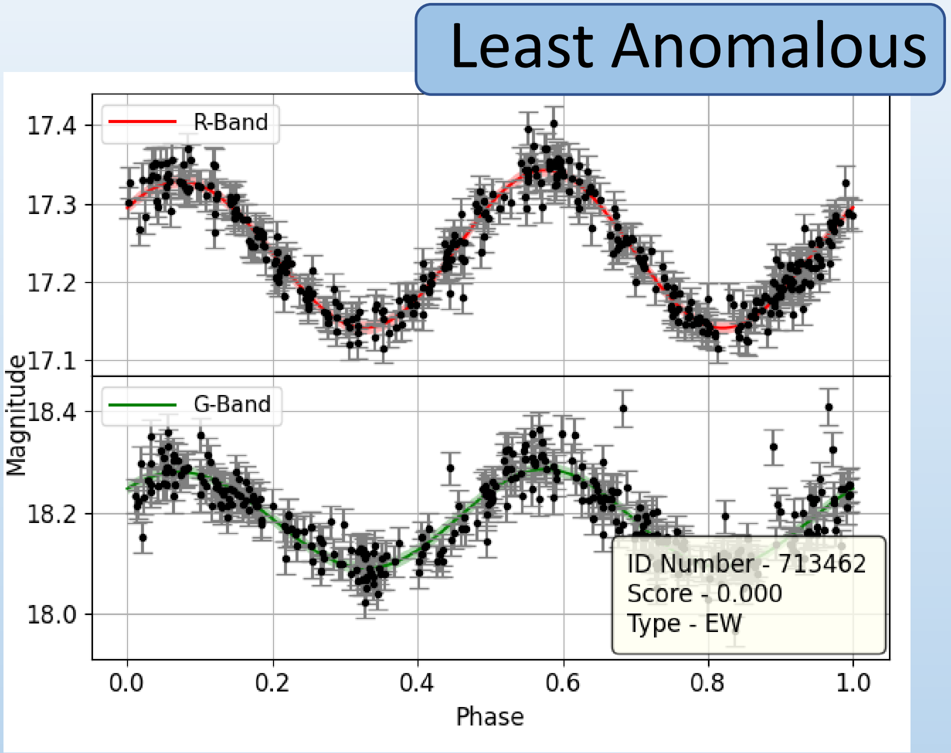
Anomalies In Cartesian Space

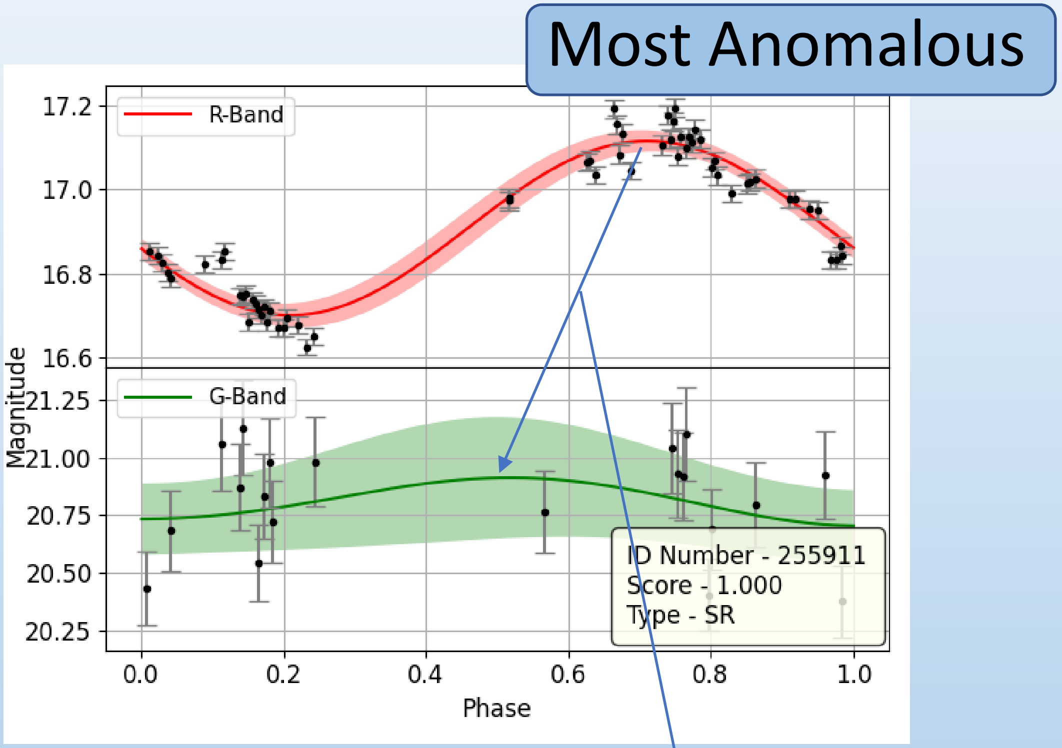




- Apparent Magnitude VS Phase [0, 1]
- Appended normalised scores and star type
- Characteristic – Irregular oscillating, multiple Fourier modes
- Mostly are SR and Mira Variables

Anomalies In Spherical Space

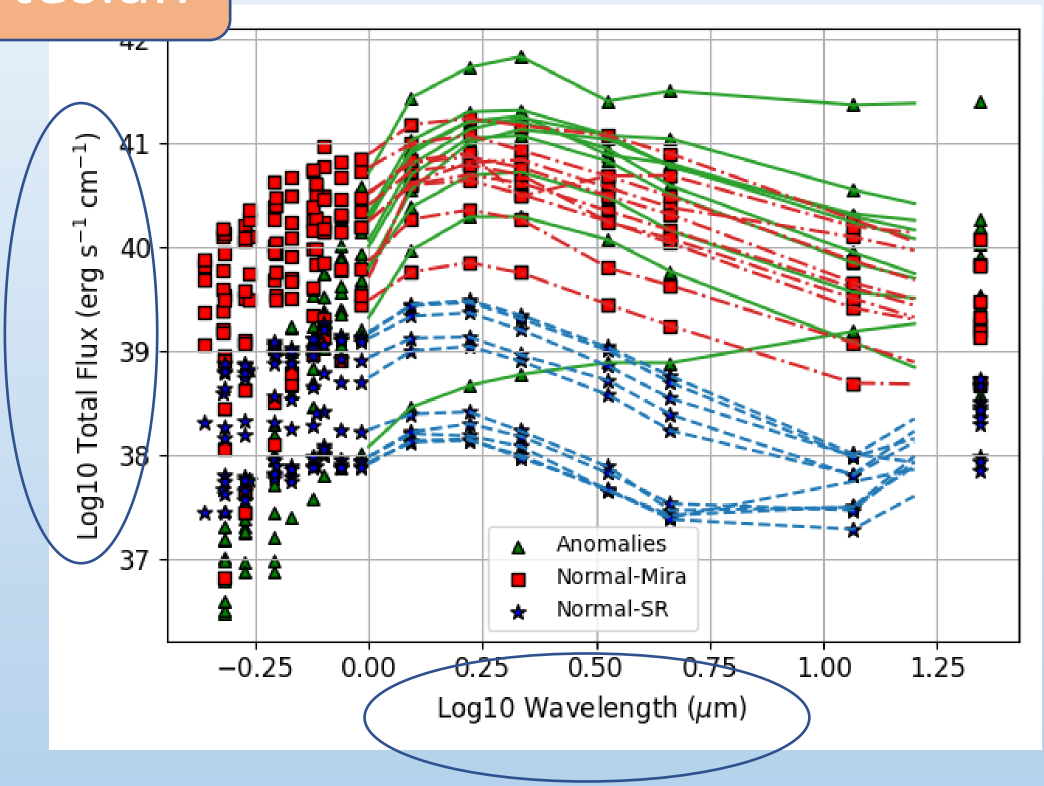




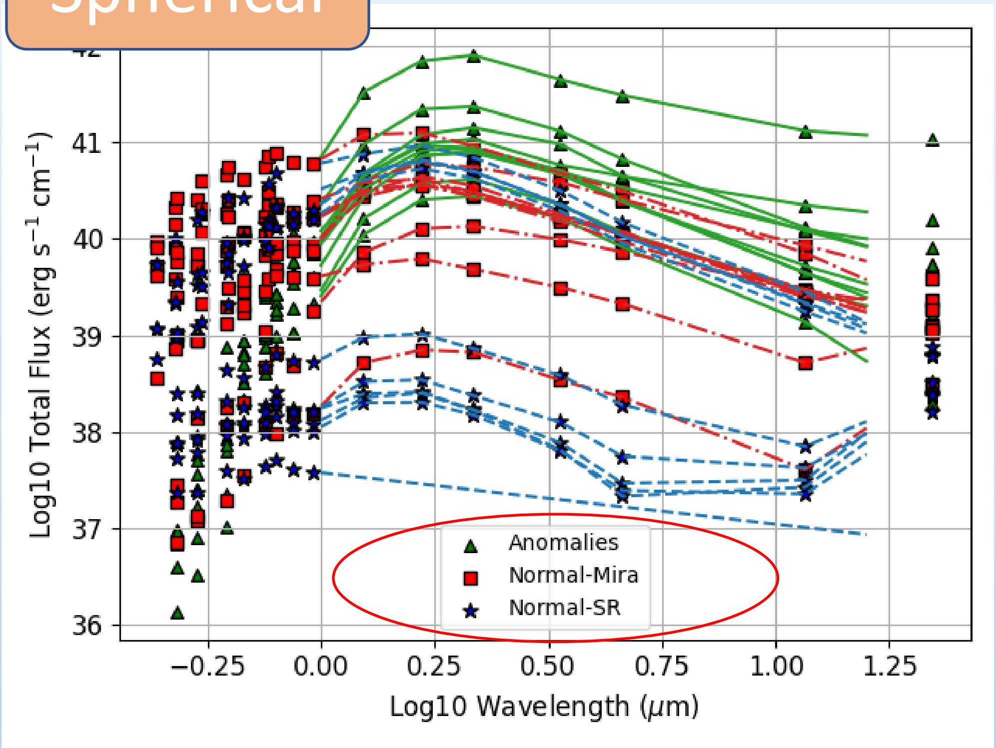
- Some anomalies are also irregular and largely variate
- Some of them even show relative phase-lag between g- and r- band

Spectral Energy Distribution

Cartesian



Spherical



- Plotted spectral energy distribution in wavelength space
- Peak around microwave region (black body temp < 2000 K)
- The normal ones are relatively hotter. Must be so many dust

Fitting The Spectral Energy Distribution

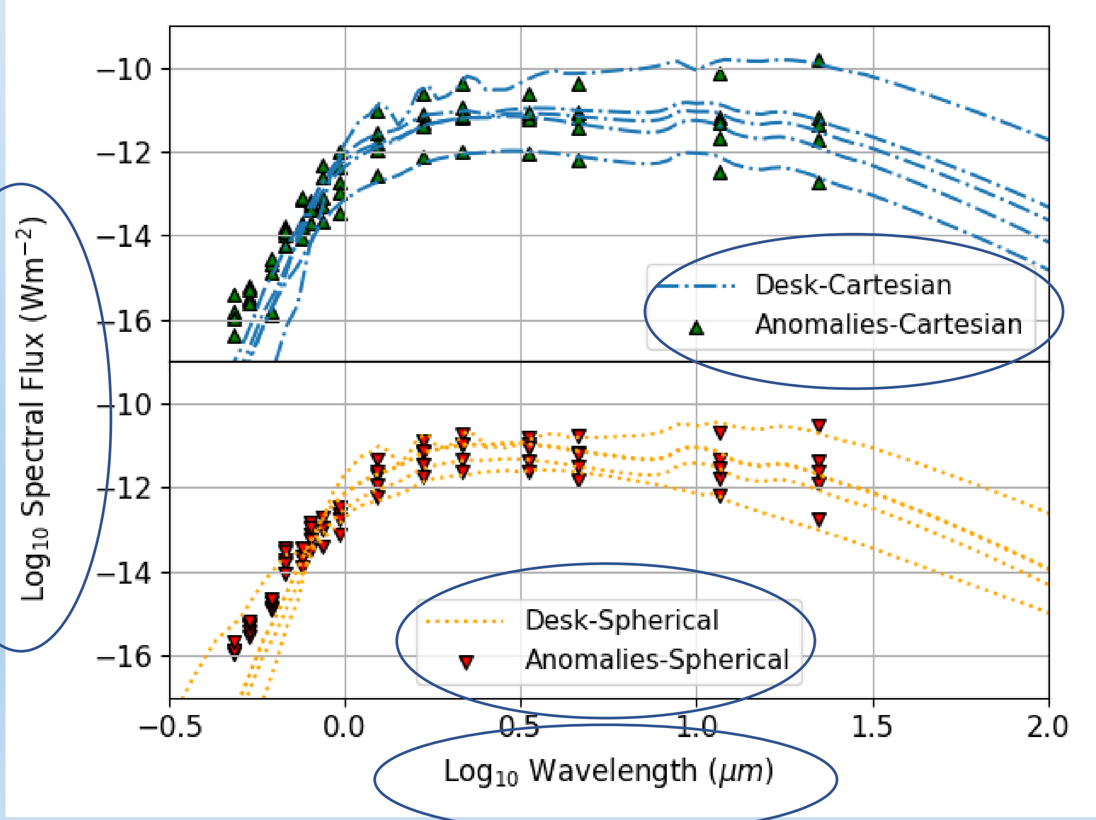
The Dusty Evolved Star Kit (DESK): A Python package for fitting the Spectral Energy Distribution of Evolved Stars

Steven R. Goldman[1]

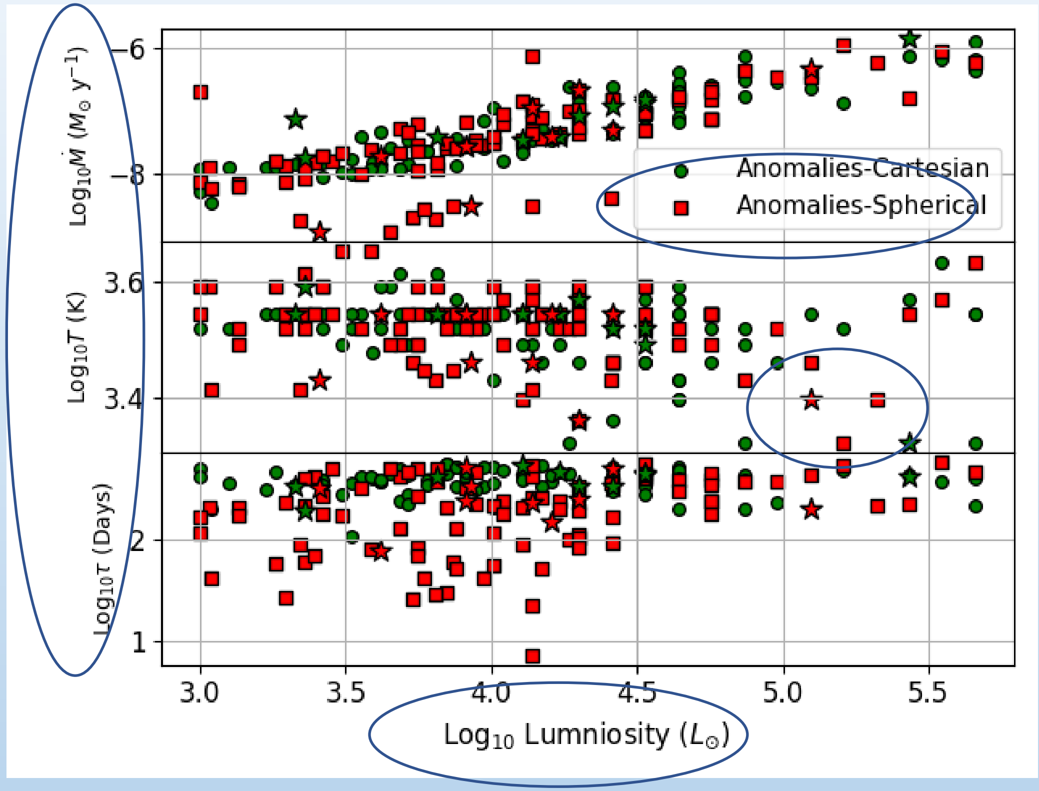
1 Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

- Open-source code – DESK (python)

- Chosen five top anomalies as example
- DUST composition: C + O
- Important parameter: τ_{ν} (at 10 μm)
- We find $\tau_{\nu} \approx 2.0$ (Wow!)
- Only a trial

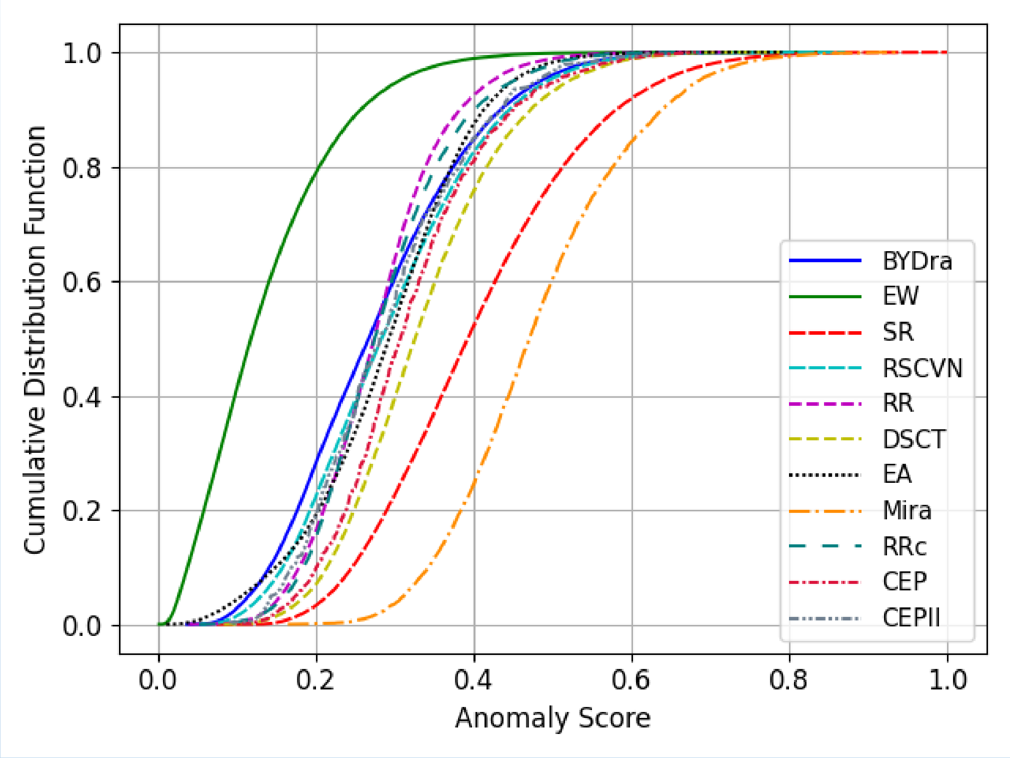


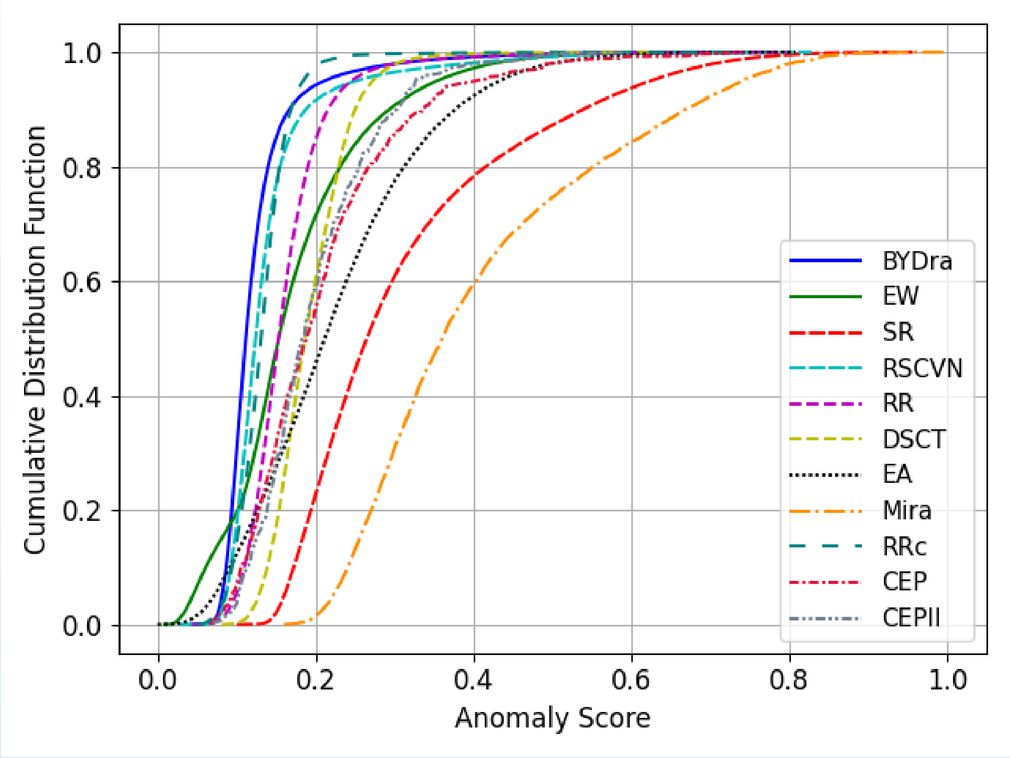
Period VS Mass Loss VS Luminosity



- Plotted: Mass loss rate/Intrinsic Temp/Period vs Intrinsic luminosity
- Stars are marked as top 10 anomalies
- Looks like AGB or RG stars?

Cumulative Distribution Function





- Normalised and ranked anomalous star according to the score