Searching for Anomalies in the ZTF Catalog of Periodic Variable Stars

Introduction and Motivation

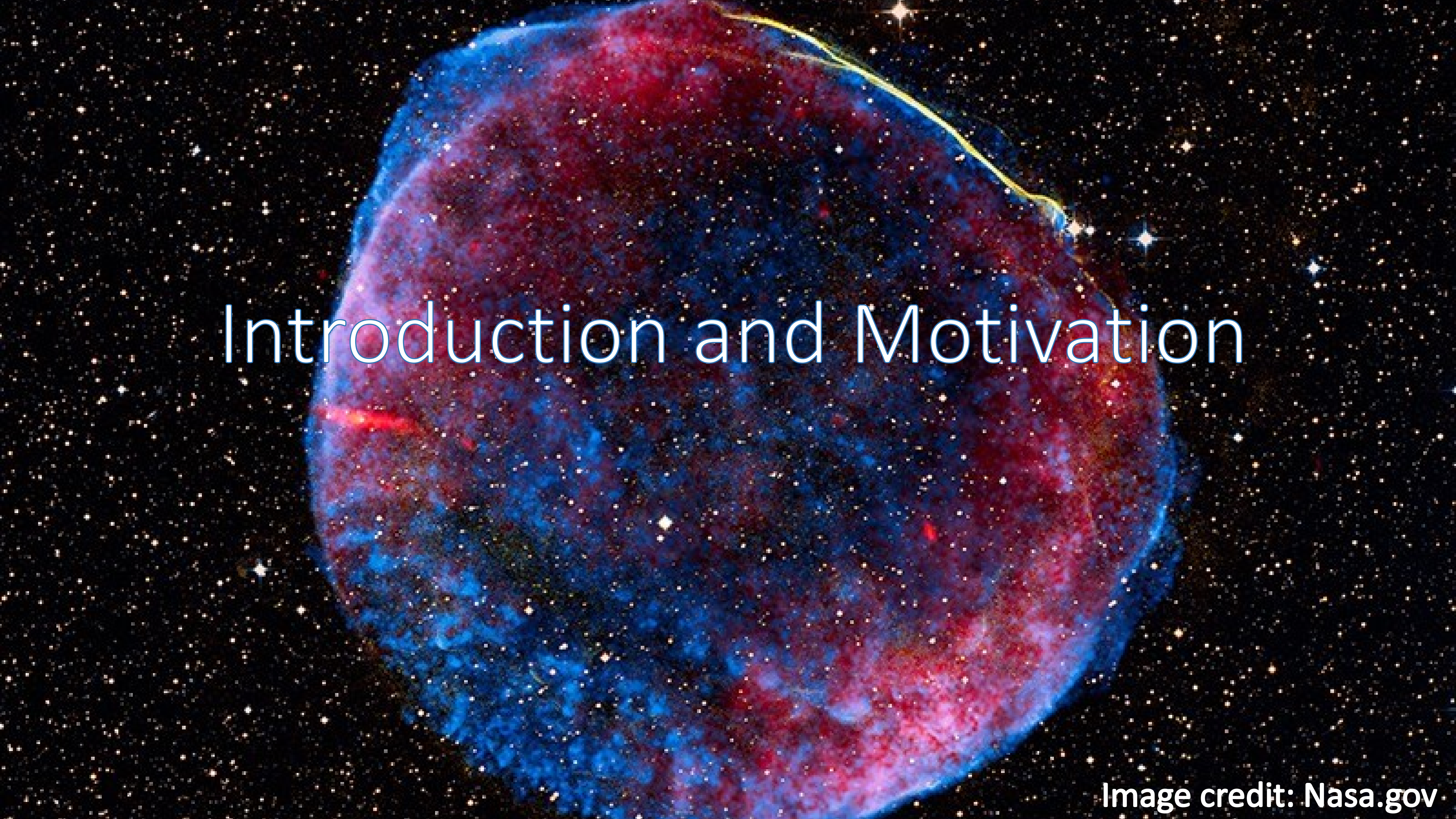


Image credit: Nasa.gov

Why care about anomalies?

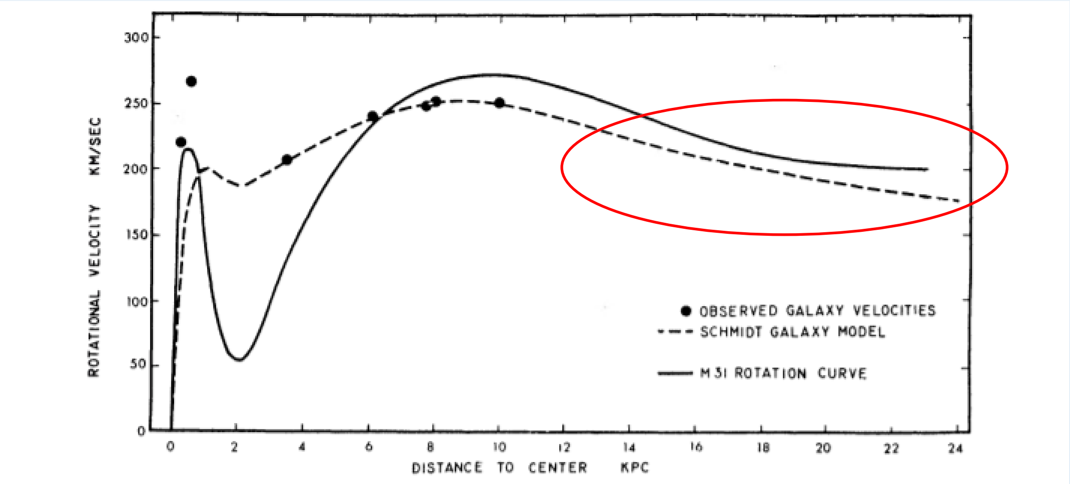
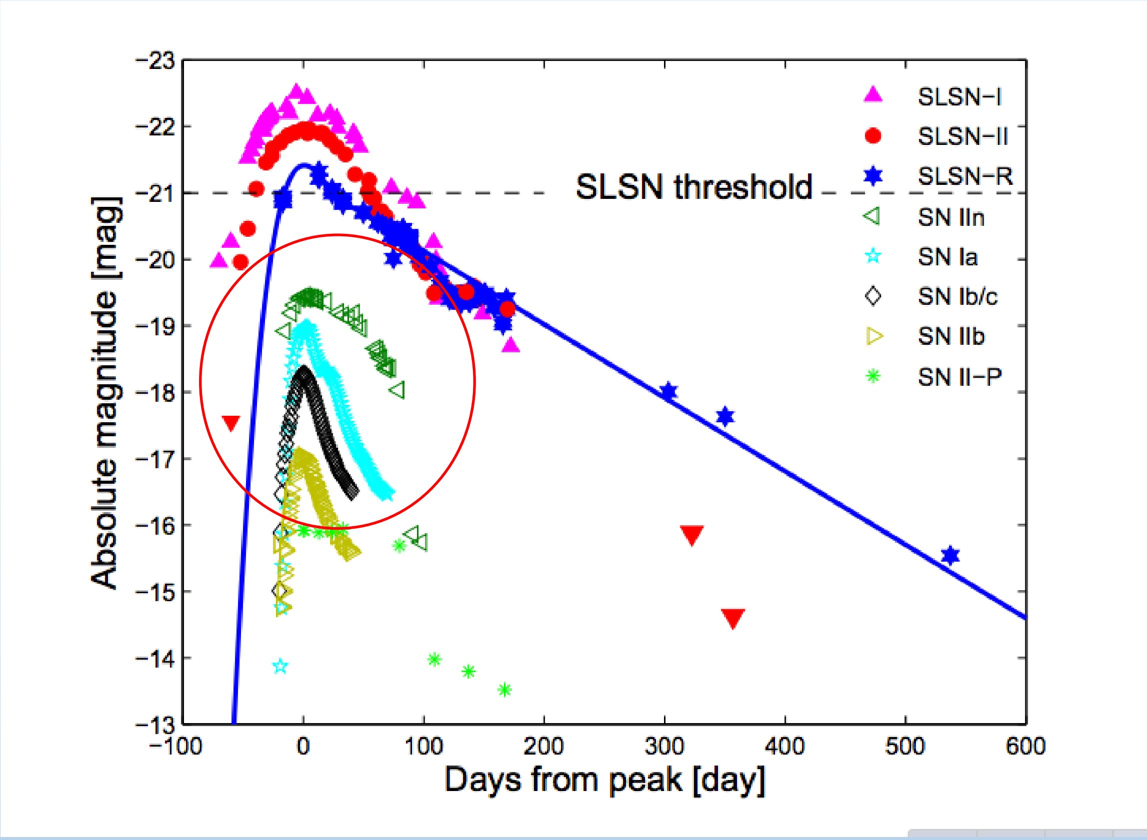


FIG. 15.—Comparison of rotation curves for M31 and the Galaxy, as a function of distance from the center. *Solid line*, rotation curve for M31 (Fig. 9); *dashed line*, rotation curve from Schmidt model of the Galaxy. *Filled circles*, observed rotational velocities for the Galaxy (Rougoor and Oort 1960; Schmidt 1965).

- Rubin et al. in 1970 discovered anomalous (flat) Galactic rotational velocity



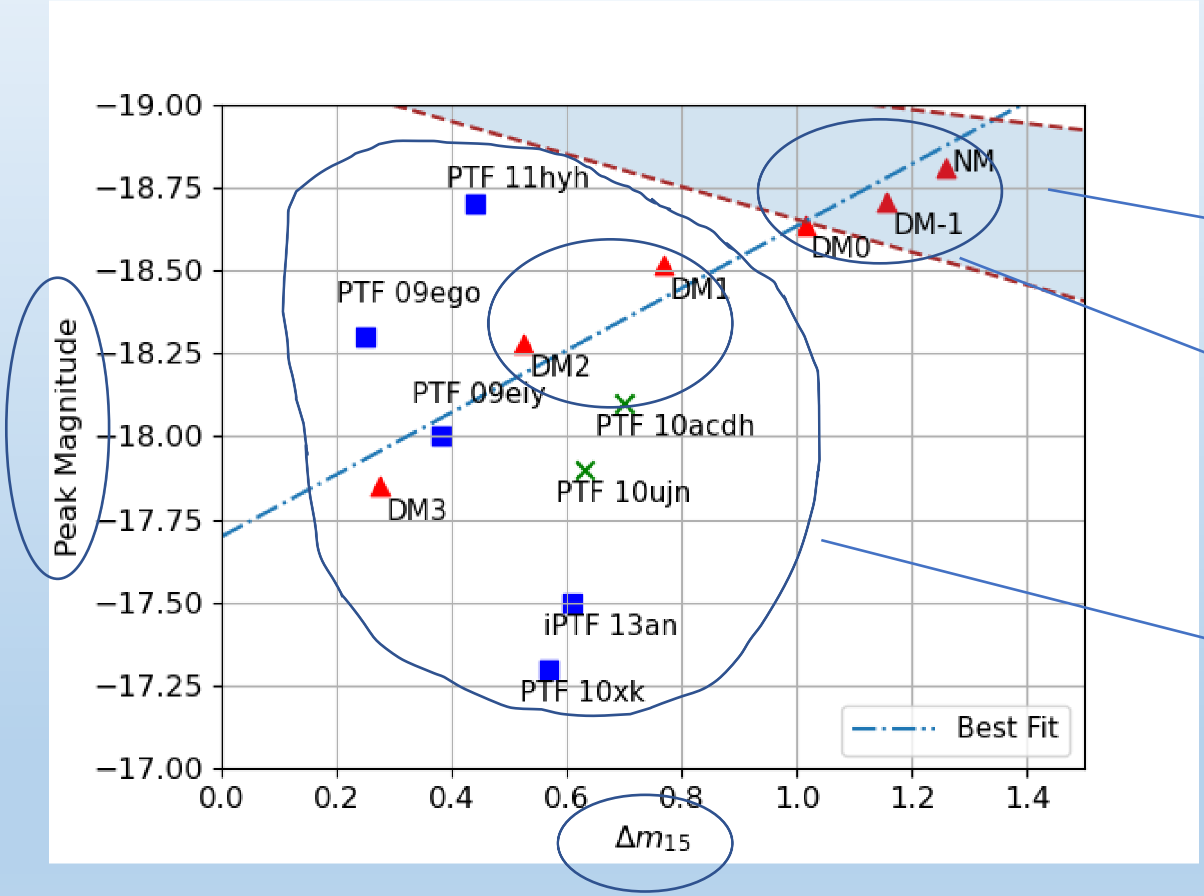
- Super-luminous supernova

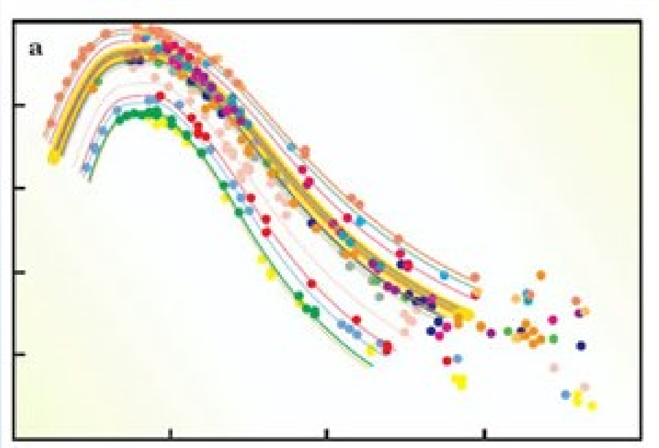
- Anomalies – unusual, out of expectation stuffs
- Have been part of the development history of Astronomy

What can anomalies tell us?

- New physics!

Chan, Ho-Sang, et al. "Delayed detonation thermonuclear supernovae with an extended dark matter component." The Astrophysical Journal 914.2 (2021): 138.





Observed R-Band Phillips Relation

DM-admixed Supernova

Low-luminosity

Slowly-declining

- Asking why and how do anomalies doesn’t behaves as expected
- Improved theories – better understanding to nature

Previous work and research gap

Anomaly detection in the Open Supernova Catalog

M. V. Pruzhinskaya,[1★] K. L. Malanchev,[1,2★] M. V. Kornilov,[1,2] E. E. O. Ishida,[3] F. Mondon,[3] A. A. Volnova[4] and V. S. Korolev[5,6]

Supernova

Extragalactic Astrophysics

A Deep-learning Approach for Live Anomaly Detection of Extragalactic Transients

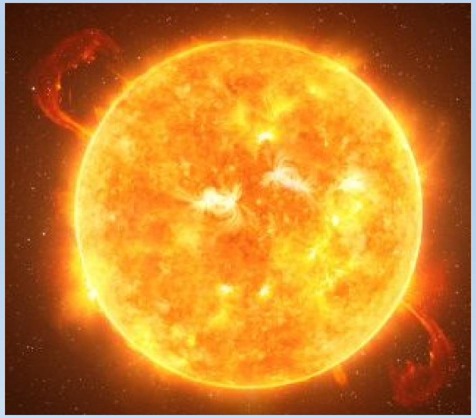
V. Ashley Villar[1,2,3,4,10], Miles Cranmer[5], Edo Berger[6,7], Gabriella Contardo[8], Shirley Ho[8], Griffin Hosseinzadeh[6,7], and Joshua Yao-Yu Lin[9]

Searching for Changing-state AGNs in Massive Data Sets. I. Applying Deep Learning and Anomaly-detection Techniques to Find AGNs with Anomalous Variability Behaviors

P. Sánchez-Sáez[1,2,3], H. Lira[1], L. Martí[1], N. Sánchez-Pi[1], J. Arredondo[2], F. E. Bauer[3,4,2,5], A. Bayo[6,7], G. Cabrera-Vives[8,2], C. Donoso-Oliva[8,2], P. A. Estévez[9,2], S. Eyheramendy[10,2], F. Förster[11,2,12,13], L. Hernández-García[2,6], A. M. Muñoz Arancibia[2,12], M. Pérez-Carrasco[8,2], M. Sepúlveda[13], and J. R. Vergara[14,2]

AGNs

- Previous work focusing on high-energy transients



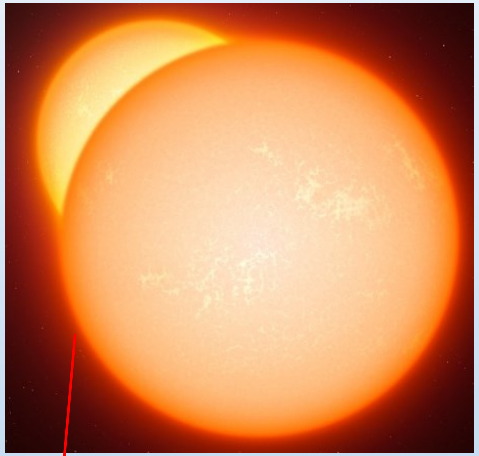
- But we don’t know too much about stars!
- Why not search for anomalous stars?

Periodic variable stars (PVSs)

Image credit: astro.wisc.edu

Or

Image credit: newatlas.com

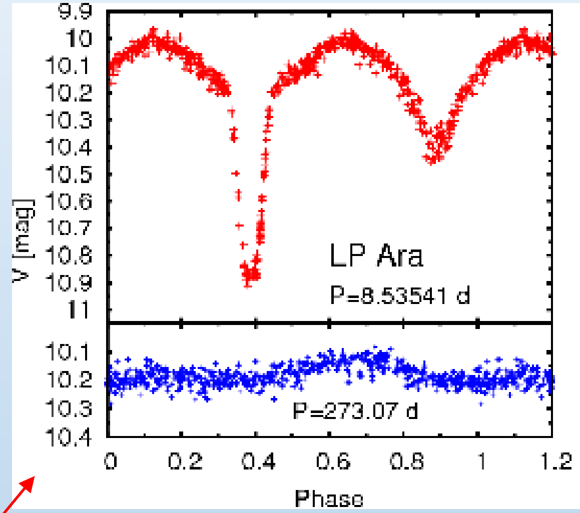


Intrinsic

Extrinsic

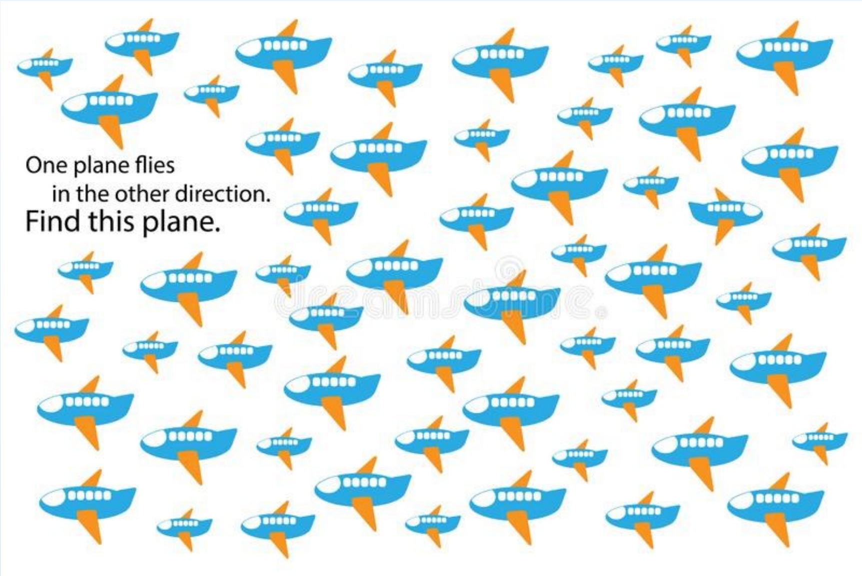
Produce

Image credit: Michalska et al. (2009)



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

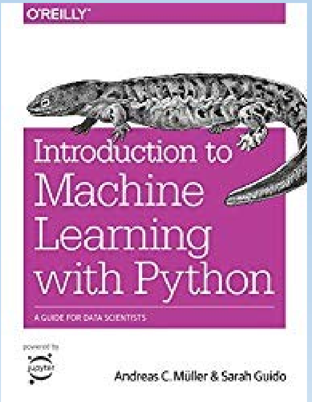
How to tackle the problem ? Machine Learning!



LSST – Commence in 2023

20 TB data/night.

- When the candidate number is small – scan one by one
- What if the number increase exponentially?



- Machine learning would be a reliable and automatic method!

Aim – Use ML to search for anomalous PVSs

Methodology

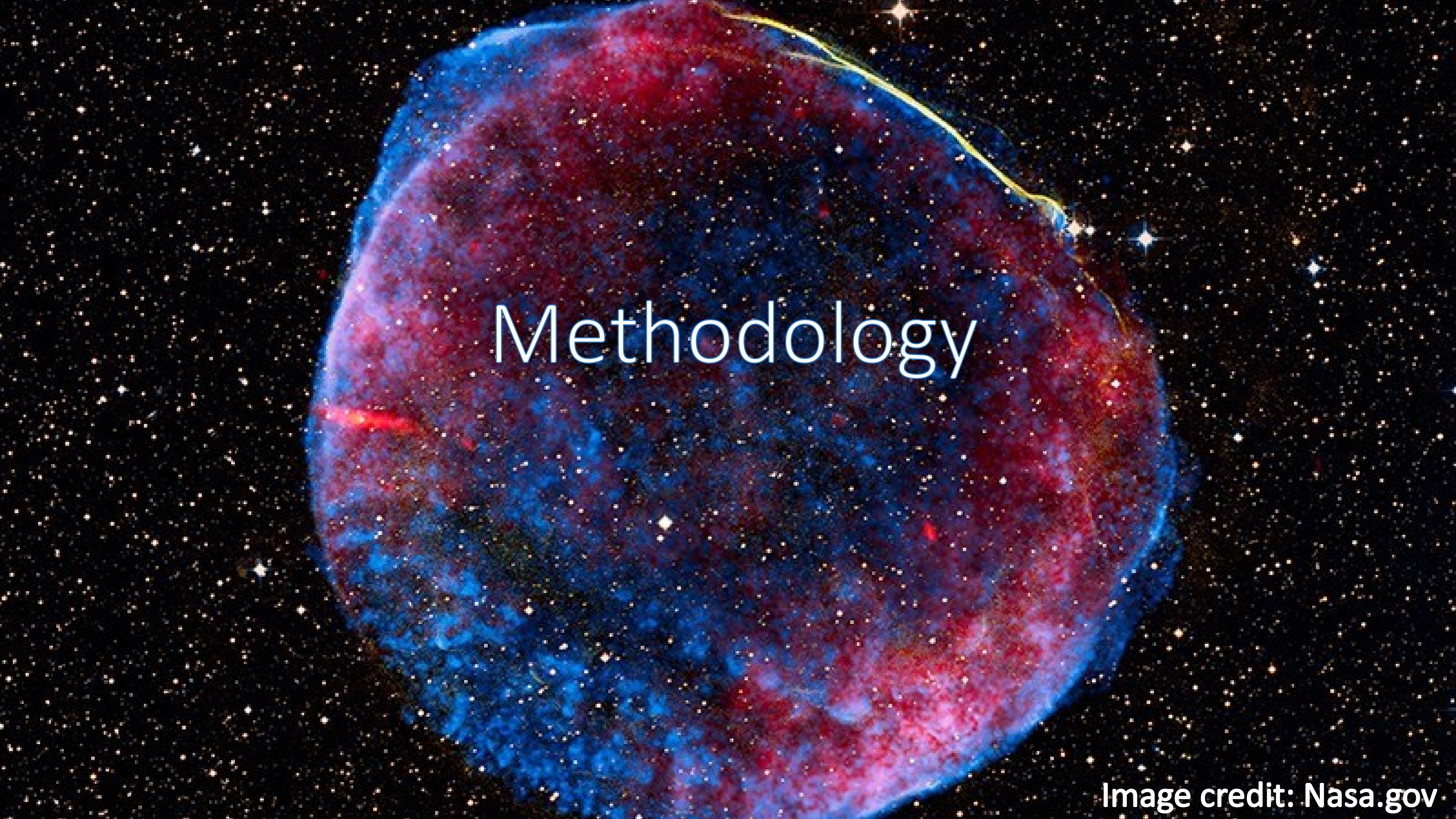


Image credit: Nasa.gov

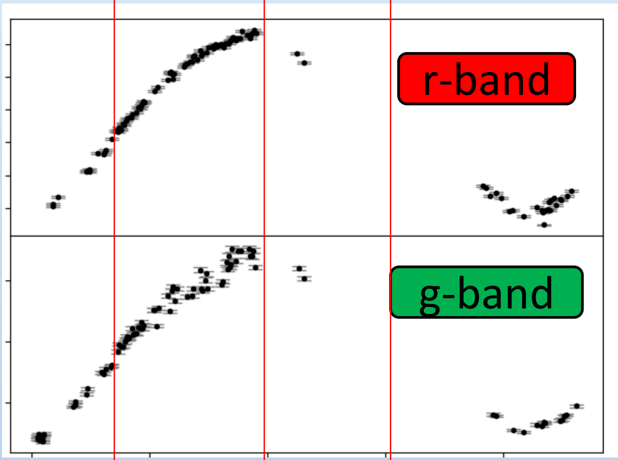
Data pre-processing – Feed meaningful info to machine

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]

- Data are given in g- and r- band filters

Raw Detections

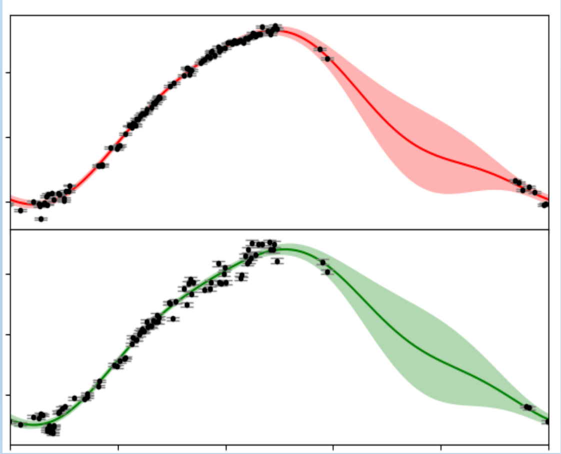


Phase-Folding

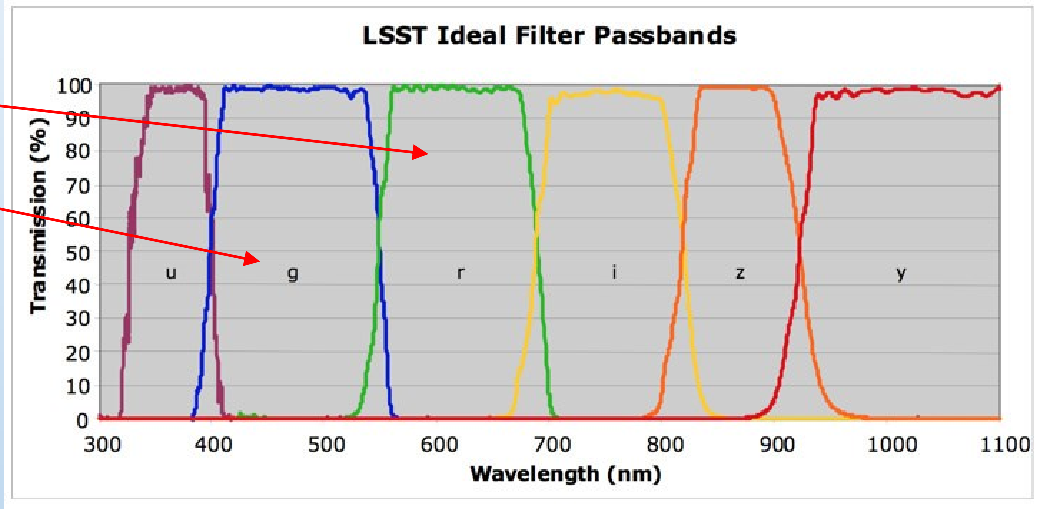
$$\phi = \frac{t}{\tau} - \left[\frac{t}{\tau}\right]$$

Period

Multivariate-GPR



Normalised



Stacked “Images”

Image credit: wordpress.com

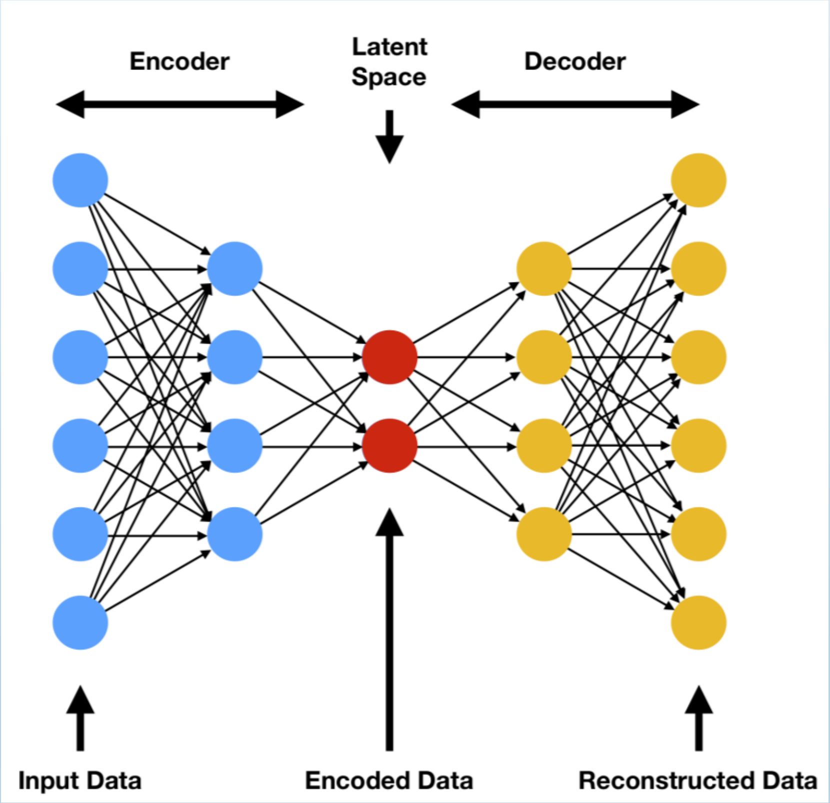
Wavelength

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

Phase

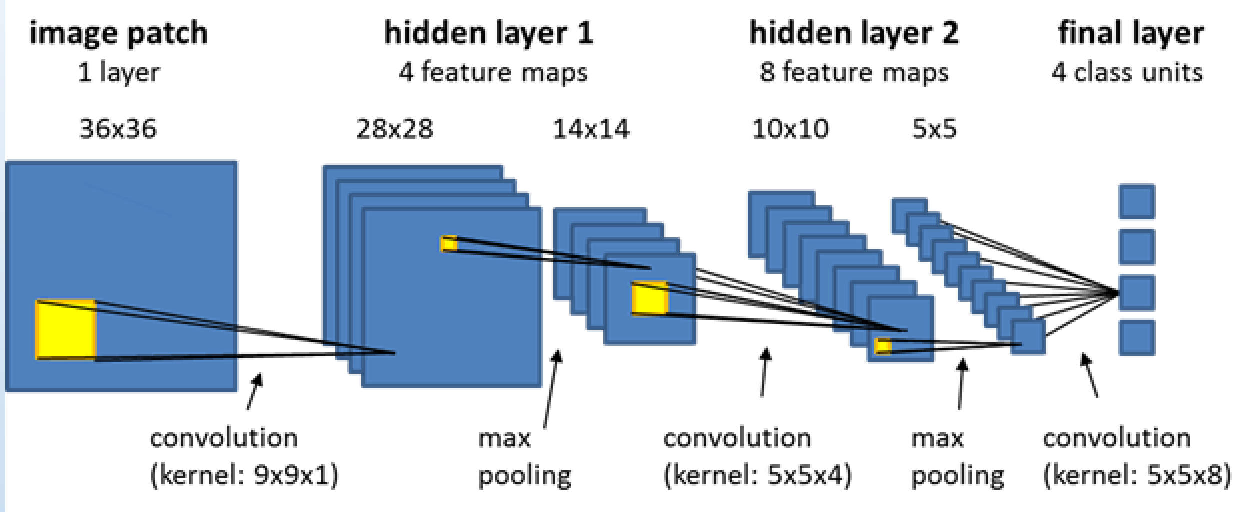
Auto-Encoder and Latent Variables

- Data volume: 2x160 per light curves
- Reduce the dimension of each LCs
- Keeping meaningful information?



- A Deep Convolutional Auto-Encoder
- Input – LCs in g- and r-band (appearances of PVSs)
- Latent variables – Resemble DNA for biological species

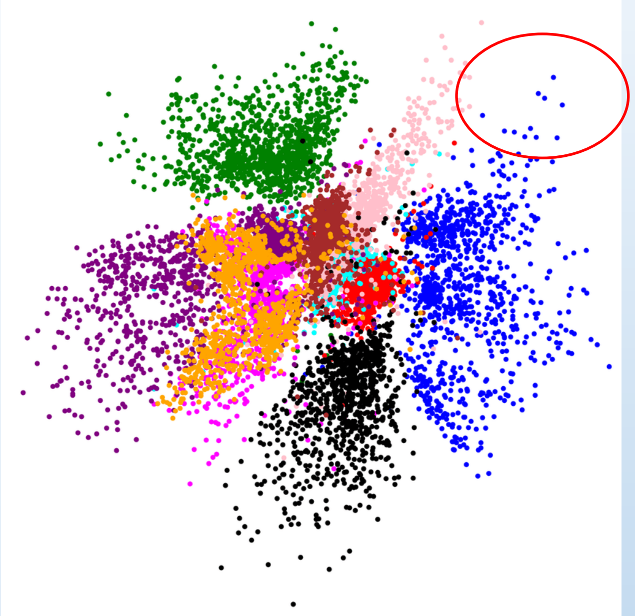
Convolutional Neural Network (CNN)



- Majorly used for image recognition
- Measure correlation between neighbourhood grid
- A blue grid (sea) probably adjacent to a blue grid (sea)
- Phase-folded light curves – no time dependence - image

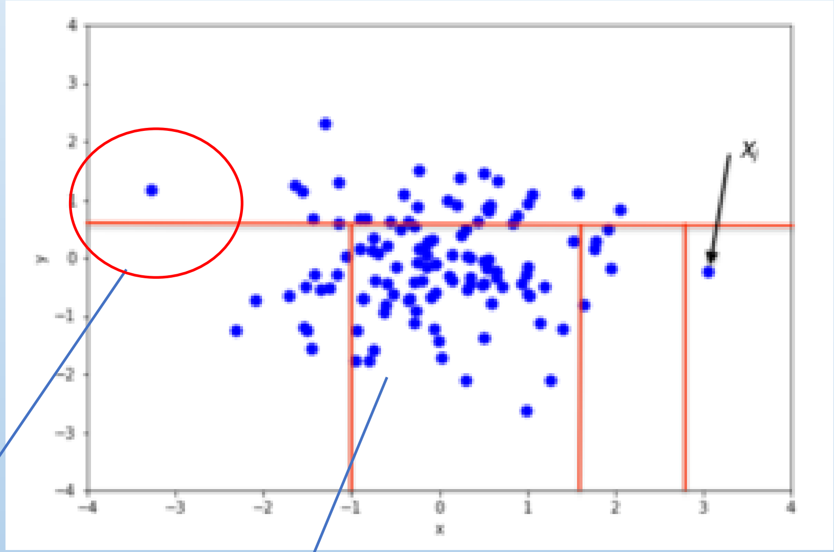
Seems to be good to use CNN for our problem

Using Isolation Forest to look for anomaies



- Assumed outliers in latent space are anomalies
- How to measure “outliers” quantitatively?

Isolation Forest – trees of cut

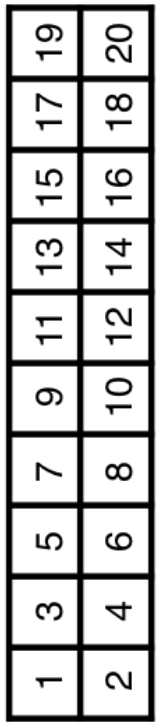


$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}},$$

Anomaly Score

- One cut only - should be a strong outlier
- Data that clustered with each other need many more cuts

Latent Features Extractions Pipeline



Stacked Images

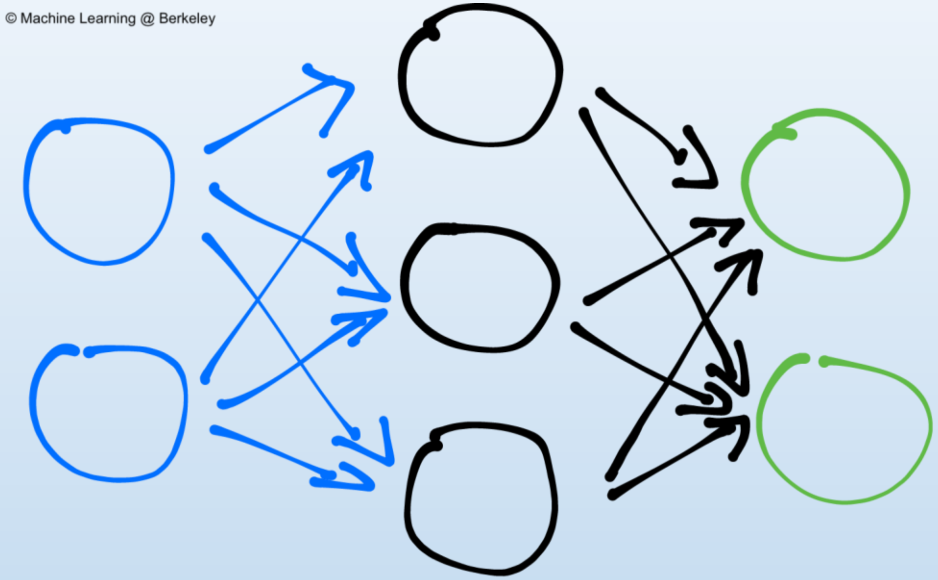


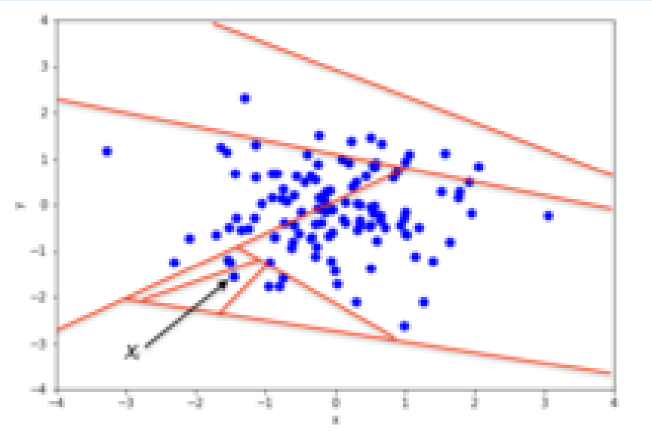
Image credit: UC Berkeley

Convolutional AutoEncoder

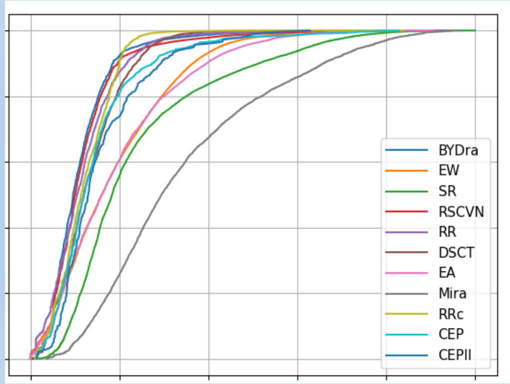
- Append colors and period

Latent Features x10

Image credit: Wikipedia.com



Isolation Forest



List of Anomalies

Results

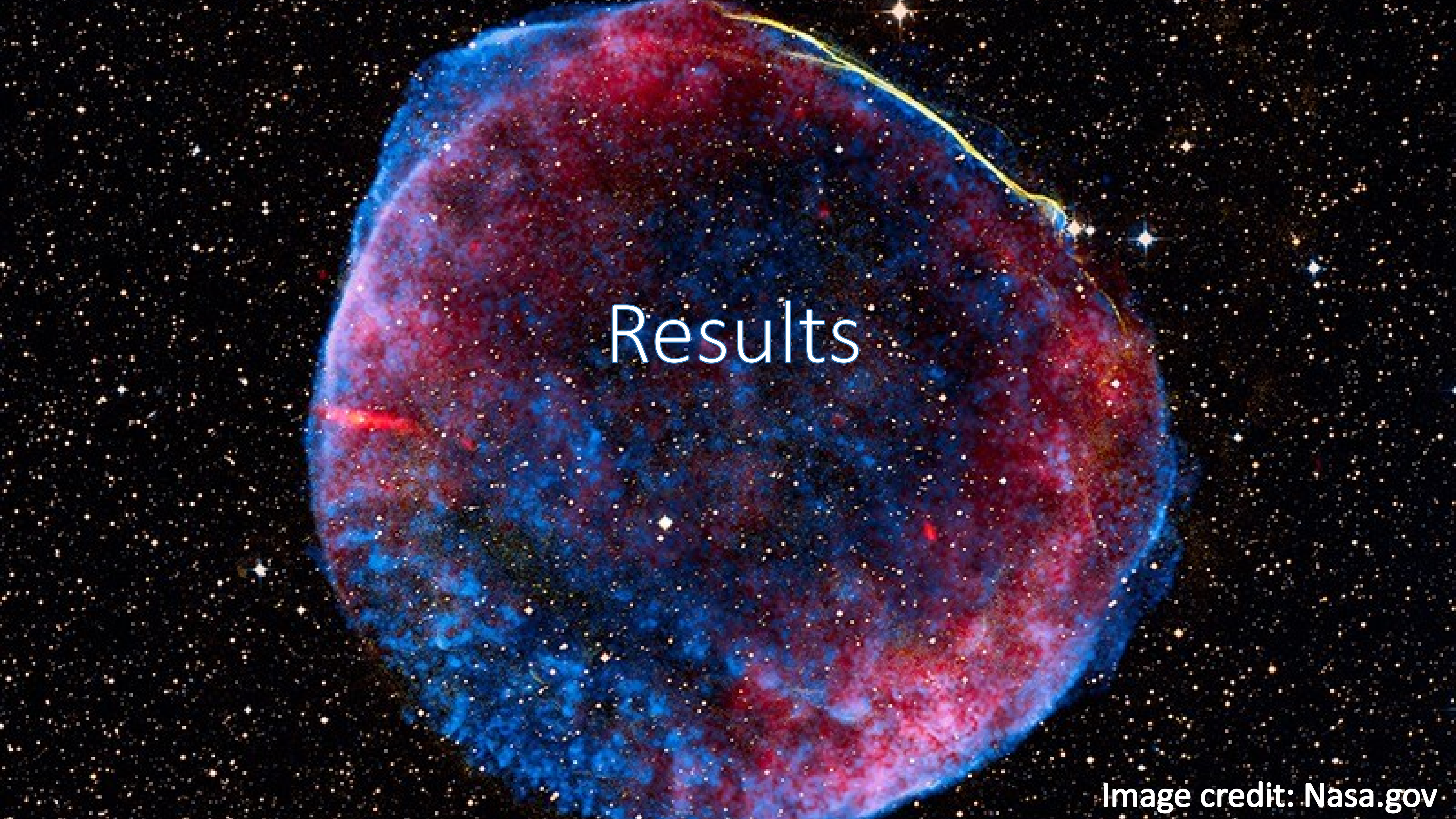
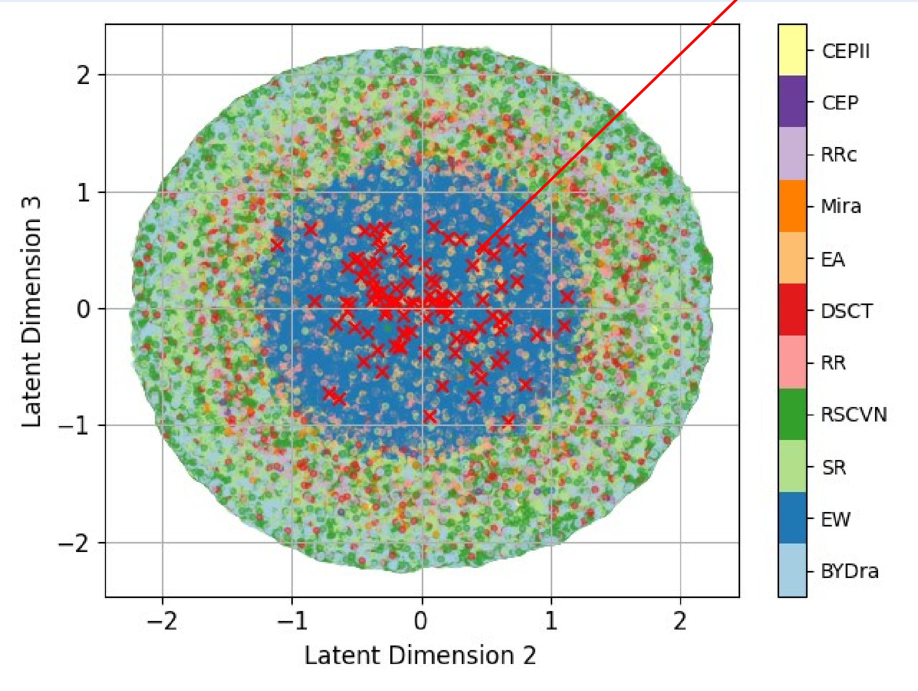


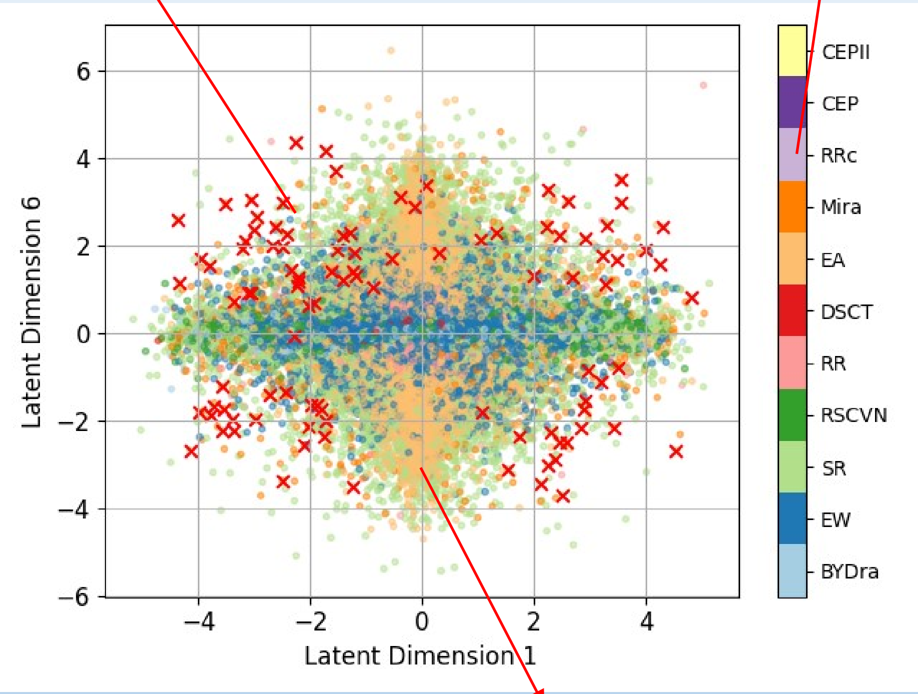
Image credit: Nasa.gov

Latent Variables Illustrations

Selected top 100 anomalies

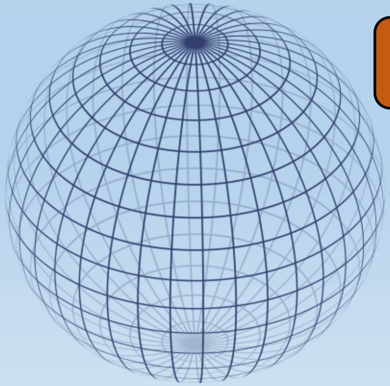
Types of variable stars





Separating from each other

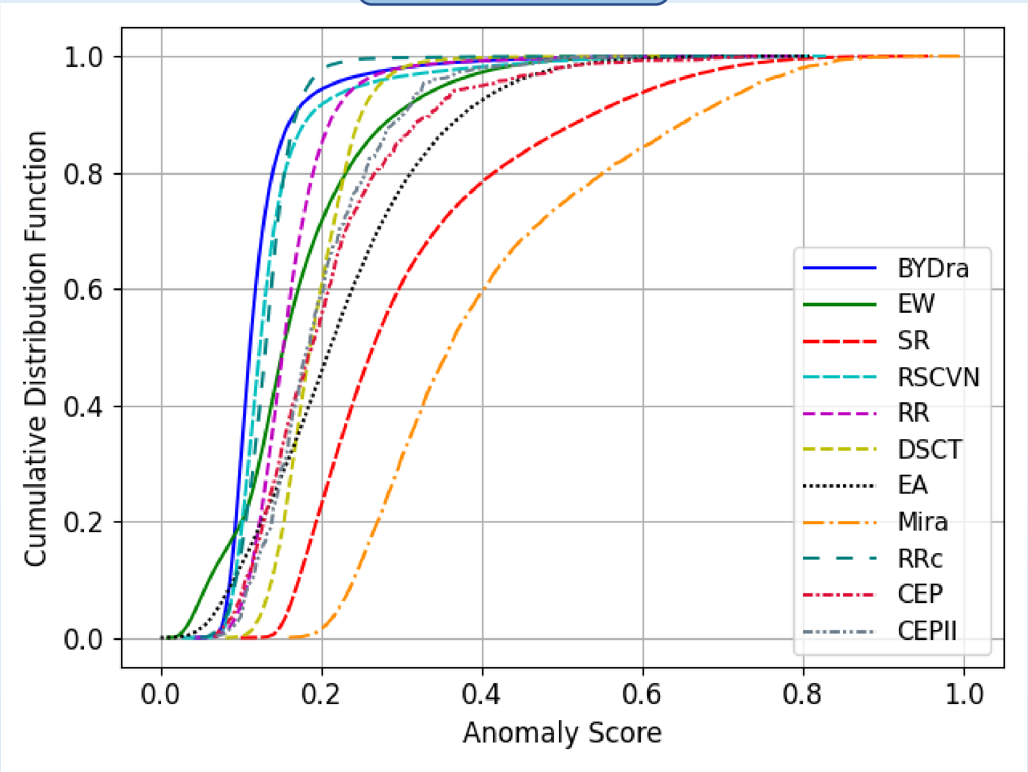
- Spherical/Annular structures
- How about transforming to a N-sphere
- Run isolation forest in the transformed space



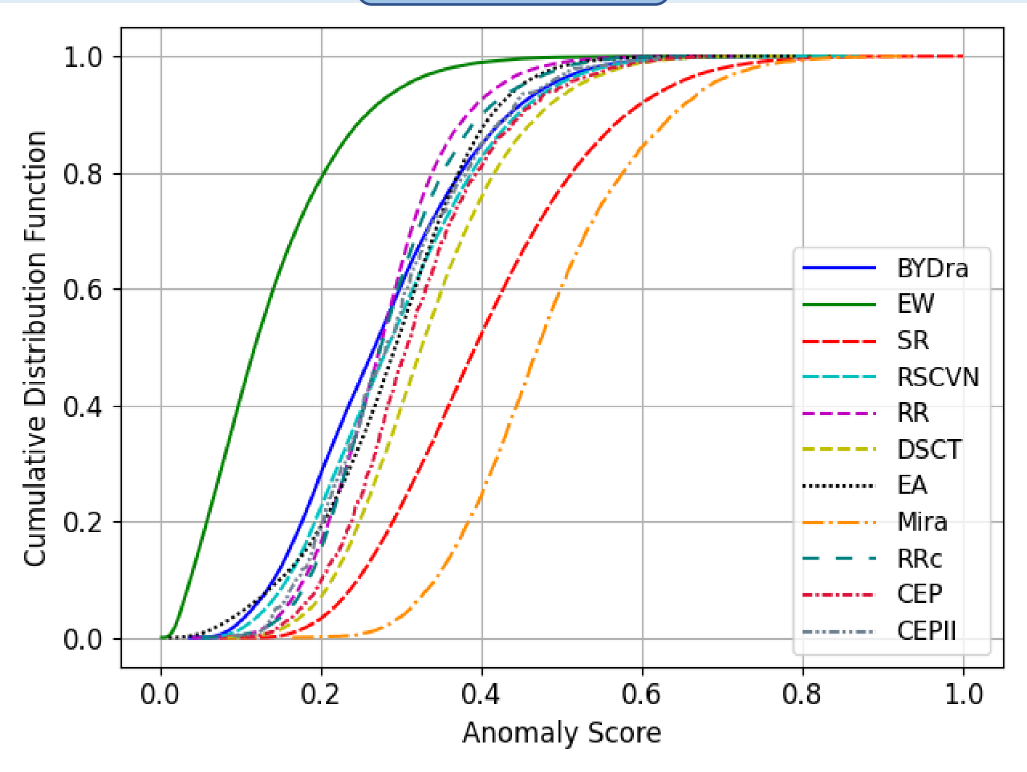
Anomaly Score Distributions

- CDF – measure fraction of objects having $s < a$

Cartesian



Spherical

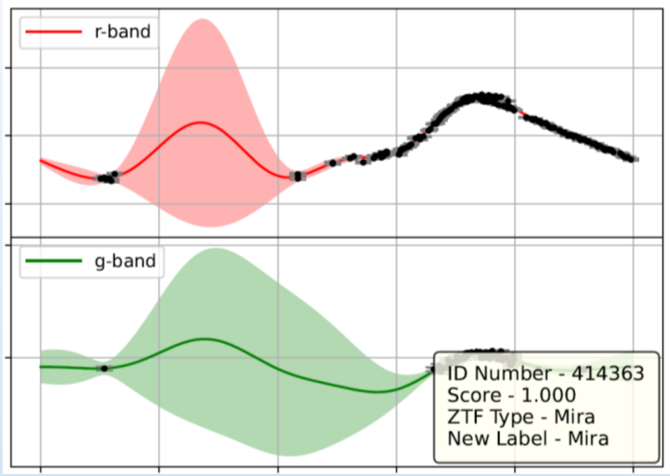


- Eclipsing W UMa (EW) is one of the least anomalous category
- Make sense – constitute most of the data
- Most anomalous – Mira and SR

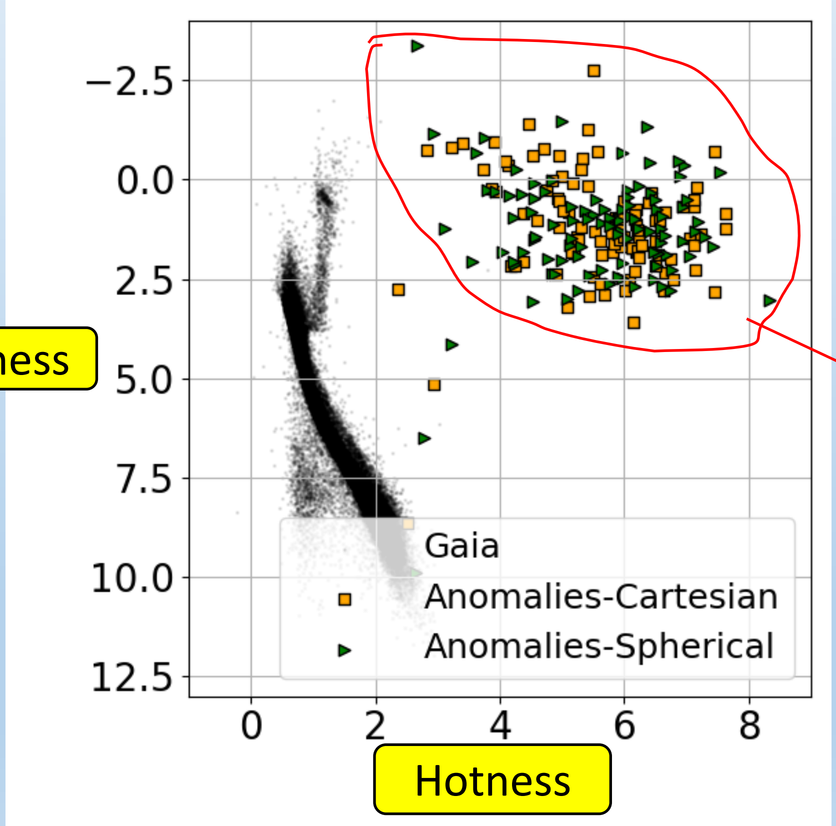
The Anomalies

- Anomalous periodic variables are
  - Irregular oscillating
  - High variability

Illustrations



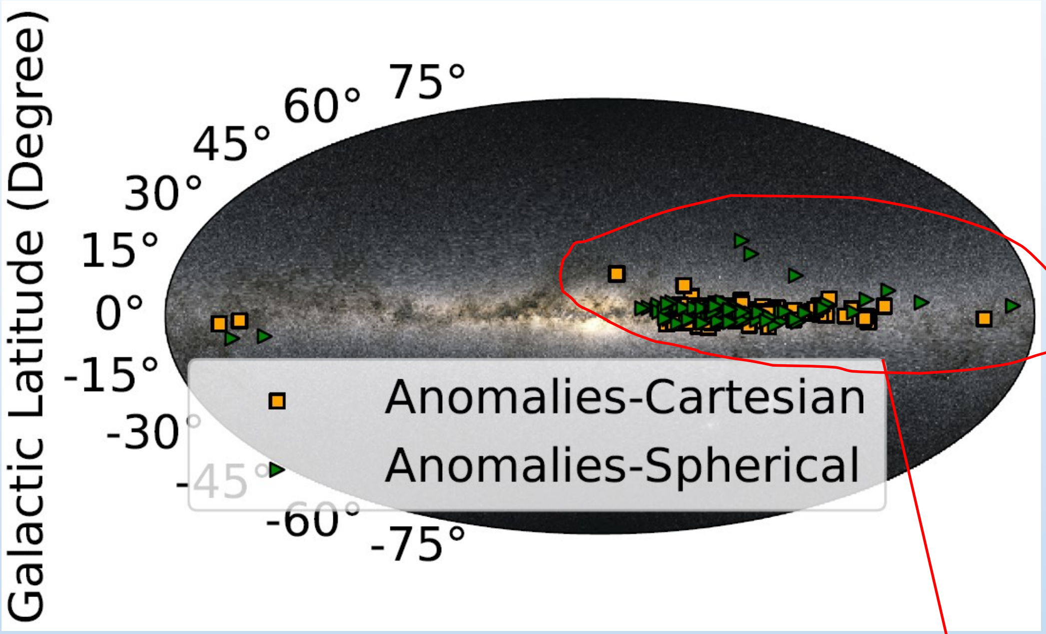
Brighter



Cooler

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

The Anomalies



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

LSST

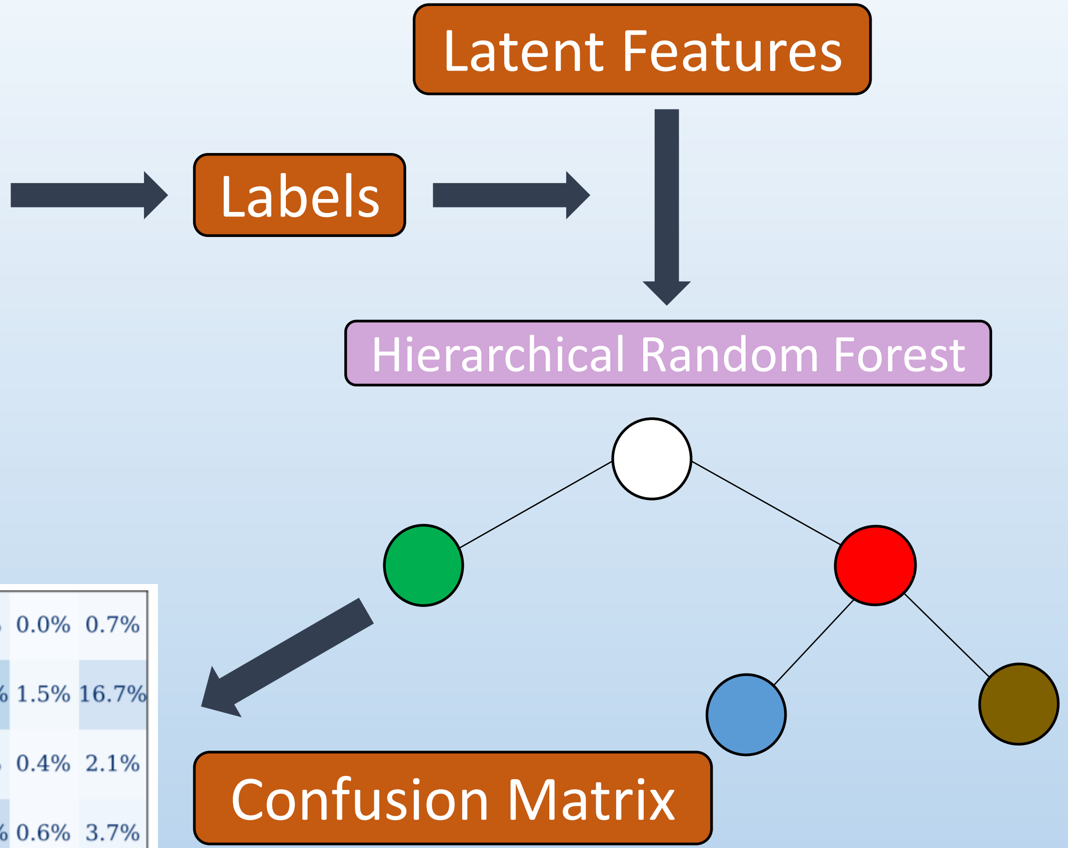
Image credit: symmetrymagazine.org

Detailed Spectroscopic Follow-Up Is Strongly Recommended!

Classifications Using The SIMBAD Labels

ZTF ⟷ SIMBAD

Crossmatch



- 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