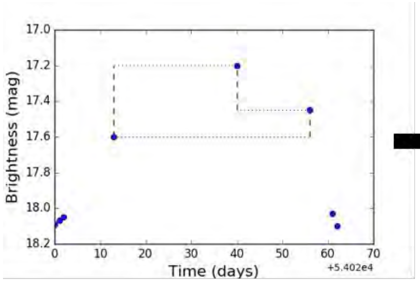
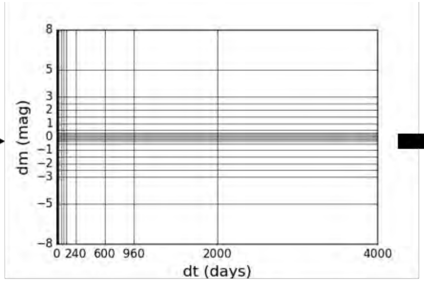
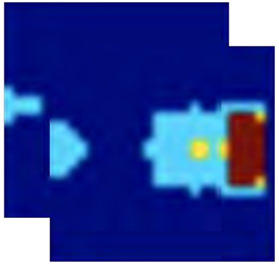
Diverse Astronomy Applications of Deep Learning



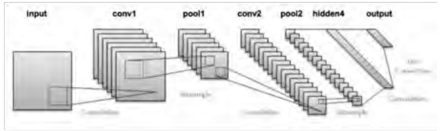
Light curves



Density representation



Equi-area images



Convolutional Neural Network

Caltech

Ashish Mahabal
AY 119, Caltech, 2025-05-06

ZTF

Broad areas (by type of data)

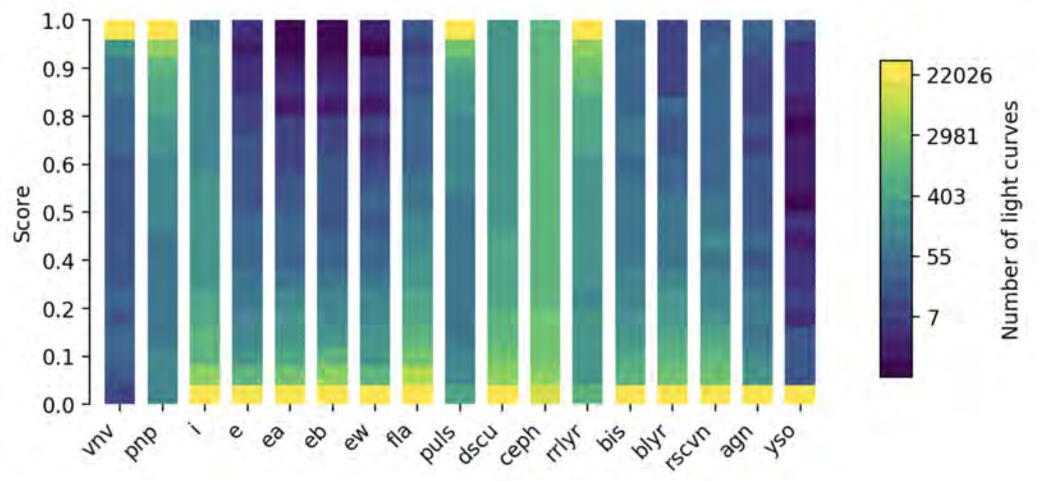
Optical astronomy

- Images
- Spectra
- Time series

Radio, X-ray etc.

Gravitational waves

Neutrino



By area of astronomy

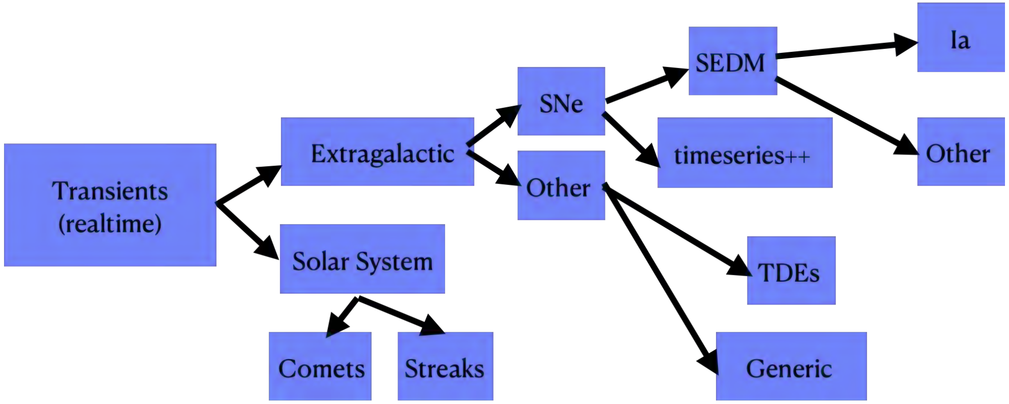
Transients

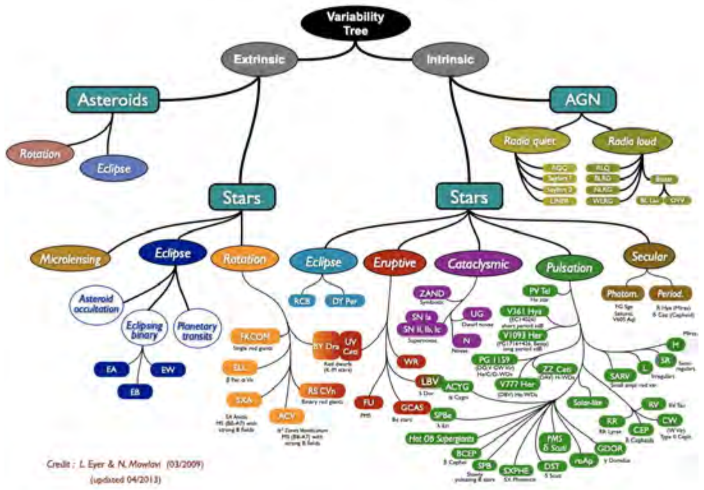
- Detection
- Classification

Variable stars

Solar System

Exoplanets





Tools

Decision trees

Random forests and variants

Convolutional Neural Networks

Variatioanal Auto Encoders

Transformers …

Always use the simplest
Do ample visualization
Check for overfitting

Optical images

SN Hunt, Galaxy Zoo ← Tools to build training samples

Zooniverse

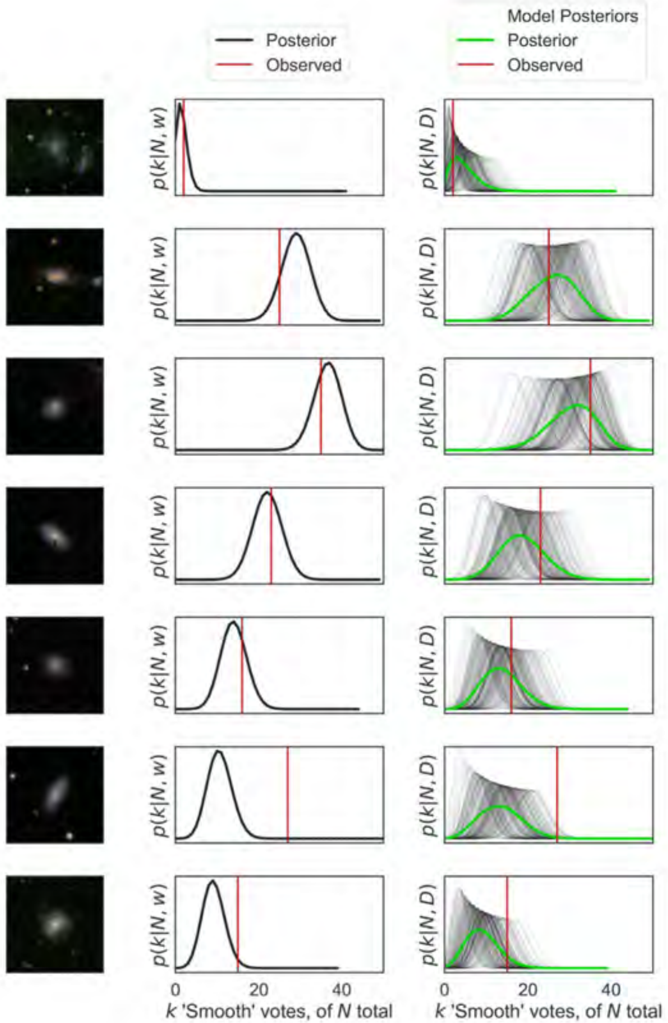
ZARTH ← Citizen science possibilities

Specialized projects close to domain expertise

- ZTF - RB
- ZTF - streaking asteroids
- Roman (example dataset on GitHub)

Bayesian CNNs

Walmsley et al. 2020

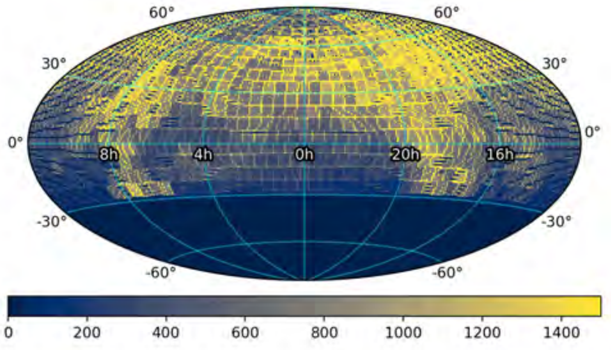


Smooth/featured galaxy - first Q in Zooniverse DT

Bayesian Active Learning by Disagreement, BALD

Zwicky Transient Facility

Systematic Exploration of the Dynamic Sky

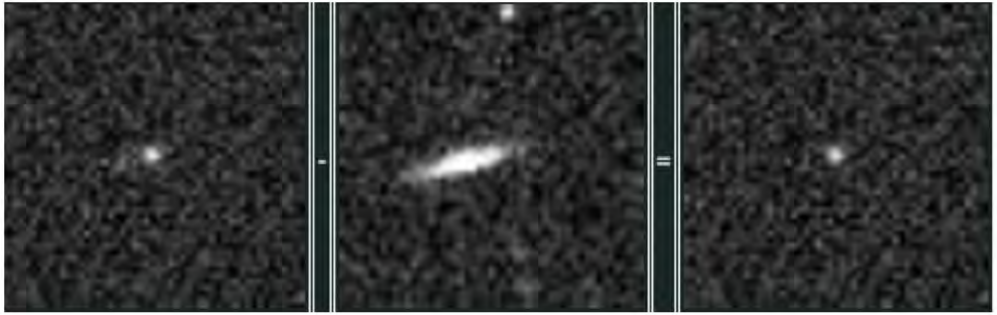


Sky coverage and number of observation epochs in ZTF.

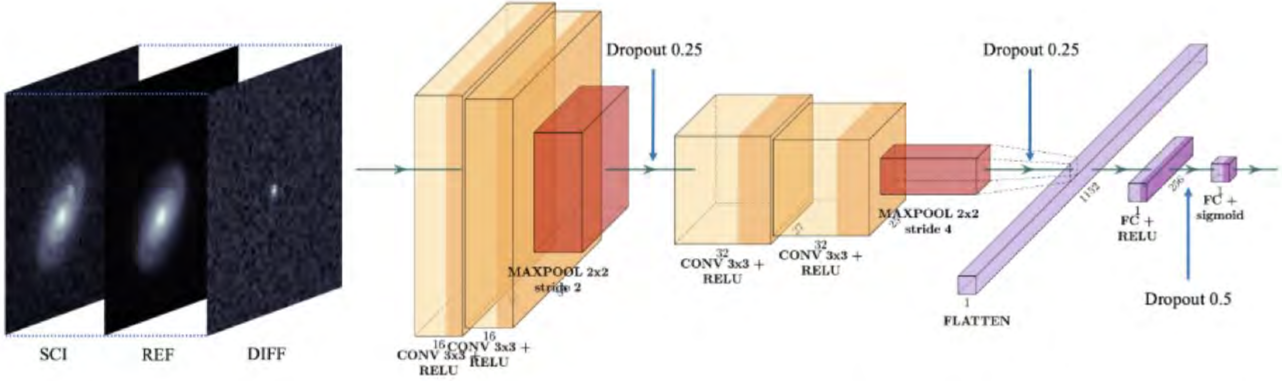
Example Query using the APIs

wget "https://irsa.ipac.caltech.edu/ibe/search/ztf/products/sci?POS=255.9302,11.8654&WHERE=obsjd>2458219.9678+AND+obsjd<2458228.8155+AND+infobits<33554432" -O out.tbl

ZTF **Real-Bogus separation ('braai')**

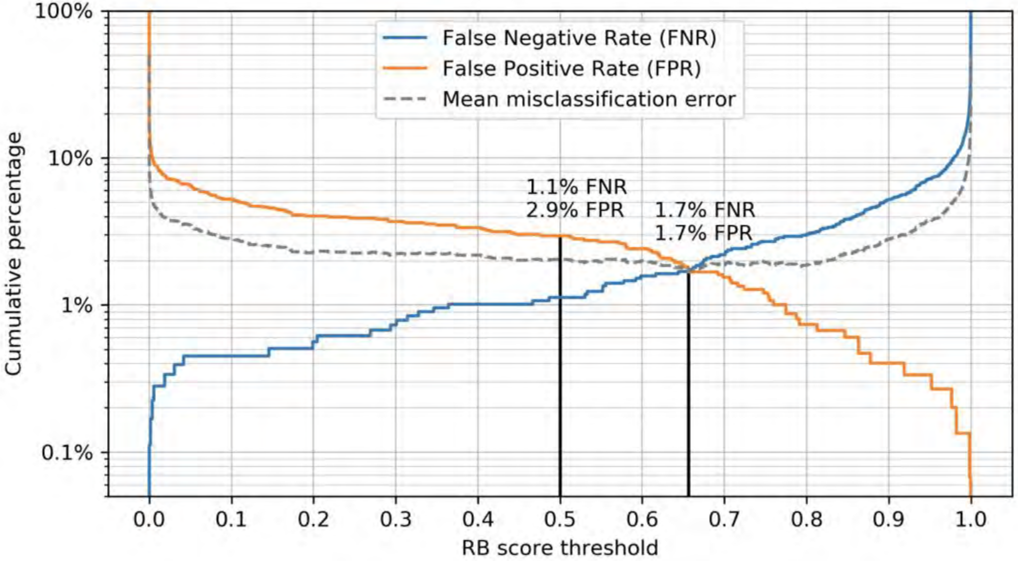


Sci **Ref** **Diff**



https://github.com/dmitryduev/braai

Duev, Mahabal, ... arXiv 1907.11259



$$A = \frac{TP + TN}{TP + FP + TN + FN}$$

Total predictions

How often is the model correct?

$$P = \frac{TP}{TP + FP}$$

Predicted positives

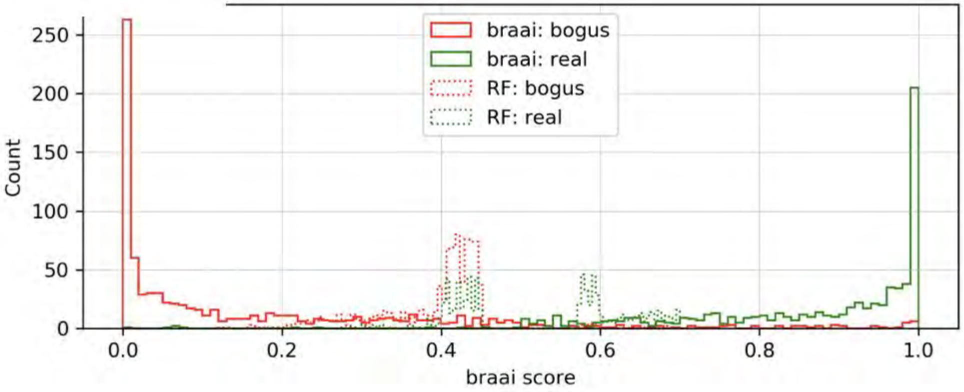
How often is the model correct when it predicts that the candidate is real?

$$R = \frac{TP}{TP + FN}$$

Actual positives

How many real candidates are predicted correctly?

ACAI filters for Fritz
b, h, n, o, v



Deep Learning with AStreaks

This is how a real asteroid would look. Short streak.

These are ghosts and dementors

A satellite trail. Note that part of it is masked out, and the unmasked trail is longer.

Another satellite trail

A masked bright star

What kind of streak do you see?

- Asteroid (short streak)
- Satellite (long streak - could be partially masked)
- Masked bright star
- Dementors and ghosts
- Cosmic rays
- Naked stars
- Yin-Yang (multiple badly subtracted stars)
- Skip (Includes 'Not Sure' and seemingly 'Blank Images')

Need some help with this task?

Done

Show the project tutorial

ZTF DeepStreaks

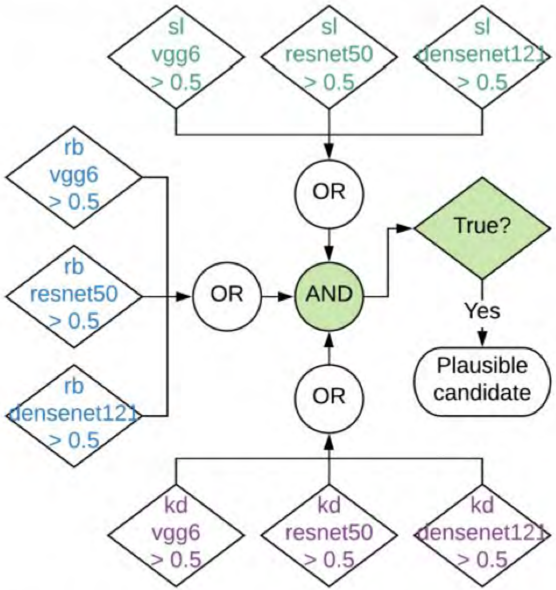
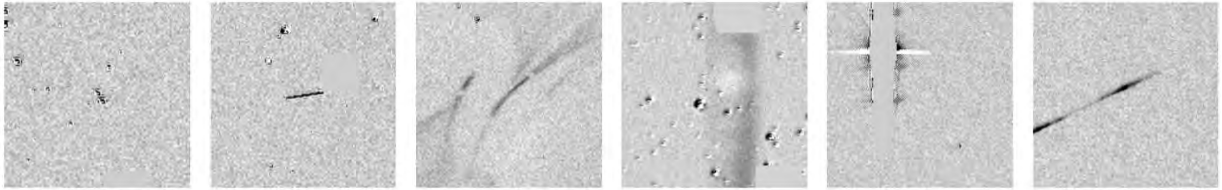
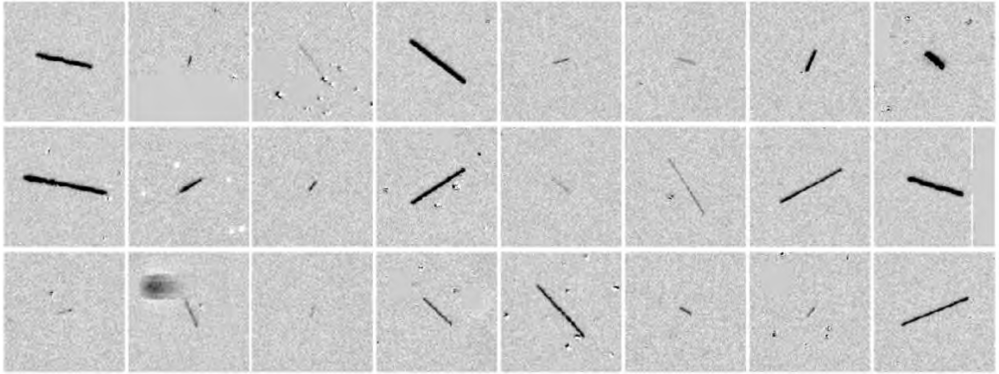


Figure 2. Decision logic used by `DeepStreaks` to identify plausible streaks. The problem is split into three simpler sub-problems, each solved by a dedicated group of classifiers assigning real vs. bogus ("rb"), short vs. long ("sl"), and keep vs. ditch ("kd") scores. At least one member of each group must output a score that passes a pre-defined threshold. See Section 2.1 for details.



(a) Bad subtraction (b) Cosmic ray (c) "Dementor" (d) "Ghost" (e) Masked star (f) Satellite trail

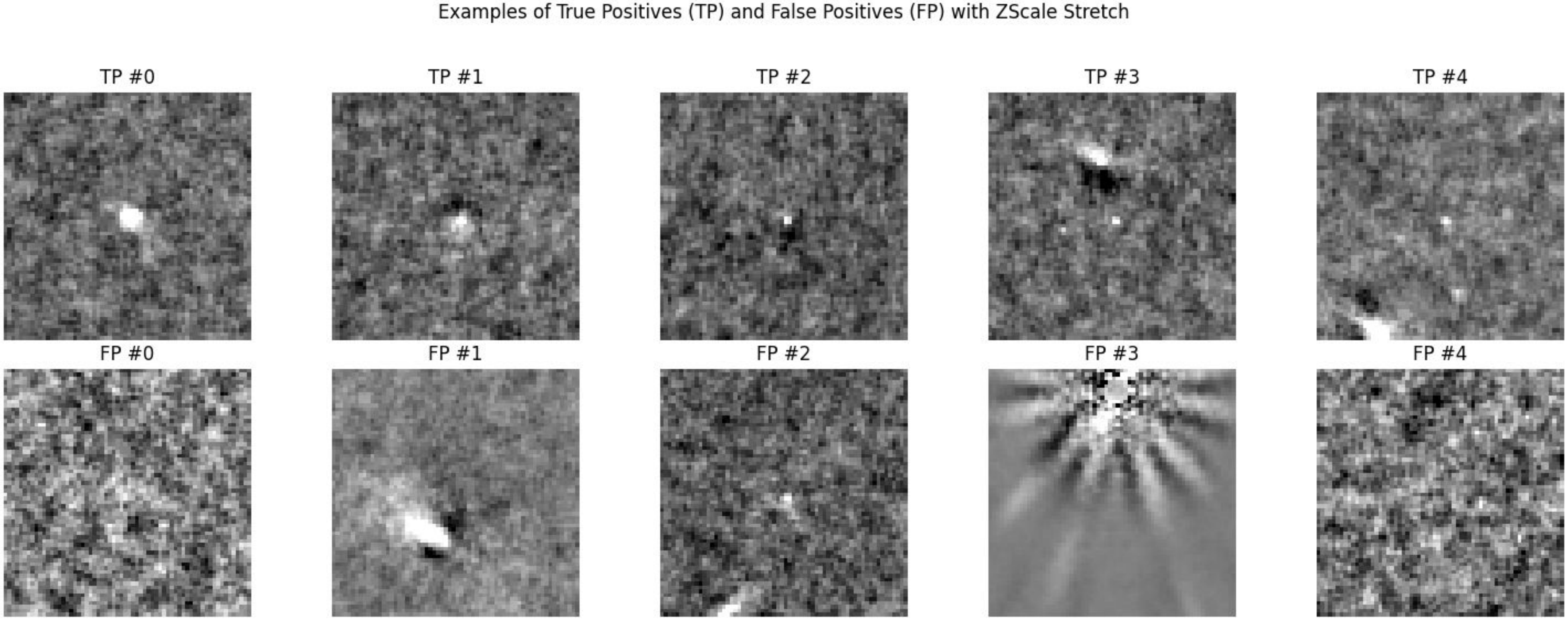


CNNs

Duev, Mahabal, ... arXiv:1904.05920

X

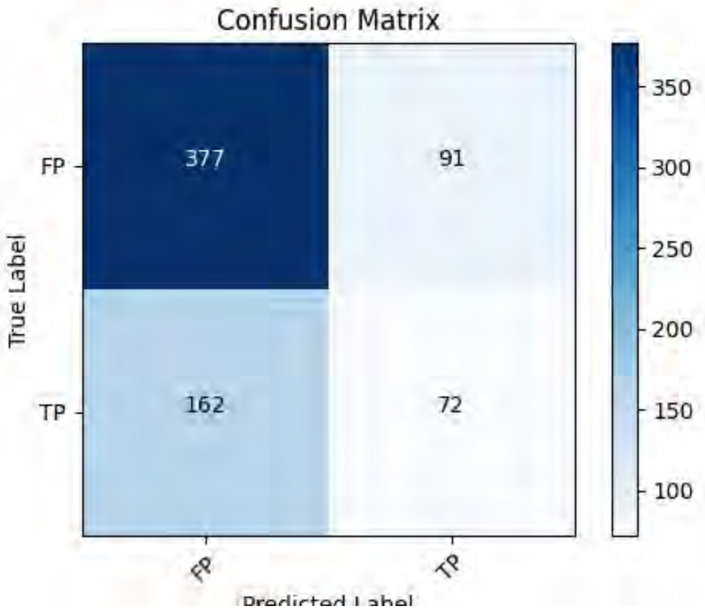
Training sample (True Positives and True Negatives)



~1500 TP and ~3000 FP used right now

Roman Simulations

Out-of-the-box poor results



Issues with sample size, purity, normalization, …

Time series

Irregular

Large gaps

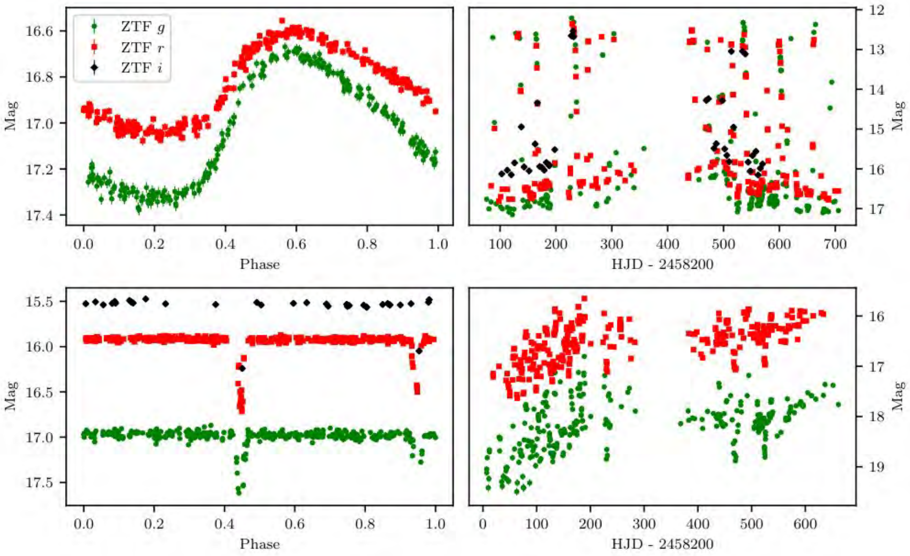
Heteroskedastic

Important tasks:

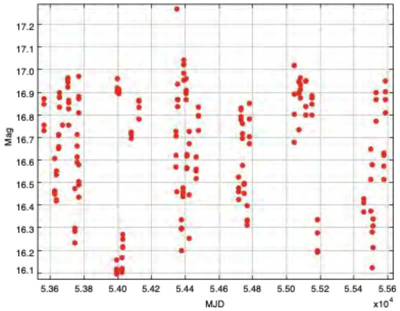
Classification

Period finding

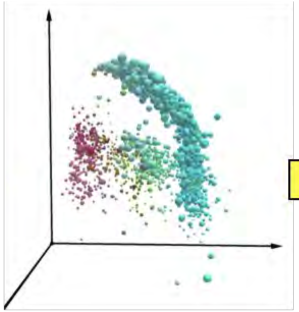
Prediction



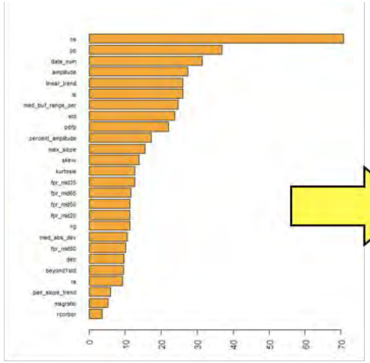
Light curves (time series) are the primary currency



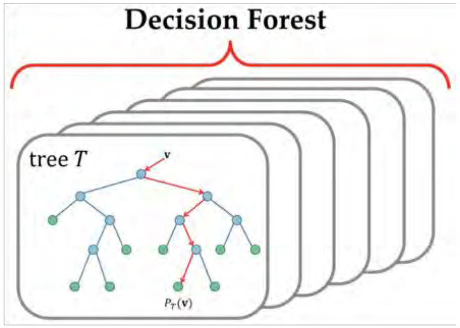
Light curves



Feature vectors



Dimensionality Reduction

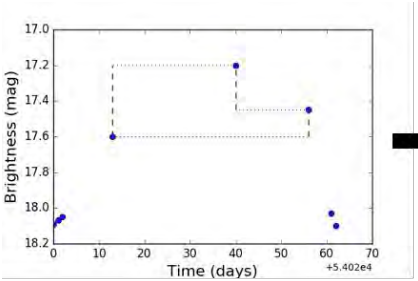


Classification

Domain knowledge/subjectivity

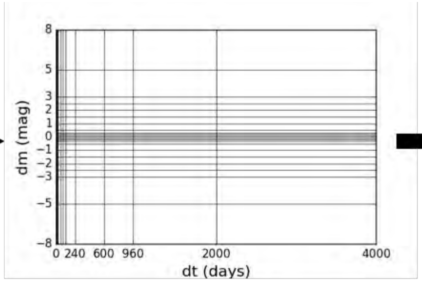
Survey Differences: area, bands, cadence, depth, exposure, ...

Light curves are typically heteroscedastic, sparse, irregular



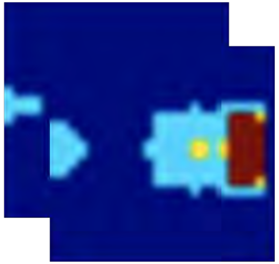
Light curves

n points

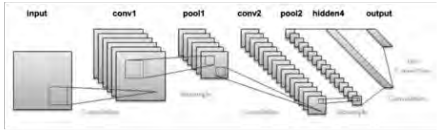


Density representation

n * (n-1)/2 points



Equi-area images



Convolutional Neural Network

Use the CNN hammer

dmdt images

RR Lyrae

RS CVn

LPV

Mahabal, Sheth et al., 1709.06257

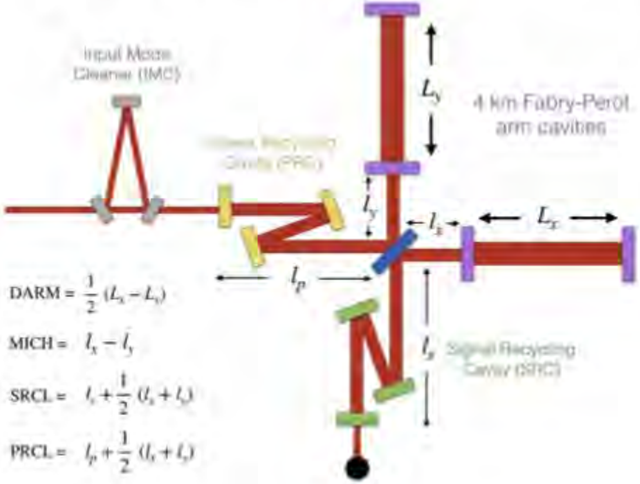
Diagnosing LIGO lockloss using auxiliary channels

Motivation

- Lockloss events due to environmental events lead to loss of observation time
- Monitor and diagnose lockloss events as they occur

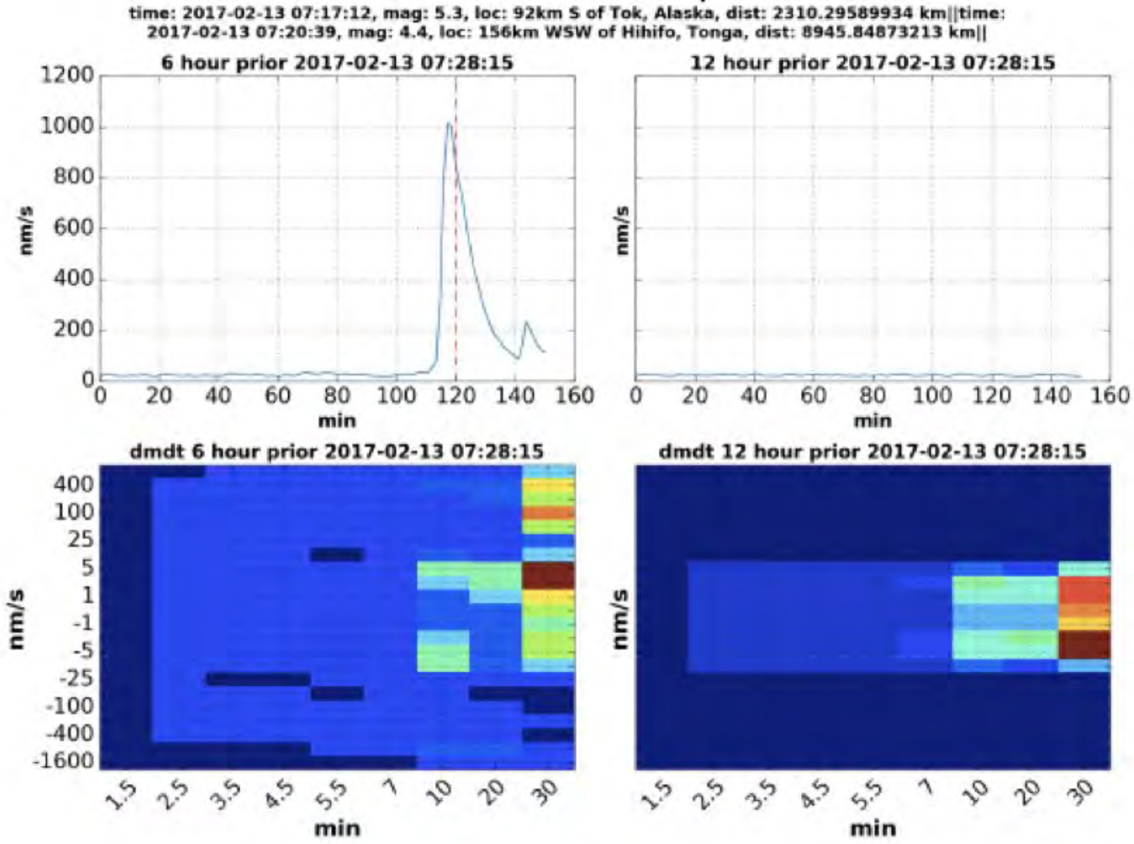
Goals

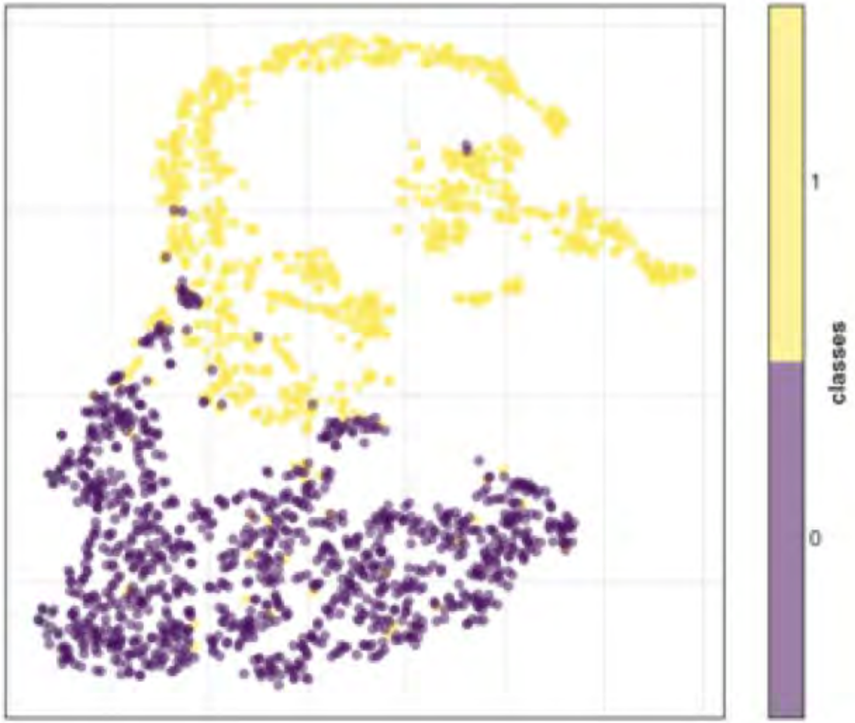
- To find a minimal set of auxiliary channels that serve as good predictors for lockloss events
- Diagnosis of interferometer behavior leading to lockloss events



With Ayon Biswas and Jess McIver

Effect of earthquakes





POP+SRCL+MITCH

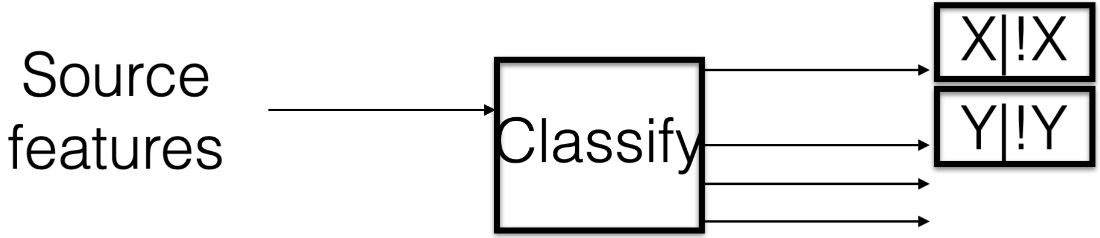
Nearly clean separation of lock-loss events in GW detectors using cavity channels

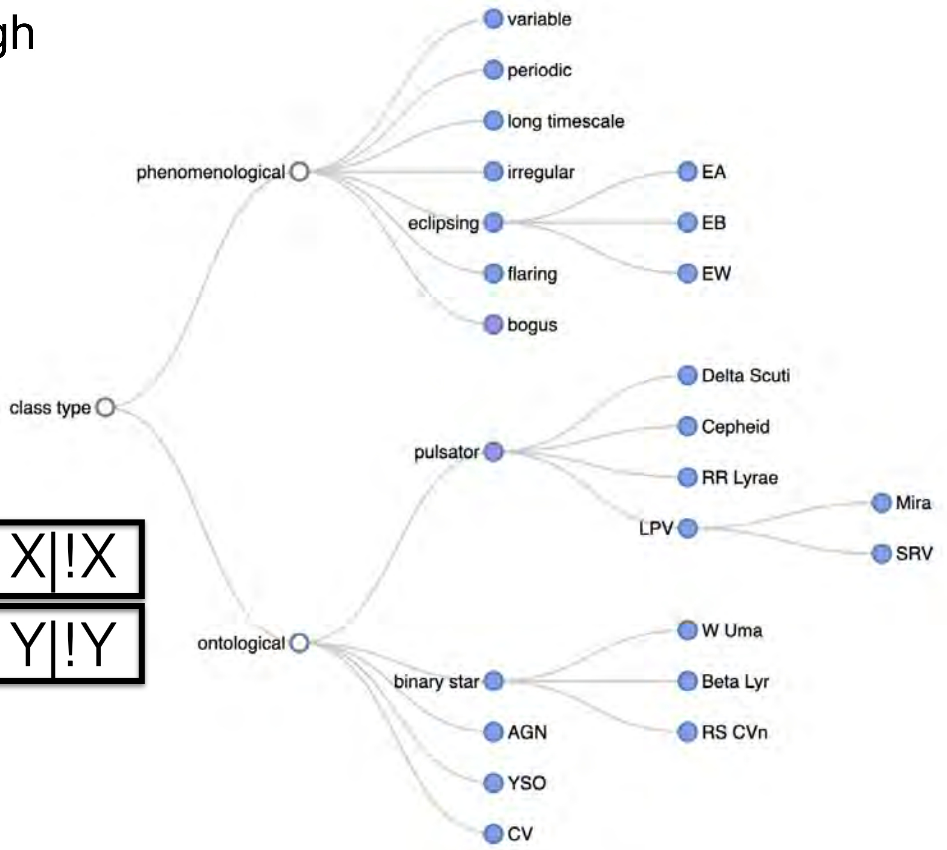
Hierarchical/stackable Classification Through Independent Binary Classifiers

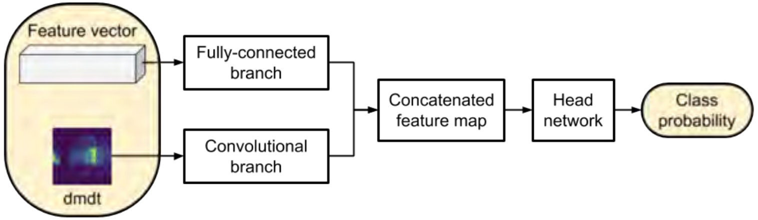
Since the classifiers are independent

no restrictions on p(sum)

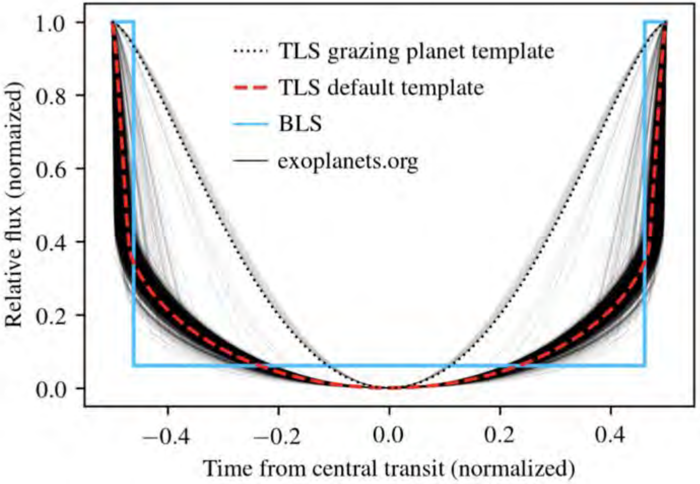
In particular it may not be 1





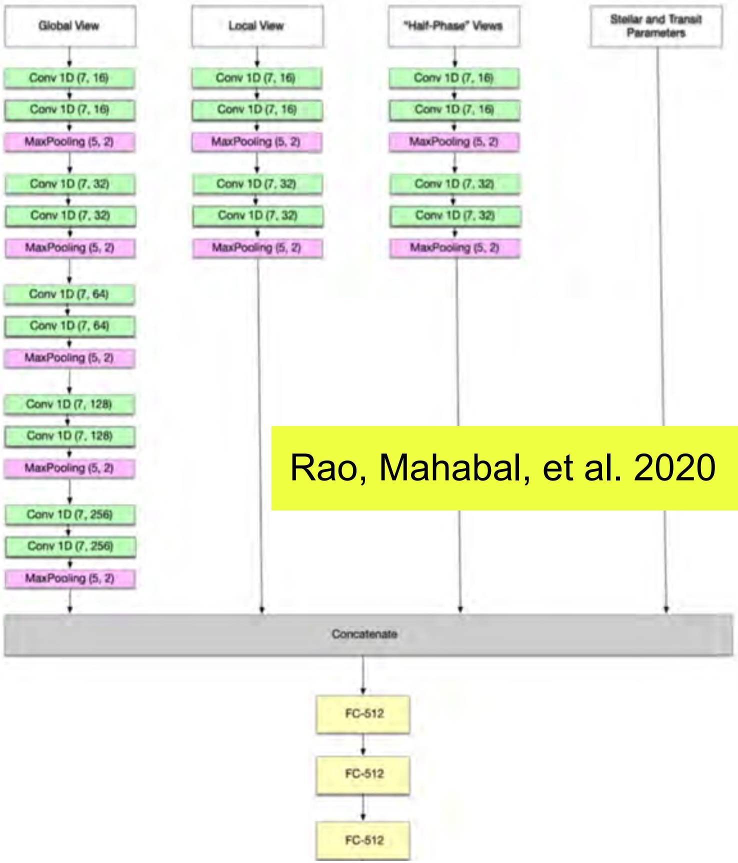


Exoplanet candidates using TESS

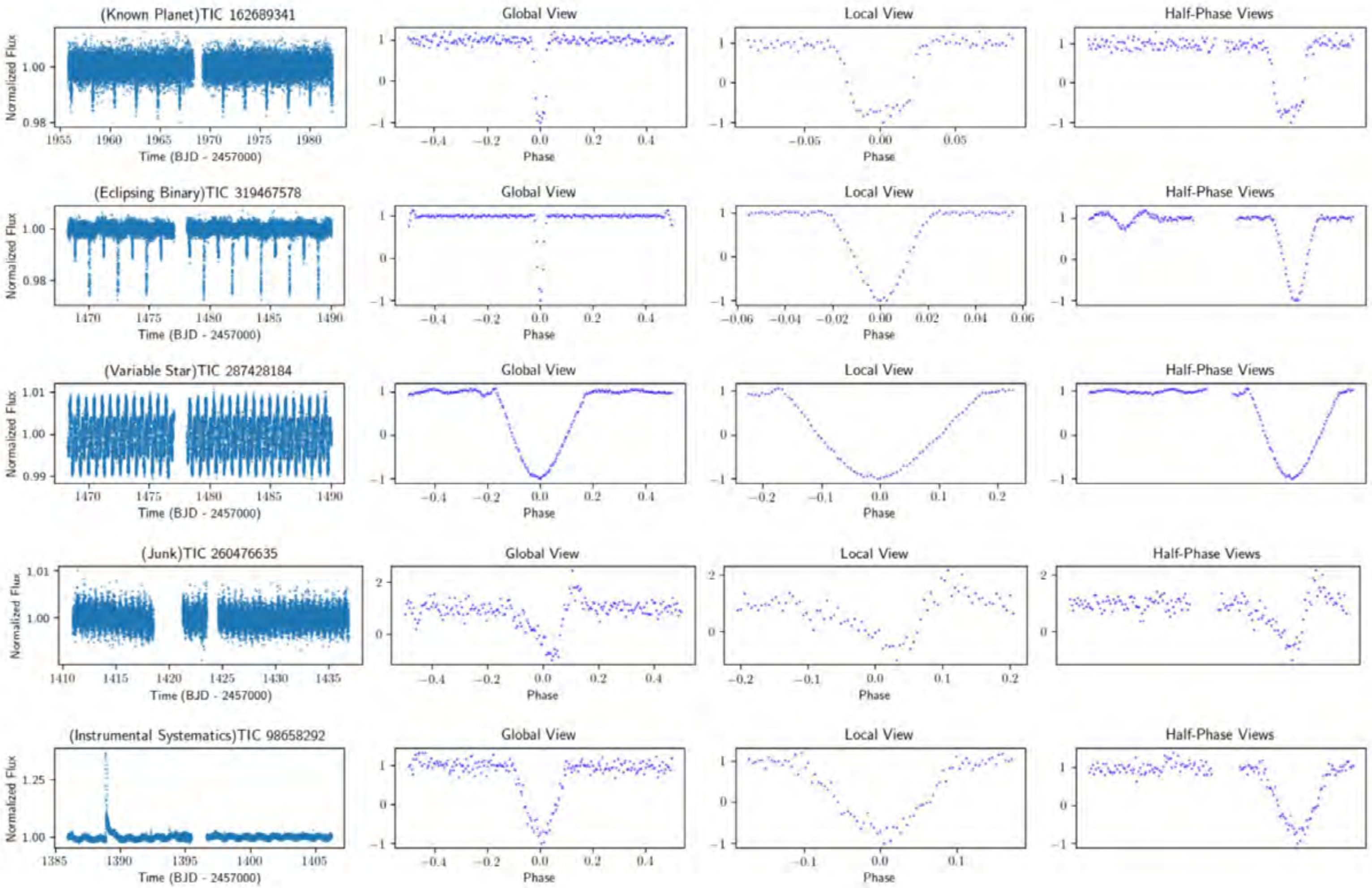


Box Least Squares (BLS - Kovács et al. 2002) - box-like transits

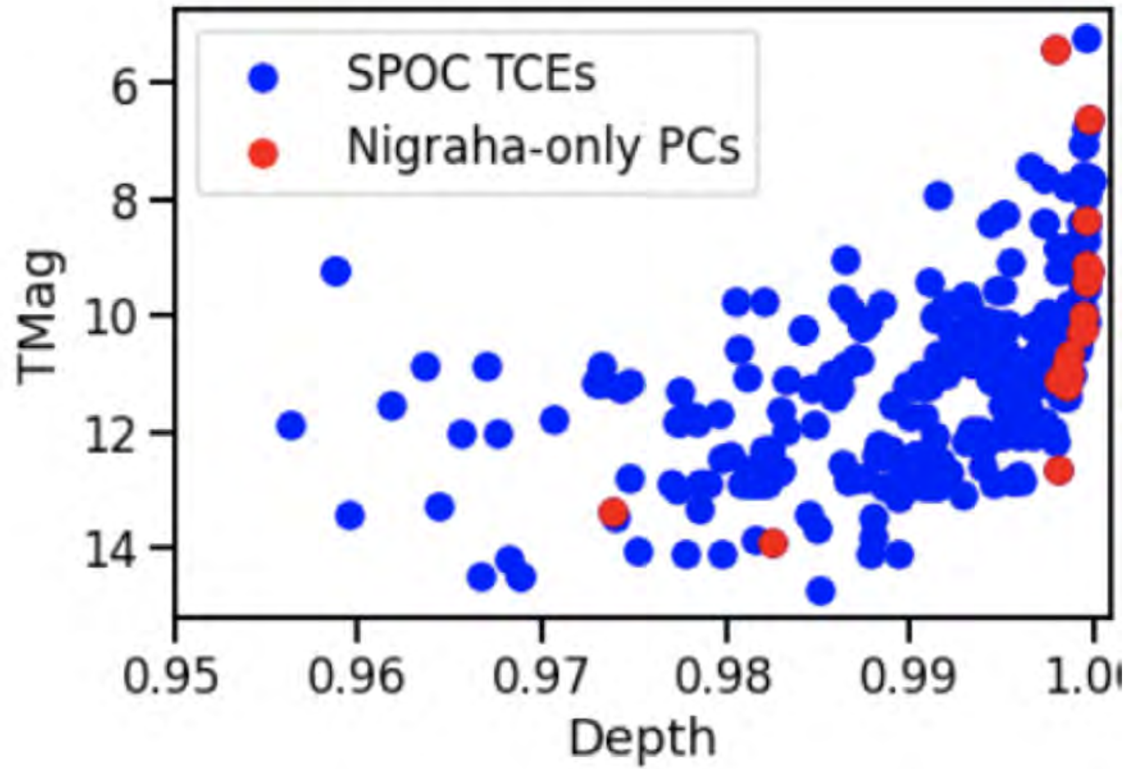
Transit Least-Squares (TLS - Hippke & Heller 2019) - shallow transits



Rao, Mahabal, et al. 2020


(Known Planet)TIC 162689341
Global View
Local View
Half-Phase Views
(Eclipsing Binary)TIC 319467578
Global View
Local View
Half-Phase Views
(Variable Star)TIC 287428184
Global View
Local View
Half-Phase Views
(Junk)TIC 260476635
Global View
Local View
Half-Phase Views
(Instrumental Systematics)TIC 98658292
Global View
Local View
Half-Phase Views
Normalized Flux
Time (BJD - 2457000)
Phase

Nigraha: Exoplanet candidates using TESS



Accuracy: 87.2%
Precision: 88.8%
Recall: 74.3%

HiRes/AO observations confirm non-binary

Rao, Mahabal, et al. 2020

Classifying spectra

Regularly spaced

Complications like redshift

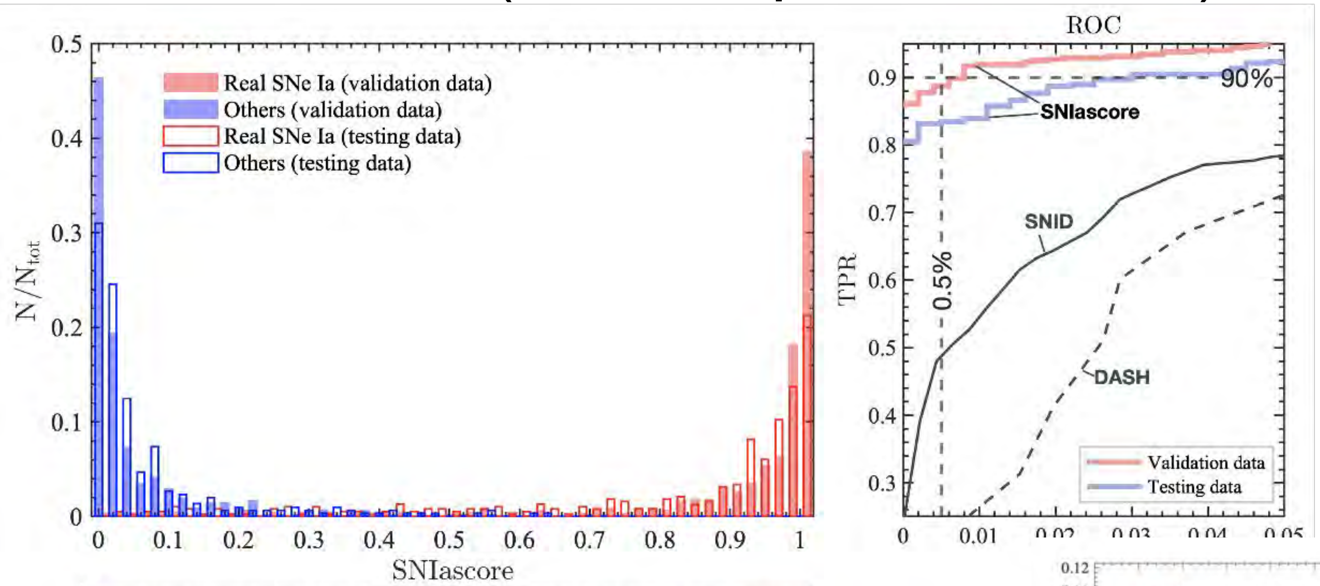
Examples:

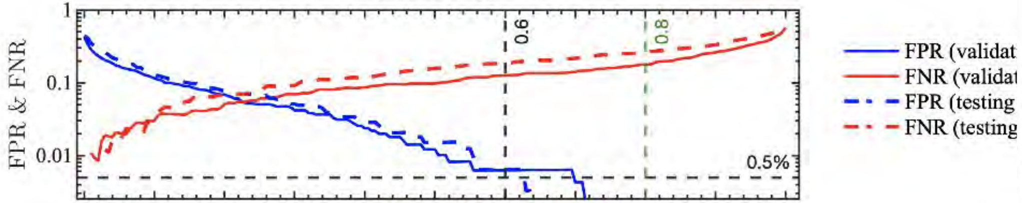
ZTF SN Ia

ZTF all SN

Sasha’s project

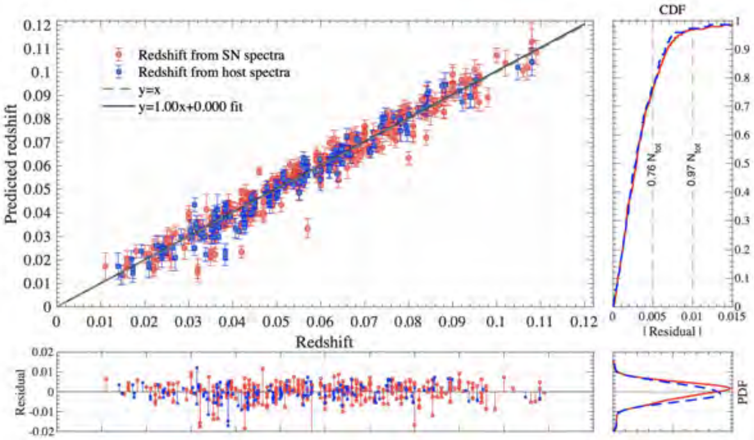
SN Ia classifier (SEDM Spectra ⇒ TNS)





C Fremling, D. Neill, Y Sharma, ...

Fremling et al. arXiv:2104.12980

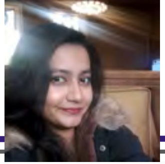


High dropout RNN

<1% FPRs
10 minutes
end-to-end

All SNe

Yashvi Sharma



Spectra

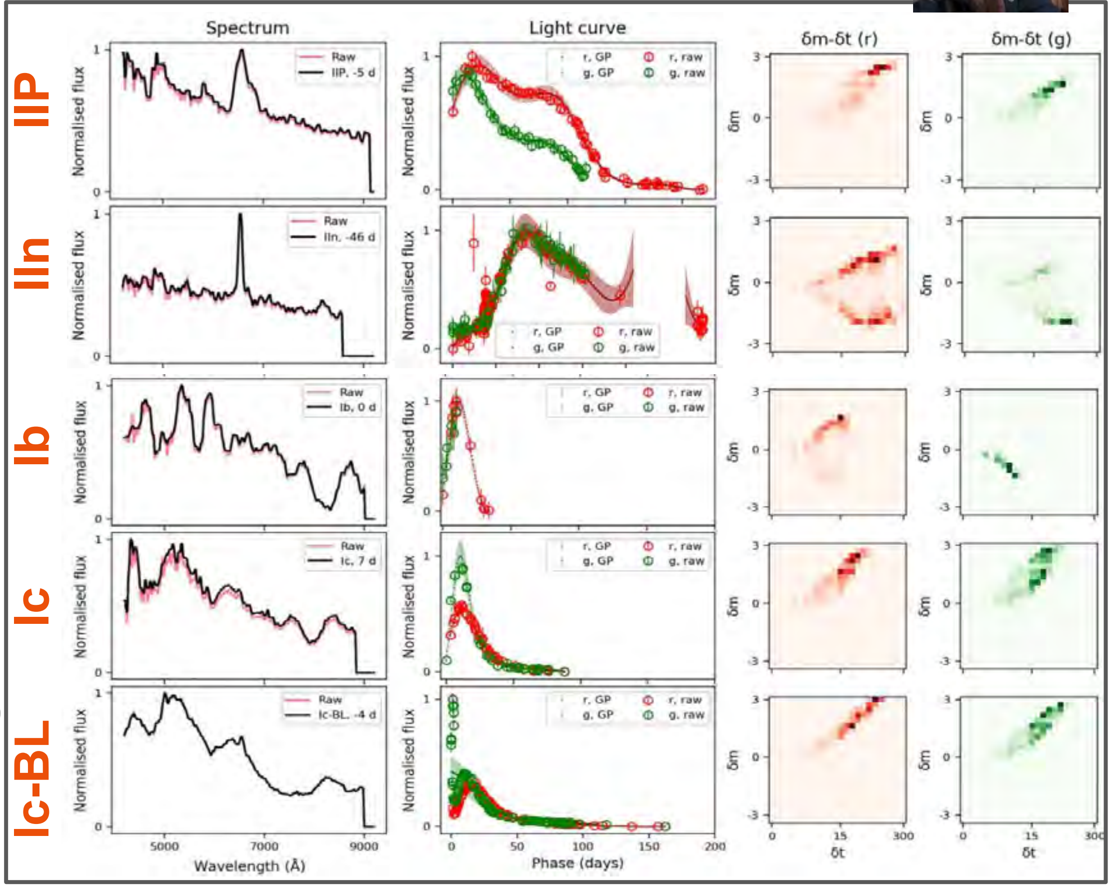
- Smoothed and normalised
- Option to deredshift

1D light curves (*g* and *r* bands)

- Interpolated to fix length (200 days) using Gaussian process regression
- Converted to linear flux
- Kept photometry with SNR > 3
- Normalised

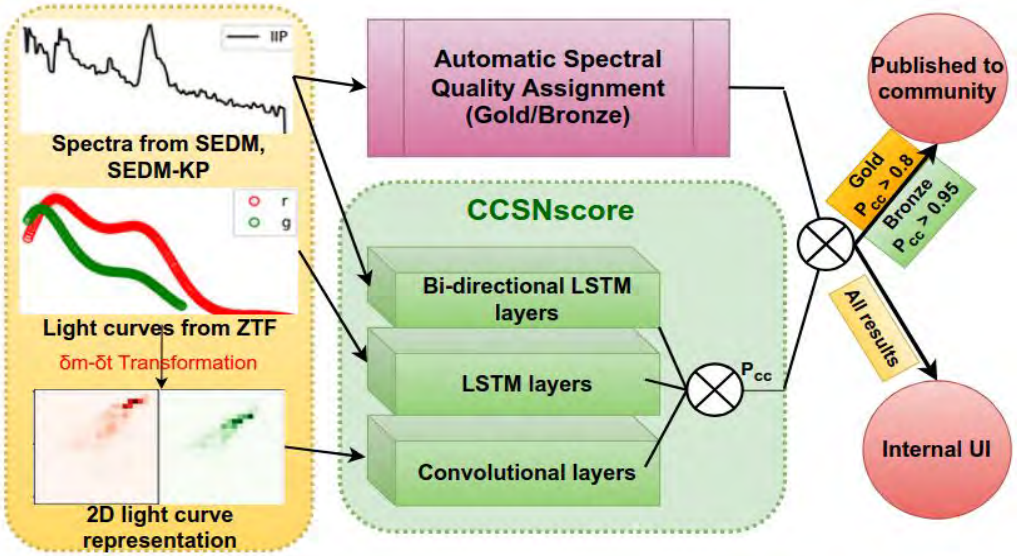
δm-δt representation

- 2D histogram created by taking pair-wise magnitude and time (phase) difference of interpolated LCs
- Normalised

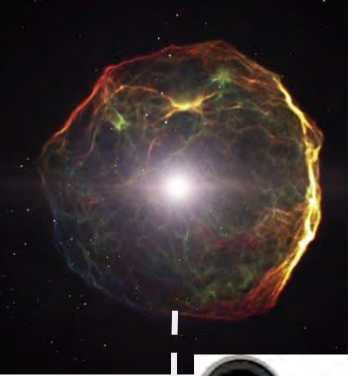


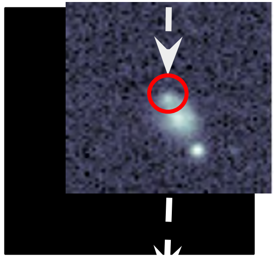
Combined model

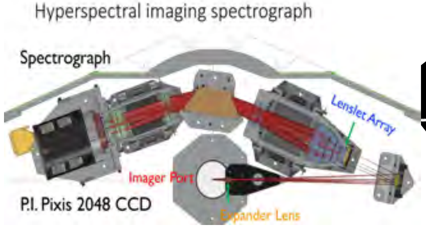
- ★ For SN II vs. SN Ibc classification task
  - With **'only spectra'** channel, 79.4% of the gold test set gets classified with >0.98 scores, out of which **98.7%** are accurate
  - With **'spectra+1D LC'**, 82.8% of the gold test set gets classified with >0.98 scores, out of which **98.0%** are accurate
- ★ Expected real-time performance (TNS reporting)
  - 0.5% of true SNe Ia likely will get misclassified as CCSNe
  - 62% of true CCSNe expected to pass TNS reporting criteria, out of which 94% will get correctly classified

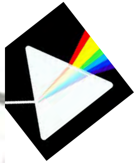


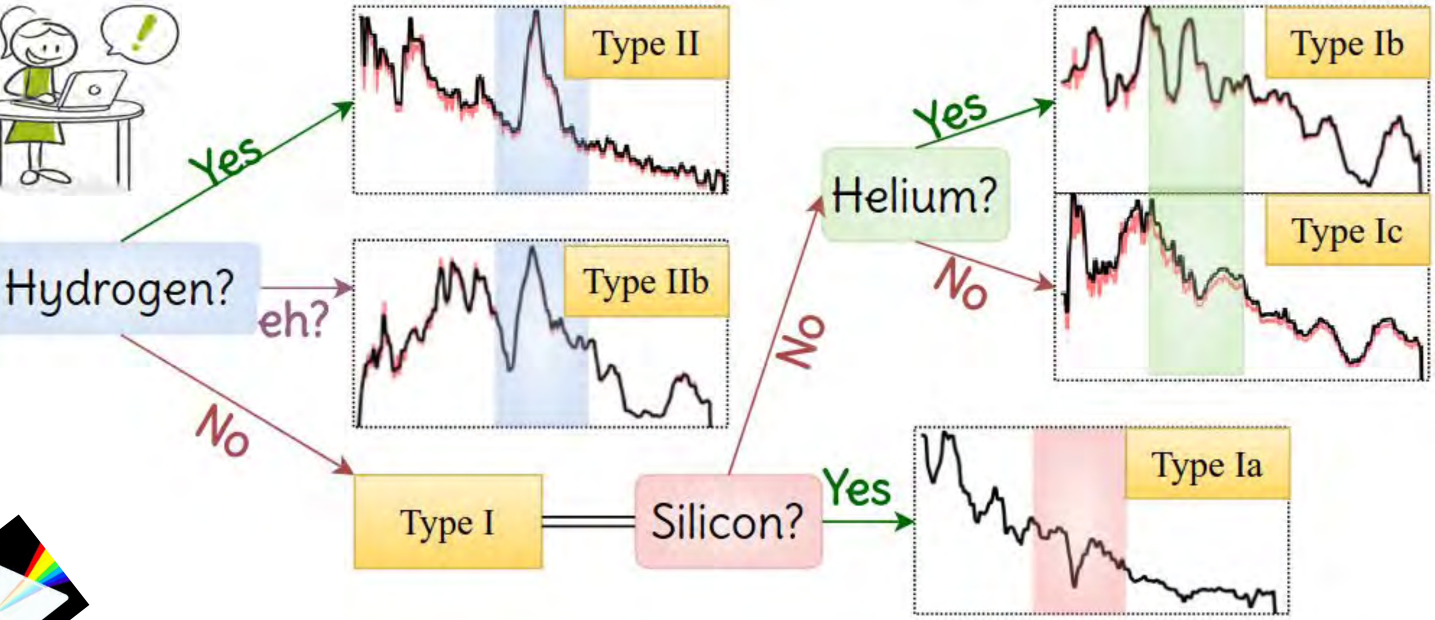
The Supernova Zoo







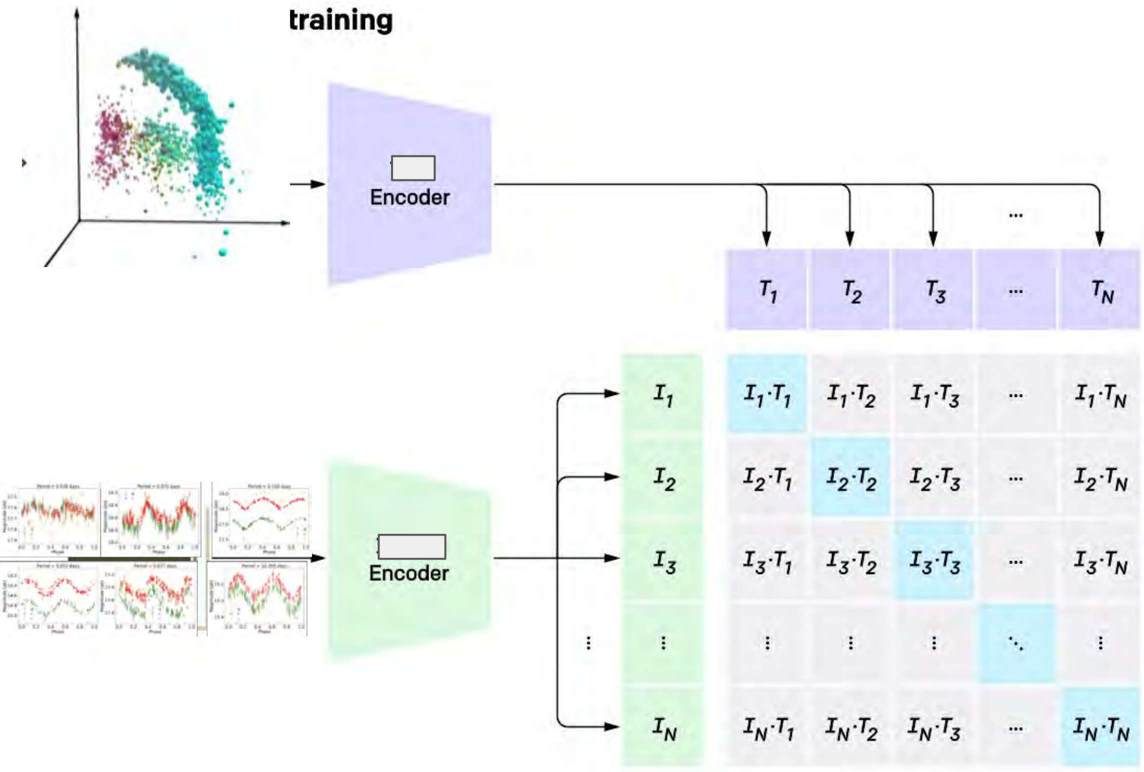




SEDM

Variational Auto Encoders

Potential for science: immense



Contrastive Language-Image Pre-training (CLIP)

https://openai.com/blog/clip/

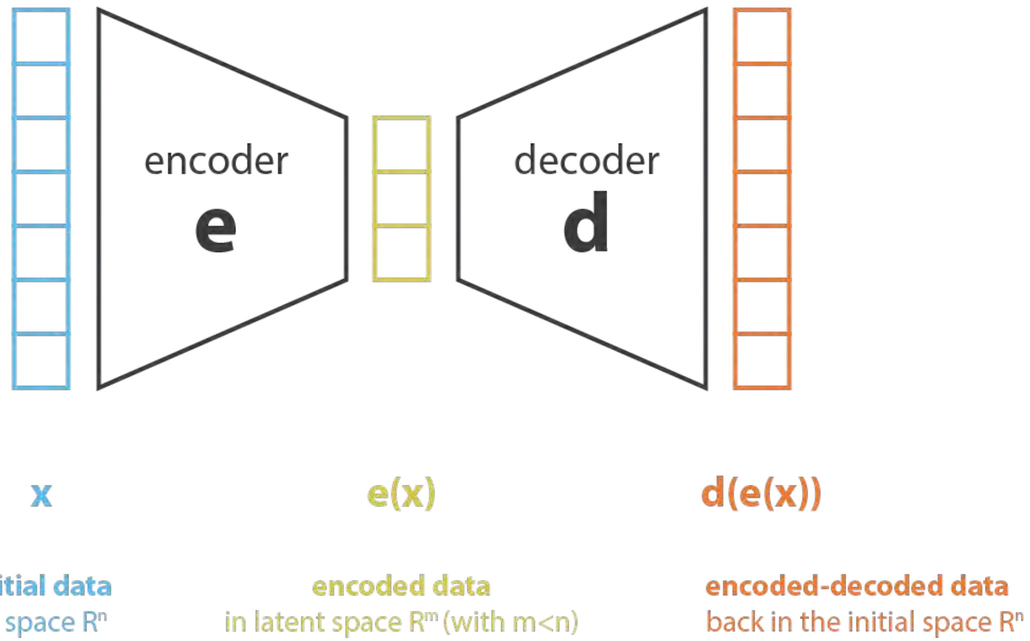
Overview of VAEs

Basic concept

- Encoder maps data to latent space
- Decoder reconstructs data from latent space

Encoder and Decoder

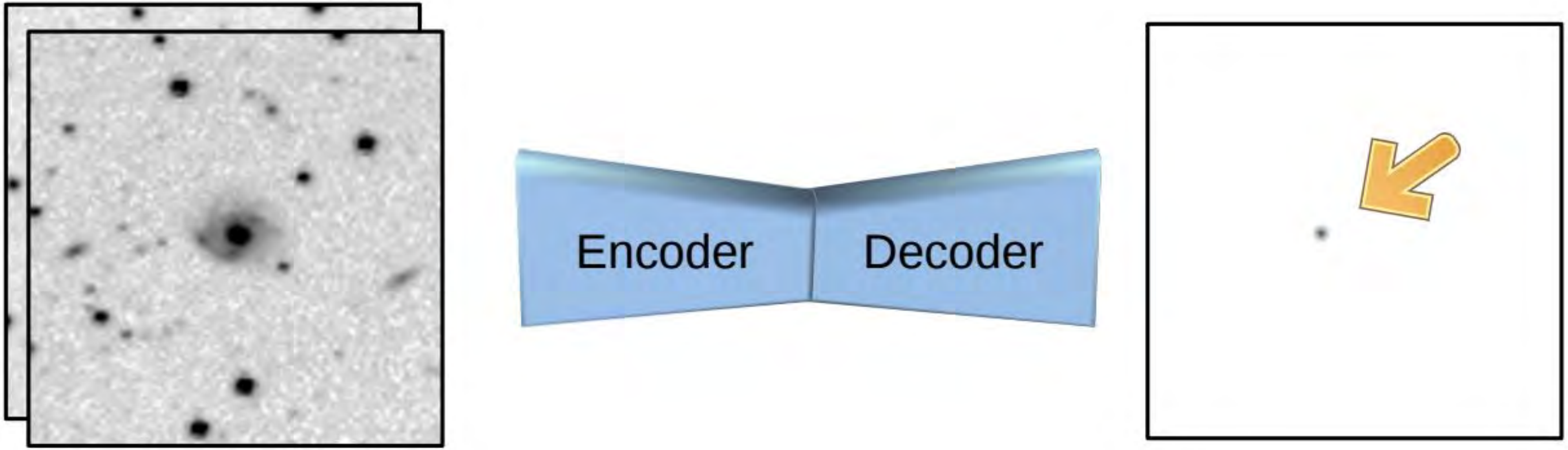
- Encoder compresses data into latent variables
- Decoder reconstructs data from latent variables
- Uses variational inference for training



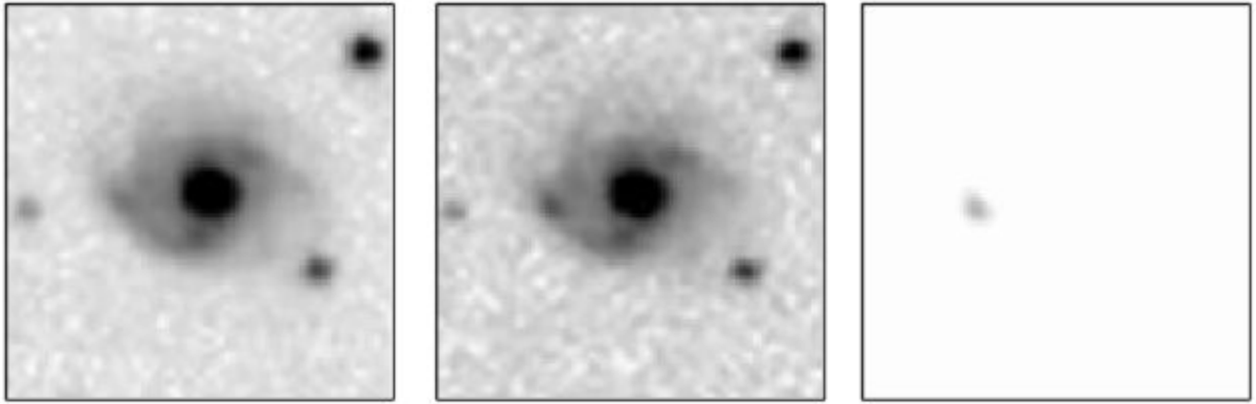
$x = d(e(x))$ ➜ **lossless encoding**
no information is lost
when reducing the
number of dimensions

$x \neq d(e(x))$ ➜ **lossy encoding**
some information is lost
when reducing the
number of dimensions and
can't be recovered later

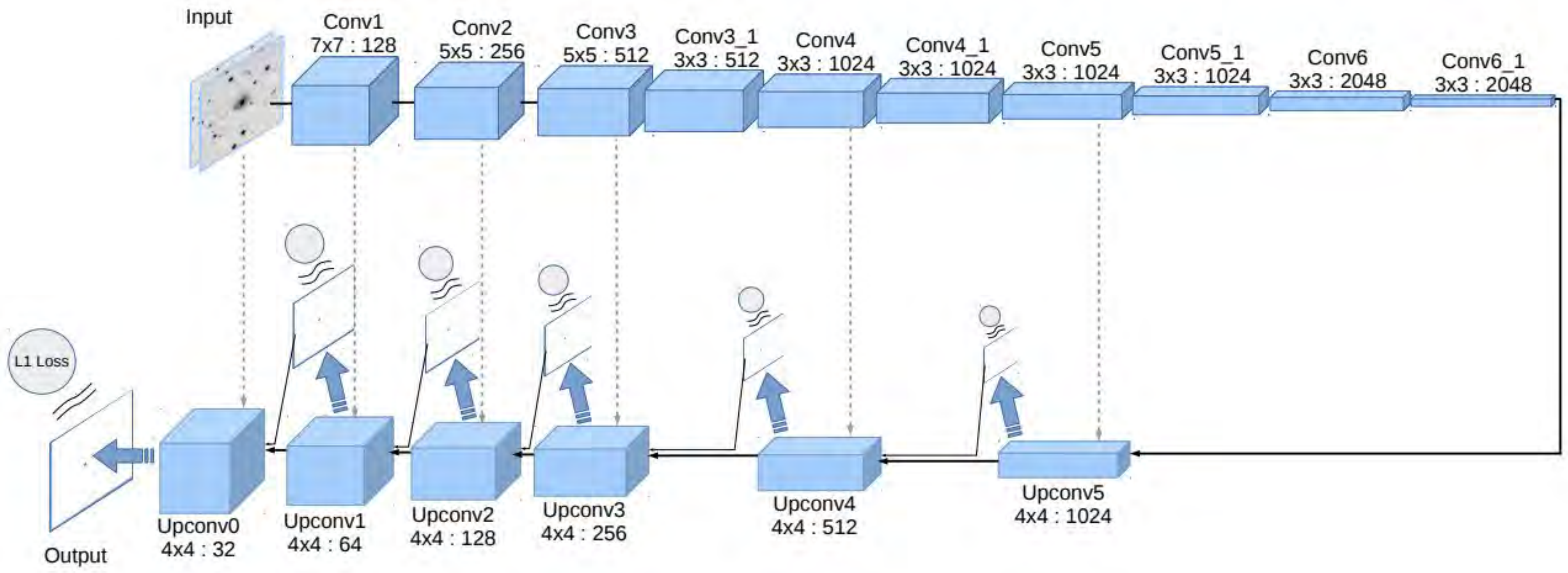
https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

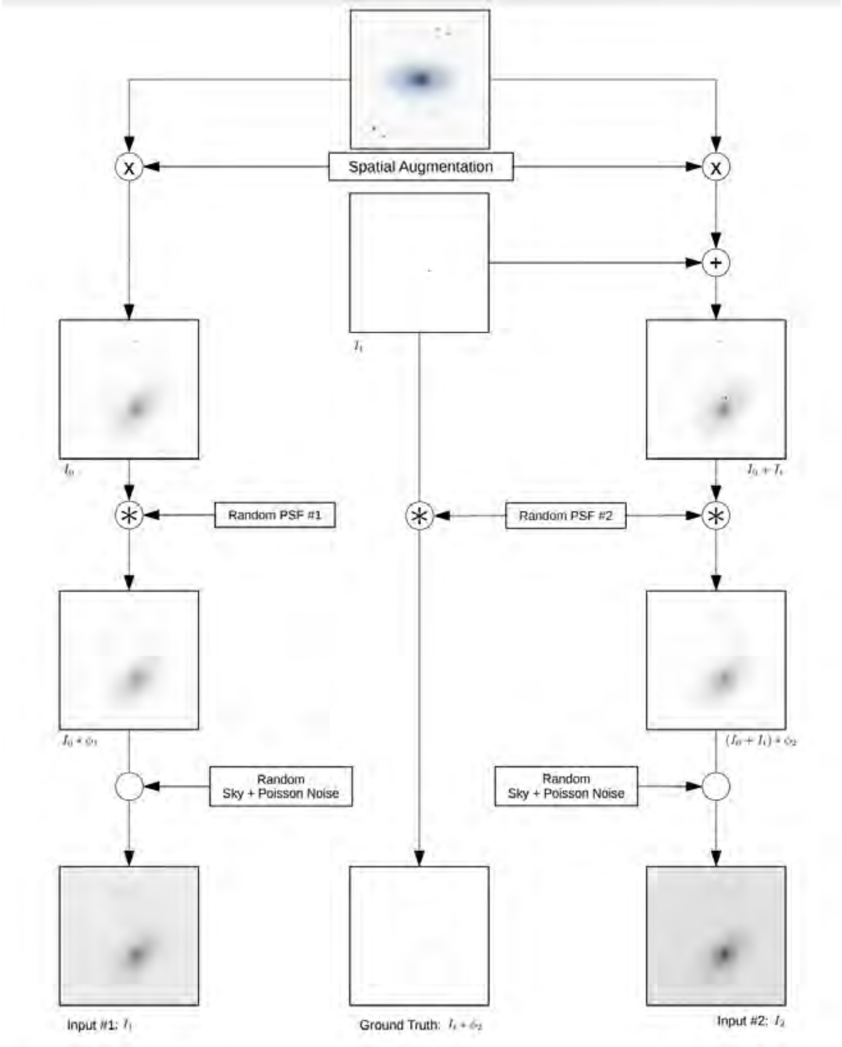


Sedaghat, Mahabal 1710.01422

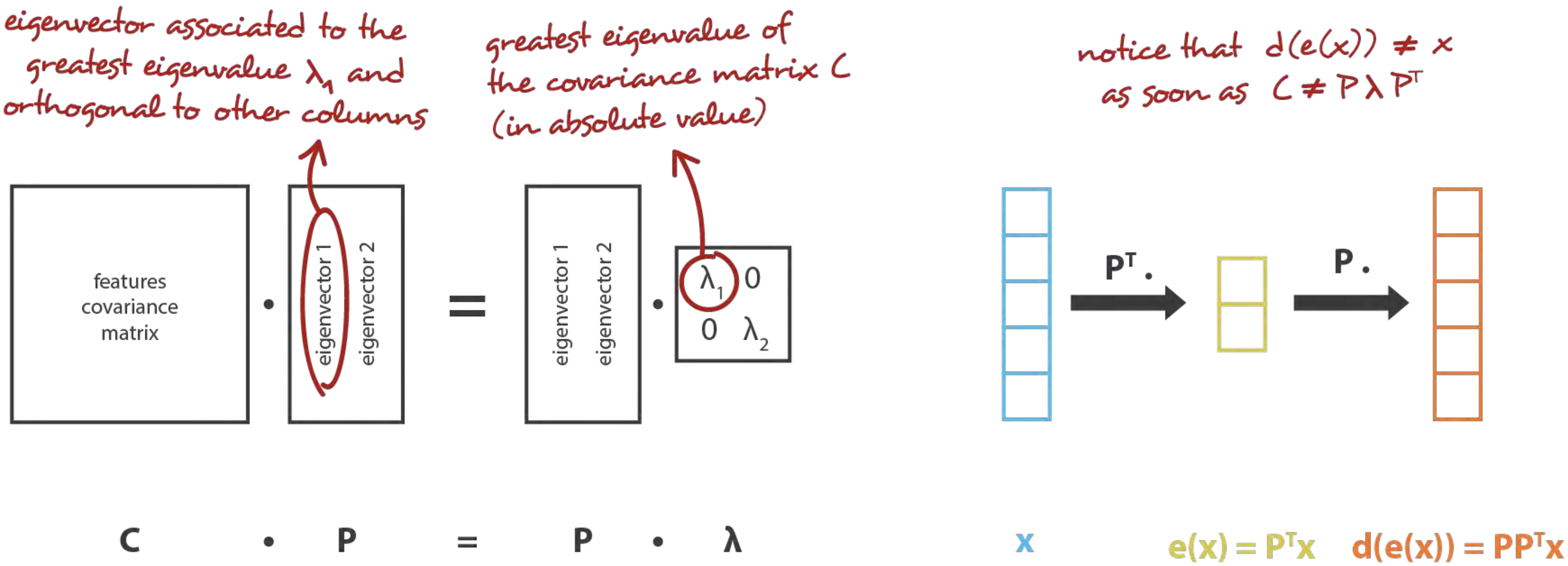


$$I1 = I0 * \varphi 1 + S1 + N1$$
$$I2 = (I0 + It) * \varphi 2 + S2 + N2$$

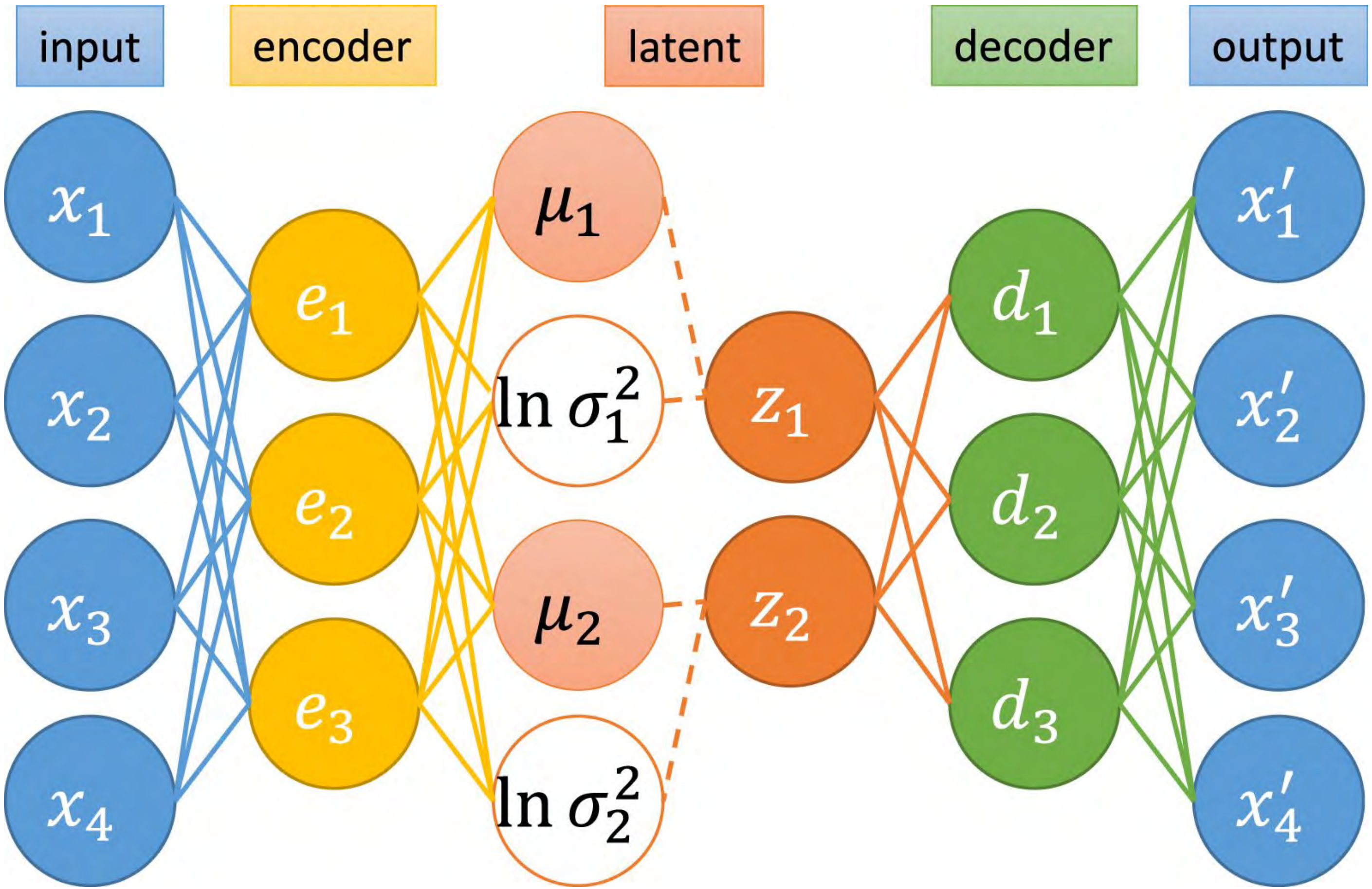

Input
Conv1
7x7 : 128
Conv2
5x5 : 256
Conv3
5x5 : 512
Conv3_1
3x3 : 512
Conv4
3x3 : 1024
Conv4_1
3x3 : 1024
Conv5
3x3 : 1024
Conv5_1
3x3 : 1024
Conv6
3x3 : 2048
Conv6_1
3x3 : 2048
L1 Loss
Output
Upconv0
4x4 : 32
Upconv1
4x4 : 64
Upconv2
4x4 : 128
Upconv3
4x4 : 256
Upconv4
4x4 : 512
Upconv5
4x4 : 1024



Has to incorporate physical conditions like the PSF



In principle anything could be used for dimensionality reduction (encoding). But neural networks are superior.



Portillo et al. 2020

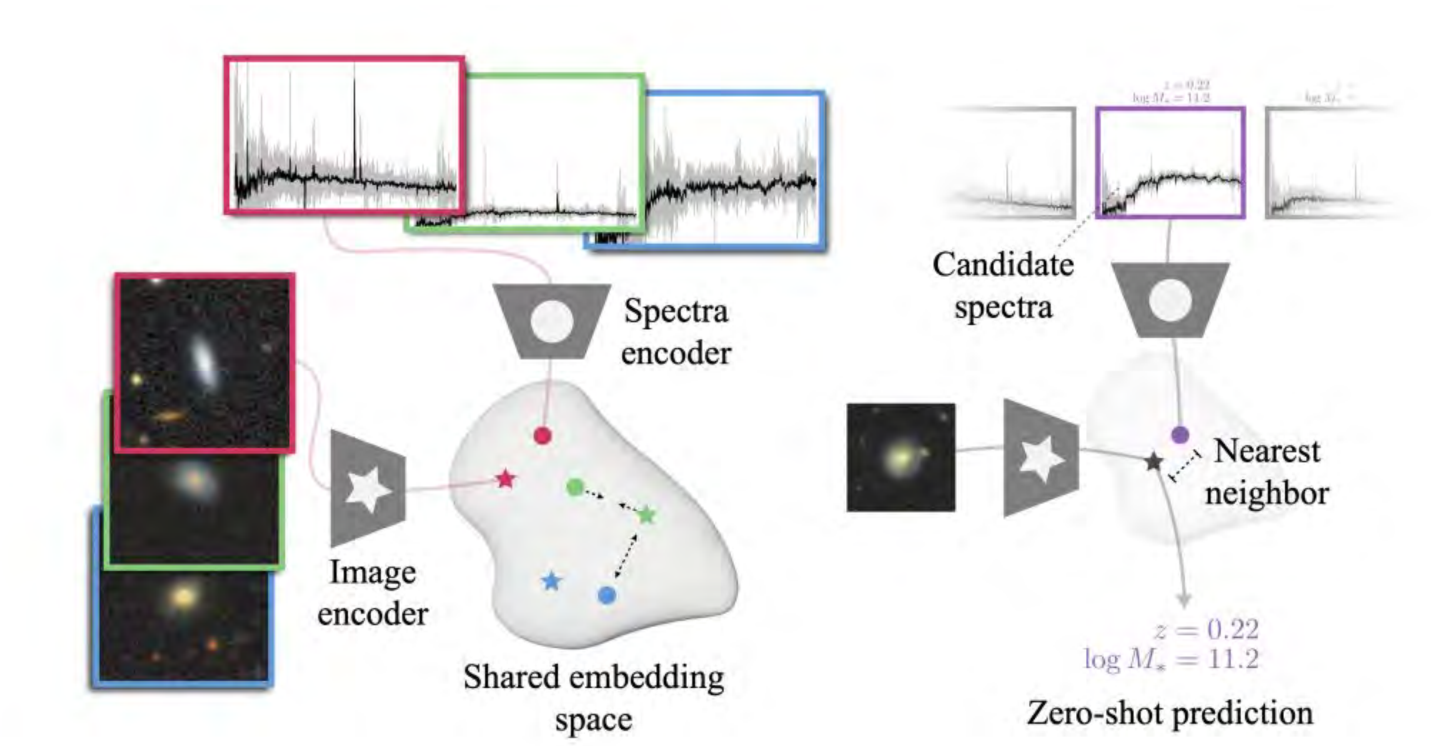
Application to SDSS spectra

Evidence lower bound (ELBO) is the objective function.
It is the sum of the reconstruction loss and the Kullback–Leibler (KL) divergence between the latent distribution for the input $q(z|x)$ and the prior $p(z)$

$$\mathrm{ELBO} = L(\boldsymbol{x}, \boldsymbol{x}') + D_{\mathrm{KL}}(q(\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{z})).$$

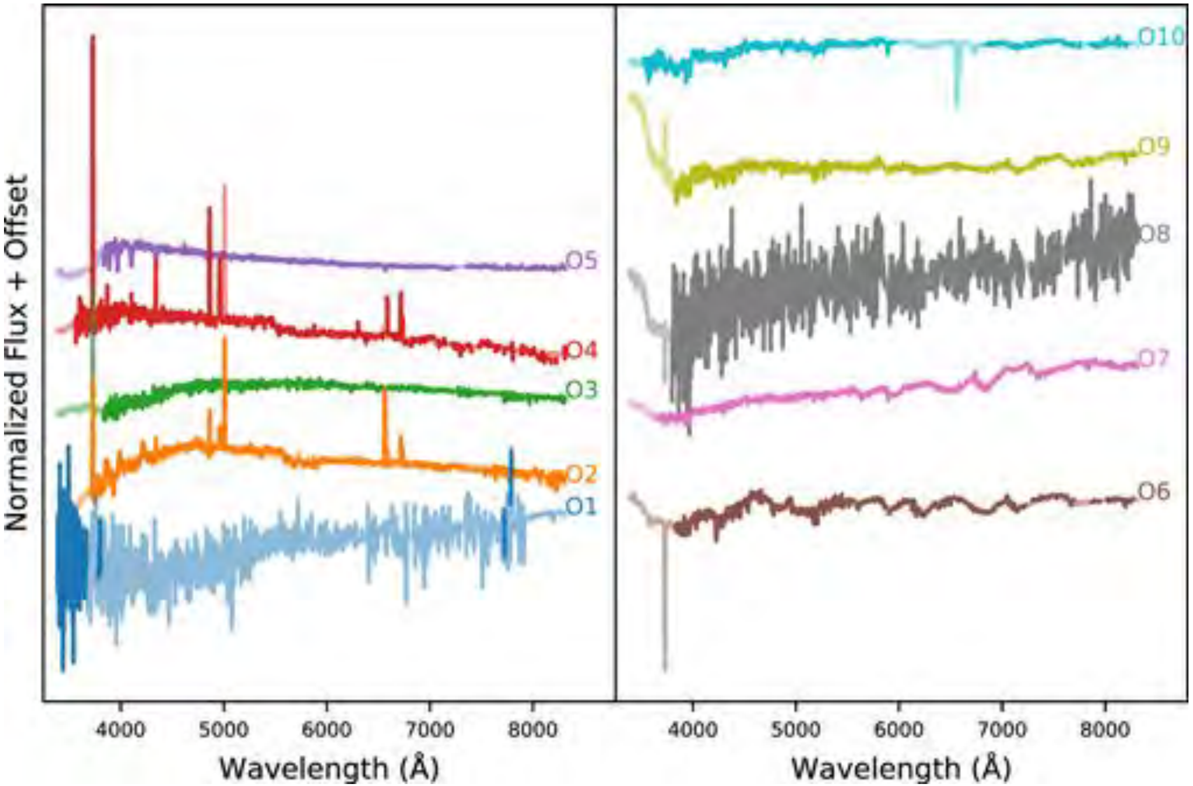
$$D_{\mathrm{KL}}(q||p) = \int q(z)\log\left(\frac{q(z)}{p(z)}\right)dz.$$

AstroCLIP (Lanusse et al. 2023) uses images and spectra of galaxies



Outlier detection

local outlier factor (LOF) algorithm (Breunig et al. 2000) is used to identify outliers. The algorithm estimates the local density of each point by using *k* nearest neighbors and then identifies points with densities much lower than their neighbors' as outliers.



Liang et al. 2023 find more interesting spectra in DESI using auto-encoders and normalizing flows

Combination of Probabilistic Modeling and Neural Networks

VAEs merge two powerful concepts: probabilistic graphical models and deep learning. This combination allows VAEs to leverage the strengths of both worlds:

- **Probabilistic Graphical Models**: These models handle uncertainty and variability in data by modeling probability distributions.
- **Deep Learning**: Neural networks, particularly deep architectures, excel at learning complex patterns and representations from high-dimensional data.

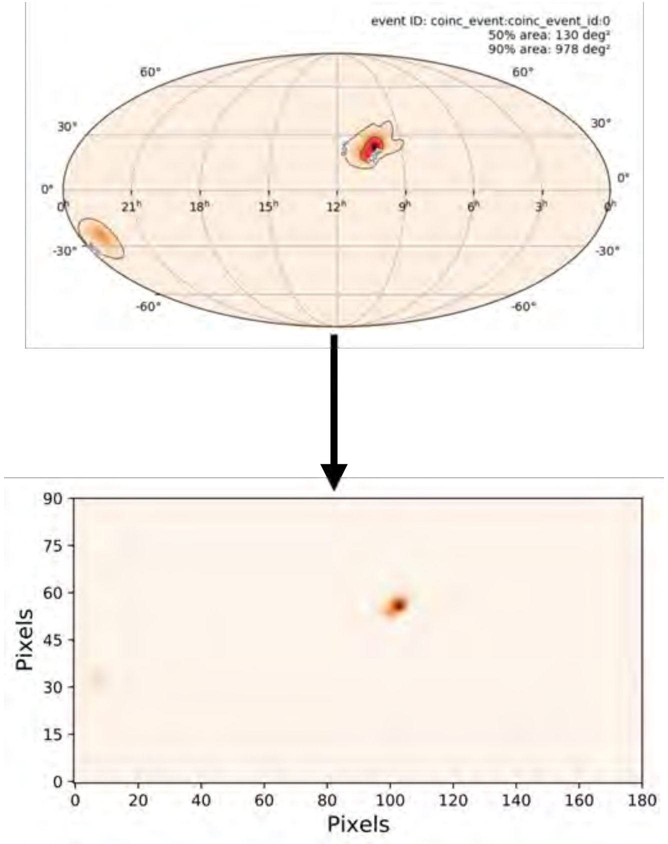
Scalability and Flexibility

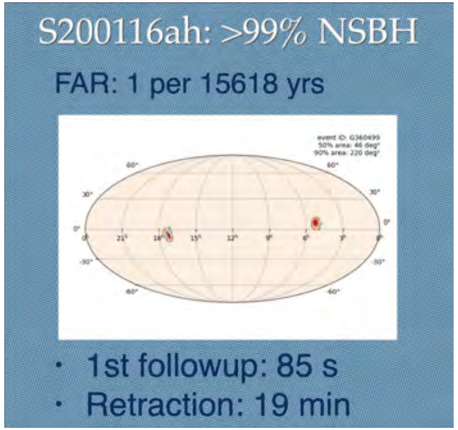
VAEs are scalable and flexible, making them applicable to various types of data:

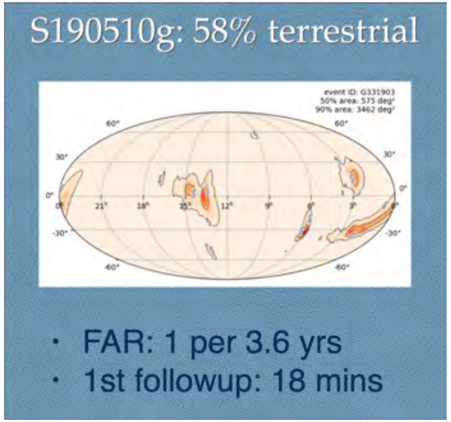
- **Different Data Types**: VAEs have been adapted to handle images, text, audio, and more.
- **Complex Architectures**: Extensions like Convolutional VAEs (for images) and Recurrent VAEs (for sequences) allow VAEs to handle complex, high-dimensional data efficiently.

GW SKyNet: a real-time classifier for public gravitational-wave candidates

arXiv:2010.11829 Cabero, Mahabal, McIver







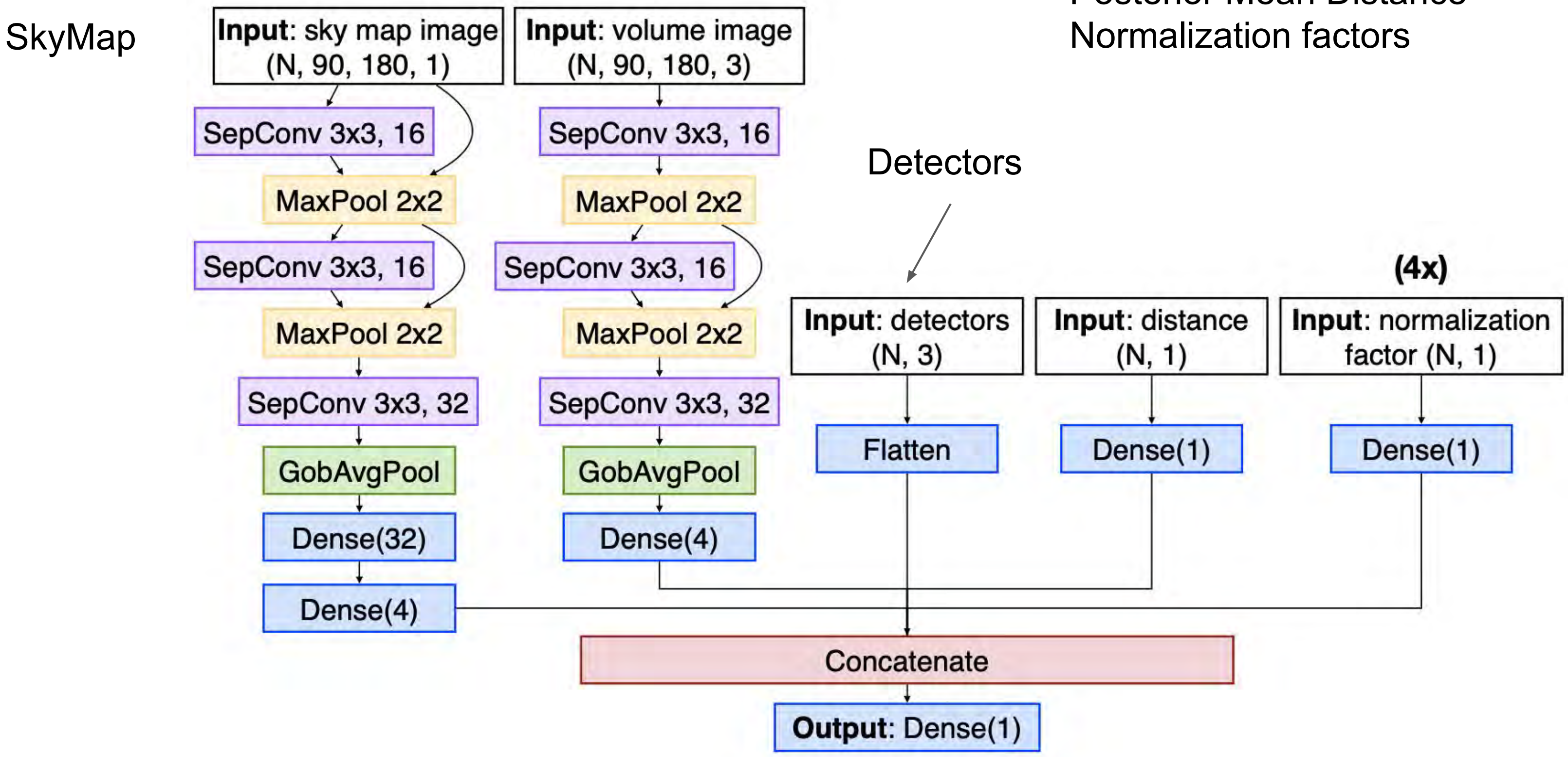
Retractions

non-astrophysical

Can we identify these based on just public data?

GW SkyNet

Sky map: 90 x 180
Stacked volume: 90 x 180 x 3
detectors 3-bits (multihot encoding)
Posterior Mean Distance
Normalization factors



GWTC-2 the O3a catalog

4 Published GW
13 new
22 confirmed
7 retracted

GWSkyNet verdict:
23 real
6 non-astrophysical

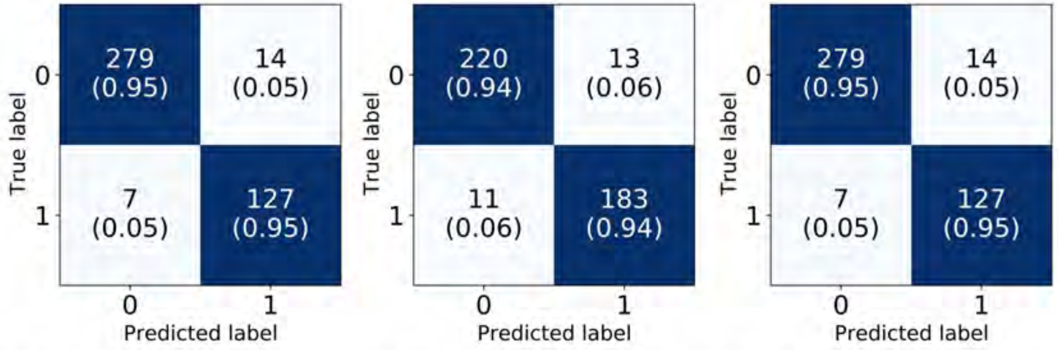
28/29 correct
no reals called bogus

23

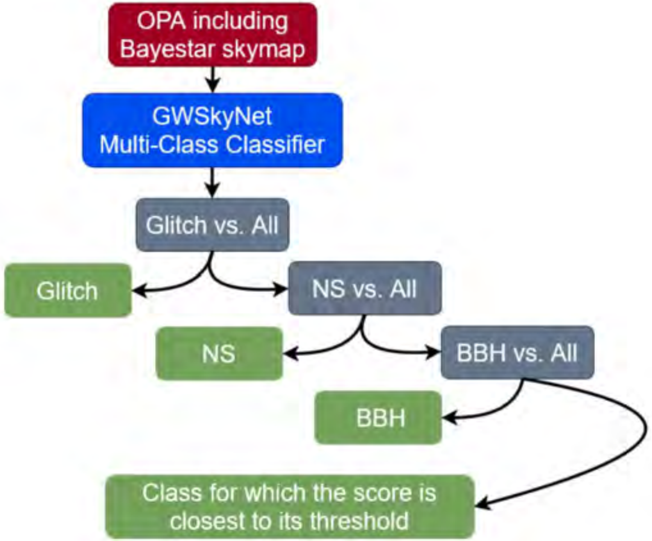
Name	Inst.	cWB		GstLAL			PyCBC			PyCBC BBH		
		FAR (yr^{-1})	SNR*	FAR (yr^{-1})	SNR	p_{astro}	FAR (yr^{-1})	SNR*	p_{astro}	FAR (yr^{-1})	SNR*	p_{astro}
GW190408_181802	HLV	$< 9.5 \times 10^{-4}$	14.8	$< 1.0 \times 10^{-5}$	14.7	1.00	$< 2.5 \times 10^{-5}$	13.5	1.00	$< 7.9 \times 10^{-5}$	13.6	1.00
GW190412	HLV	$< 9.5 \times 10^{-4}$	19.7	$< 1.0 \times 10^{-5}$	18.9	1.00	$< 3.1 \times 10^{-5}$	17.9	1.00	$< 7.9 \times 10^{-5}$	17.8	1.00
GW190413_052954	HLV	–	–	–	–	–	–	–	–	7.2×10^{-2}	8.6	0.98
GW190413_134308	HLV	–	–	3.8×10^{-1}	10.0	0.95	–	–	–	4.4×10^{-2}	9.0	0.98
GW190421_213856	HL	3.0×10^{-1}	9.3	7.7×10^{-4}	10.6	1.00	1.9×10^{0}	10.2	0.89	6.6×10^{-3}	10.2	1.00
GW190424_180648	L			$7.8 \times 10^{-1\dagger}$	10.0	0.91						
GW190425	LV			$7.5 \times 10^{-4\dagger}$	13.0	–						
GW190426_152155	HLV	–	–	1.4×10^{0}	10.1	–	–	–	–	–	–	–
GW190503_185404	HLV	1.8×10^{-3}	11.5	$< 1.0 \times 10^{-5}$	12.1	1.00	3.7×10^{-2}	12.2	1.00	$< 7.9 \times 10^{-5}$	12.2	1.00
GW190512_180714	HLV	8.8×10^{-1}	10.7	$< 1.0 \times 10^{-5}$	12.3	1.00	3.8×10^{-5}	12.2	1.00	$< 5.7 \times 10^{-5}$	12.2	1.00
GW190513_205428	HLV	–	–	$< 1.0 \times 10^{-5}$	12.3	1.00	3.7×10^{-4}	11.8	1.00	$< 5.7 \times 10^{-5}$	11.9	1.00
GW190514_065416	HL	–	–	–	–	–	–	–	–	5.3×10^{-1}	8.3	0.96
GW190517_055101	HLV	6.5×10^{-3}	10.7	9.6×10^{-4}	10.6	1.00	1.8×10^{-2}	10.4	1.00	$< 5.7 \times 10^{-5}$	10.2	1.00
GW190519_153544	HLV	3.1×10^{-4}	14.0	$< 1.0 \times 10^{-5}$	12.0	1.00	$< 1.8 \times 10^{-5}$	13.0	1.00	$< 5.7 \times 10^{-5}$	13.0	1.00
GW190521	HLV	2.0×10^{-4}	14.4	1.2×10^{-3}	14.7	1.00	1.1×10^{0}	12.6	0.93	–	–	–
GW190521_074359	HL	$< 1.0 \times 10^{-4}$	24.7	$< 1.0 \times 10^{-5}$	24.4	1.00	$< 1.8 \times 10^{-5}$	24.0	1.00	$< 5.7 \times 10^{-5}$	24.0	1.00
GW190527_092055	HL	–	–	6.2×10^{-2}	8.9	0.99	–	–	–	–	–	–
GW190602_175927	HLV	1.5×10^{-2}	11.1	1.1×10^{-5}	12.1	1.00	–	–	–	1.5×10^{-2}	11.4	1.00
GW190620_030421	LV			$2.9 \times 10^{-3\dagger}$	10.9	1.00						
GW190630_185205	LV			$< 1.0 \times 10^{-5}$	15.6	1.00						
GW190701_203306	HLV	5.5×10^{-1}	10.2	1.1×10^{-2}	11.6	1.00	–	–	–	–	–	–
GW190706_222641	HLV	$< 1.0 \times 10^{-3}$	12.7	$< 1.0 \times 10^{-5}$	12.3	1.00	6.7×10^{-5}	11.7	1.00	$< 4.6 \times 10^{-5}$	12.3	1.00
GW190707_093326	HL	–	–	$< 1.0 \times 10^{-5}$	13.0	1.00	$< 1.0 \times 10^{-5}$	12.8	1.00	$< 4.6 \times 10^{-5}$	12.8	1.00
GW190708_232457	LV			$2.8 \times 10^{-5\dagger}$	13.1	1.00						
GW190719_215514	HL	–	–	–	–	–	–	–	–	1.6×10^{0}	8.0	0.82
GW190720_000836	HLV			$< 1.0 \times 10^{-5}$	11.7	1.00	$< 2.0 \times 10^{-5}$	10.6	1.00	$< 3.7 \times 10^{-5}$	10.5	1.00
GW190727_060333	HLV	8.8×10^{-2}	11.4	$< 1.0 \times 10^{-5}$	12.3	1.00	3.5×10^{-3}	11.5	1.00	$< 3.7 \times 10^{-5}$	11.8	1.00
GW190728_064510	HLV	–	–	$< 1.0 \times 10^{-5}$	13.6	1.00	$< 1.6 \times 10^{-5}$	13.4	1.00	$< 3.7 \times 10^{-5}$	13.4	1.00
GW190731_140936	HL	–	–	2.1×10^{-1}	8.5	0.97	–	–	–	2.8×10^{-1}	8.2	0.96
GW190803_022701	HLV	–	–	3.2×10^{-2}	9.0	0.99	–	–	–	2.7×10^{-2}	8.6	0.99
GW190814	LV			$< 1.0 \times 10^{-5}$	22.2	1.00						
GW190828_063405	HLV	$< 9.6 \times 10^{-4}$	16.6	$< 1.0 \times 10^{-5}$	16.0	1.00	$< 1.5 \times 10^{-5}$	15.3	1.00	$< 3.3 \times 10^{-5}$	15.3	1.00
GW190828_065509	HLV	–	–	$< 1.0 \times 10^{-5}$	11.1	1.00	5.8×10^{-5}	10.8	1.00	$< 3.3 \times 10^{-5}$	10.8	1.00
GW190909_114149	HL	–	–	1.1×10^{0}	8.5	0.89	–	–	–	–	–	–
GW190910_112807	LV			$1.9 \times 10^{-5\dagger}$	13.4	1.00						
GW190915_235702	HLV	$< 1.0 \times 10^{-3}$	12.3	$< 1.0 \times 10^{-5}$	13.1	1.00	8.6×10^{-4}	13.0	1.00	$< 3.3 \times 10^{-5}$	12.7	1.00
GW190924_021846	HLV	–	–	$< 1.0 \times 10^{-5}$	13.2	1.00	$< 6.3 \times 10^{-5}$	12.5	1.00	$< 3.3 \times 10^{-5}$	12.4	1.00
GW190929_012149	HLV	–	–	2.0×10^{-2}	9.9	1.00	–	–	–	–	–	–
GW190930_133541	HL	–	–	5.8×10^{-1}	10.0	0.92	3.4×10^{-2}	9.7	1.00	3.3×10^{-2}	9.8	0.99

TABLE IV. Gravitational wave candidate event list. We find 39 candidate events passing the FAR threshold of 2.0 yr^{-1} in at

Excellent performance



(a) Glitch-versus-all confusion matrix. (b) NS-versus-all confusion matrix. (c) BBH-versus-all confusion matrix.



Candidate Name	GraceDB Label	GWSkyNet Binary	Glitch Score (%)	NS Score (%)	BBH Score (%)	Hierarchical Prediction
S191105e	BBH	Glitch	0.1	0	98.7	BBH
S191109d	BBH	Real	15.6	0	100	BBH
S191110x	RETRACTED	Real	90.2	86.6	17.9	Glitch[†]
S191117j	RETRACTED	Real	100	70.2	6.5	Glitch[†]
S191120aj	RETRACTED	Glitch	92.8	34	0.1	Glitch
S191120at	RETRACTED	Real	99.4	98.7	27	Glitch[†]
S191124be	RETRACTED	Glitch	95.1	91.2	5.6	Glitch[†]
S191129u	BBH	Real	0.8	0.5	77.1	BBH
S191204r	BBH	Real	0.1	3.1	84.1	BBH
S191205ah	NSBH	Glitch	50.4	97	0.1	Glitch[†]
S191212q	RETRACTED	Real	49.8	100	1.9	Glitch[†]
S191213ai	RETRACTED	Glitch	98.3	92.1	0	Glitch[†]
S191213g	BNS	Real	3	2.4	2.7	BBH*
S191215w	BBH	Real	0.1	0	100	BBH
S191216ap	BBH	Real	1.7	83.4	43.2	NS[†]

GW SKyNet-Multi

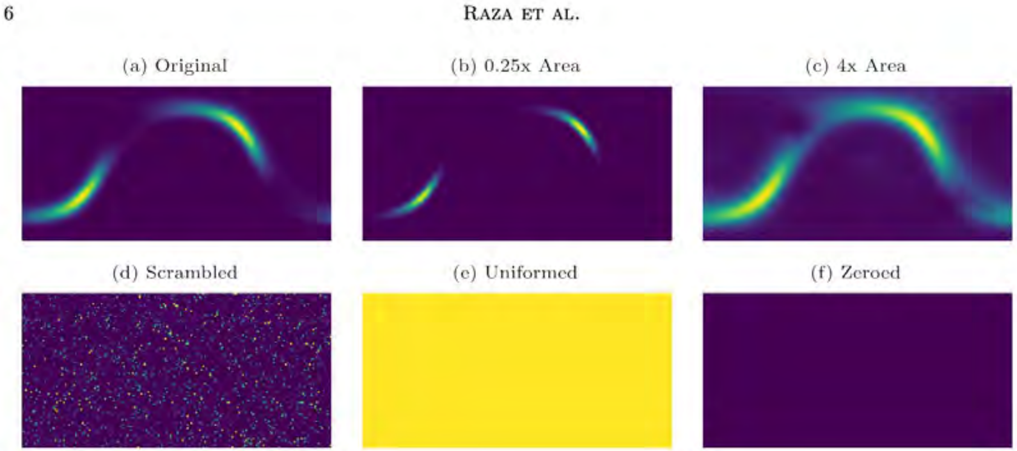
Abbott et al. 2011.04015

Does great! But WHY??

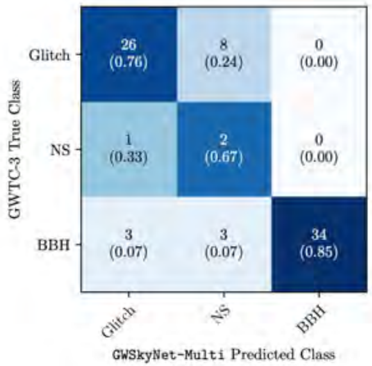
Explaining the GWSkyNet-Multi machine learning classifier predictions for gravitational-wave events

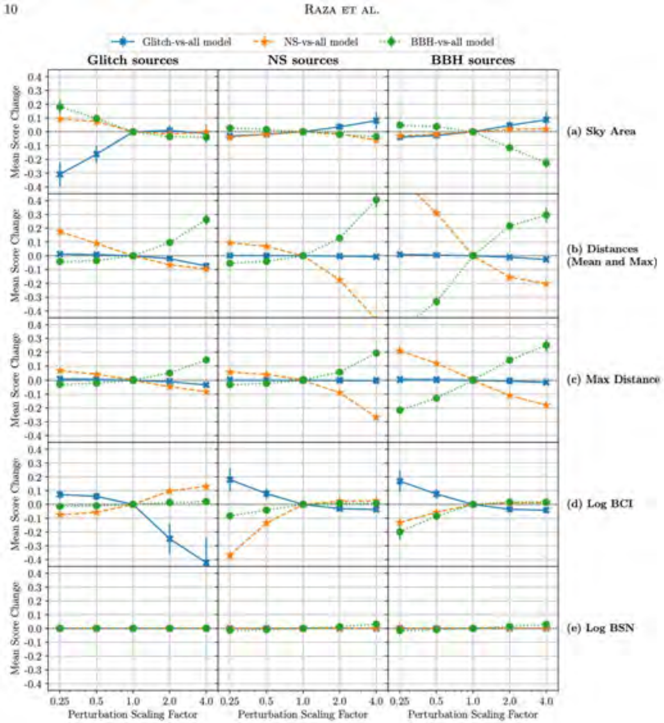
Nayyer Raza[1,2], Man Leong Chan[3], Daryl Haggard[1,2], Ashish Mahabal[4,5], Jess McIver[3], Thomas C. Abbott[1,2], Eitan Buffaz[1,2] and Nicholas Vieira[1,2]

Raza et al. 2308.12357

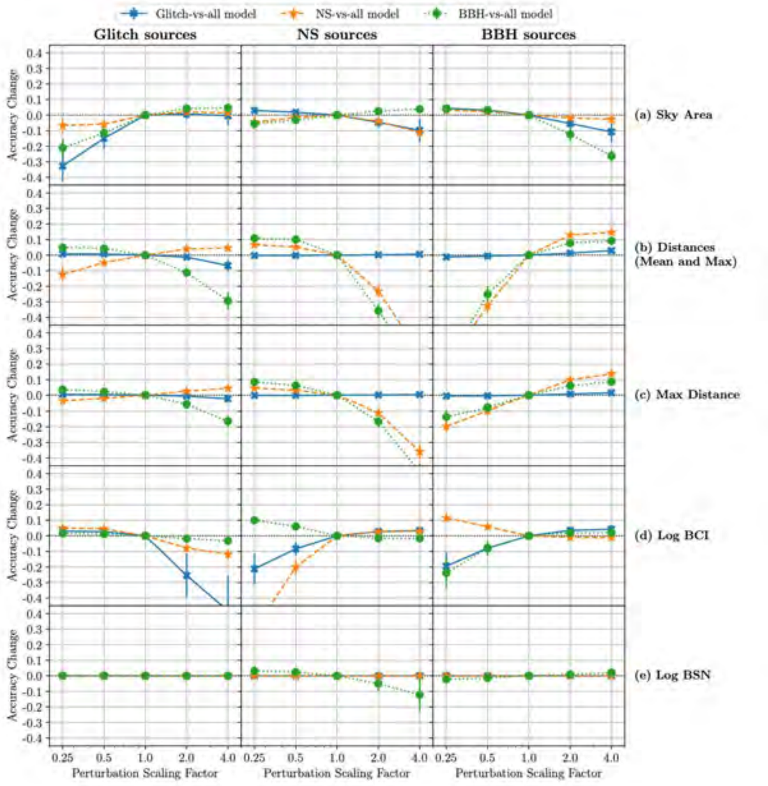


O3 events





Perturbations



Strong glitch predictors:

- Localization area of the 2D sky maps
- Computed coherence versus incoherence Bayes factors

Strong real subclassifier:

- Estimated distance to the source

Note used:

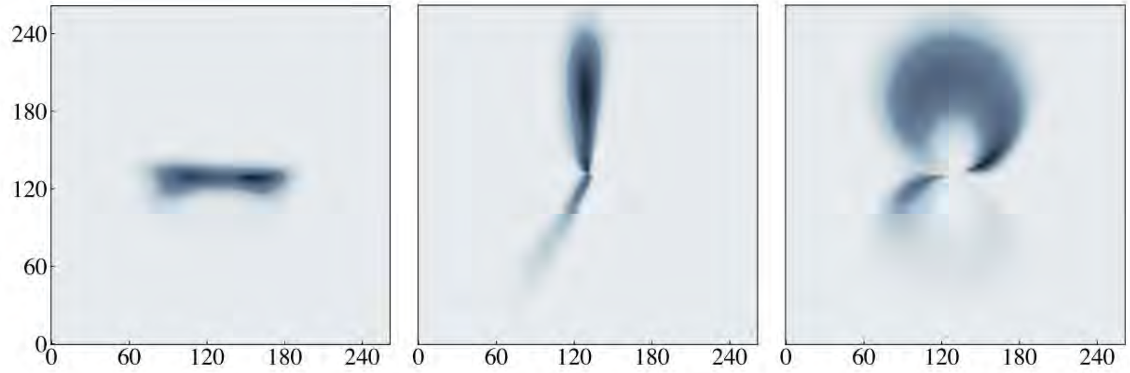
- Signal versus noise Bayes factors

Helps us understand our models better

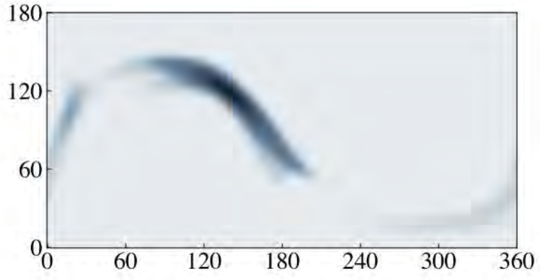
Better distributions needed for training!

Refined GWSkyNet part of LVK's low latency pipeline

Chan et al. 2408.06491

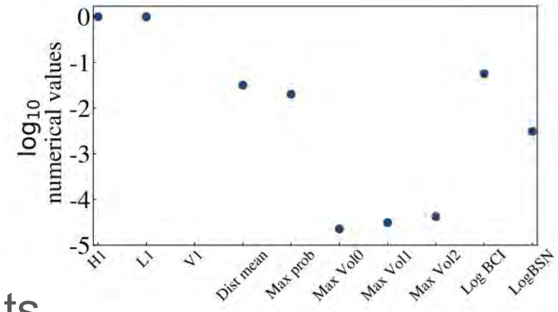


(a) Vol0 (b) Vol1 (c) Vol2

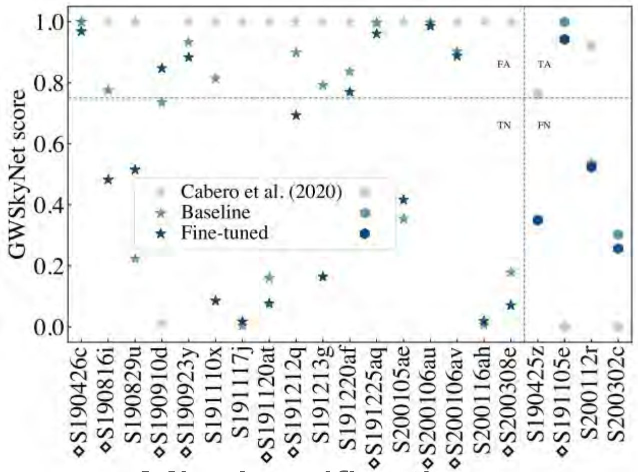


(d) Sky localization

Inputs



(e) Numerical data



Misclassifications

The Fine-tuned model rejects >80% noise while capturing >93% astrophysical signals

Other

Neutrino

Cosmology

(graph neural networks)

(normalizing flows)

Interpretability and explainability

Linardatos, Papastefanopoulos, Kotsiantis 2020

Post-hoc

LIME: Local Interpretable Model-agnostic Explanations

SHAP: Shapley Additive Explanations

Fairness