Real-time ML-powered transient discovery with GOTO and Kilonova Seekers

Tom Killestein

Prize Fellow

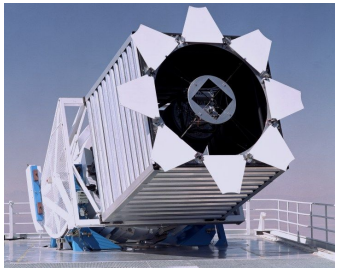
University of Warwick

Astronomy seminar, Southampton (2025-11-25)

UNIVERSITY OF WARWICK

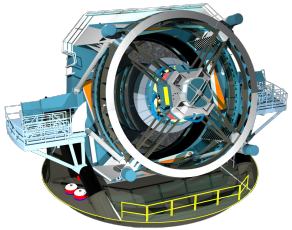
The deluge of data

> Astronomy is now among the most data-heavy sciences.

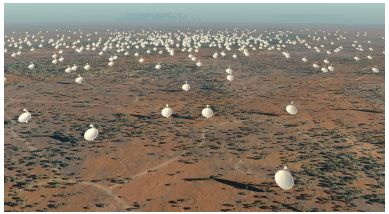


SDSS (1990s)
200GB / night

GOTO (now)
1TB / night



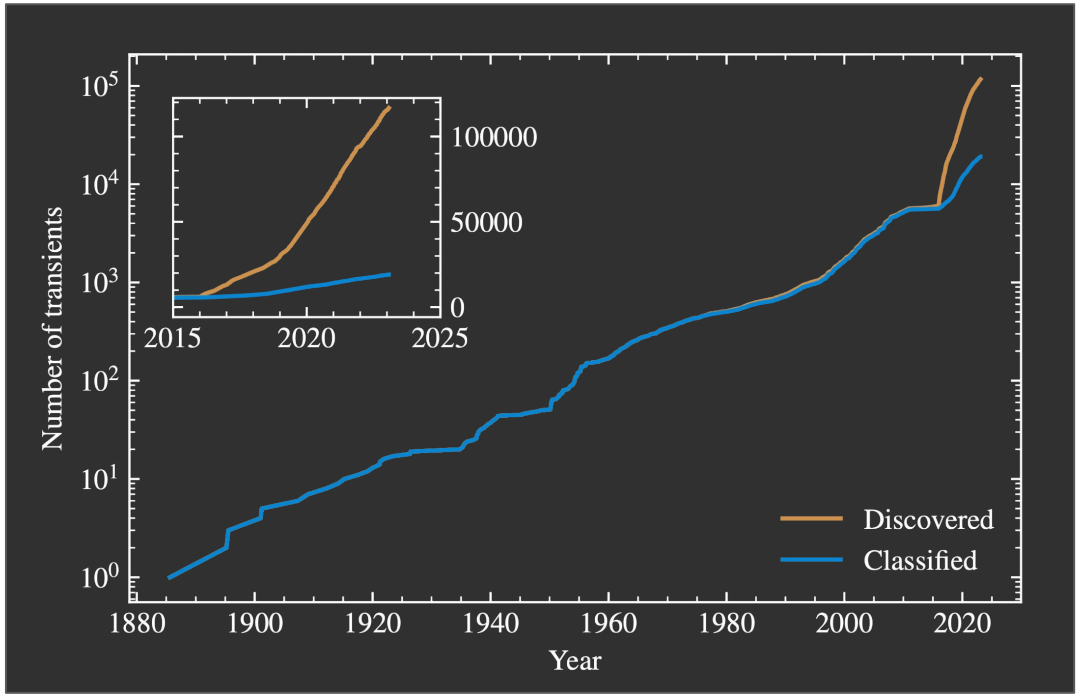
Rubin Observatory
(2024-)
20TB / night



SKA (2030s)
160TB / second

> All of the above are reliant of machine learning to manage the data volumes involved!

Growing survey capacity: growing discovery



Data from IAU Transient Name Server

Massive step-changes in gradient from new surveys coming online.

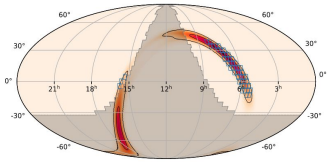
176901 transient discoveries as of this morning

Rapidly-widening gap between discovery and classification

The Gravitational-wave Optical Transient Observer

Each telescope/node:
8x40cm astrographs, combined FoV of **40 square degrees** - reaches **L ~ 20.5 in 180s** enabling rapid and deep surveys of GW localisation regions in real-time.

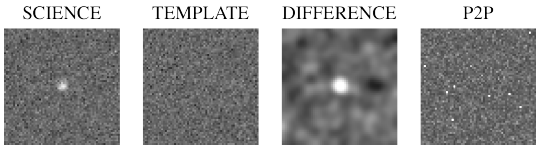
Real-time response to GW/GRB/neutrino alerts, 2-3 day cadence over whole sky.

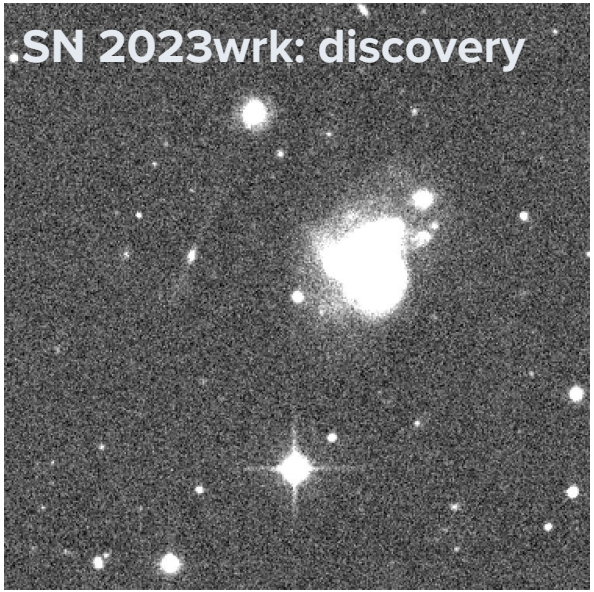


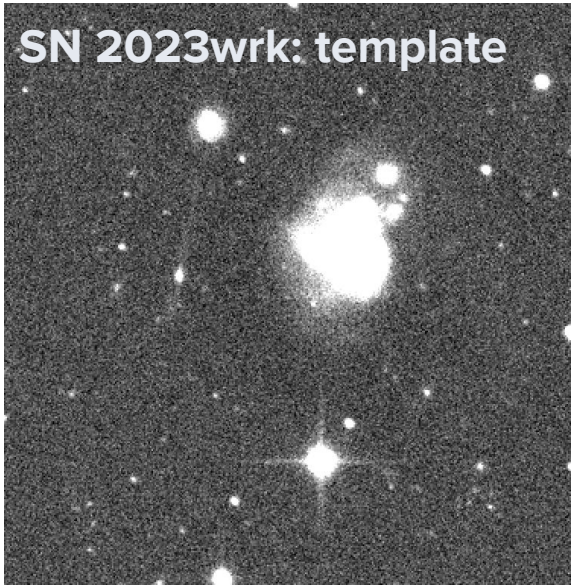
Rapidly fading (~hours) GW counterparts necessitates **real-time, accurate** source classification **-** only possible with high-performance ML/DL.

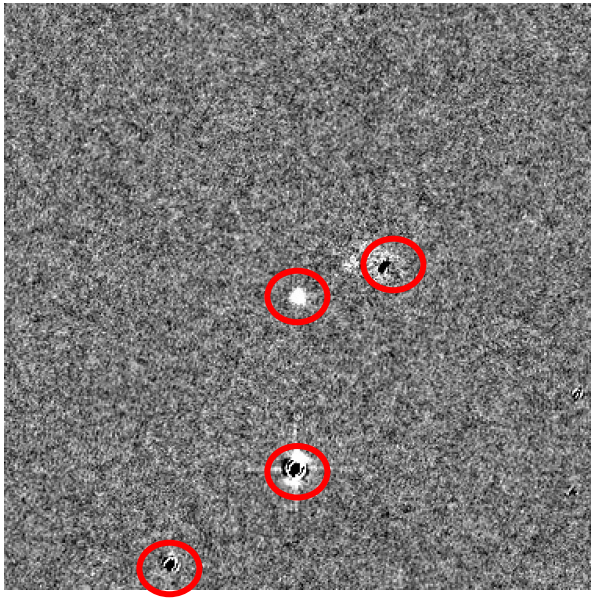
The significant data volume generated by GOTO (and other surveys) provides an ideal **data-rich proving ground** for novel ML/DL algorithms applied to a diverse range of data modalities.

Difference imaging









Many algorithms, none perfect: have to match complex PSF + background + photometric normalisation between two images. Intrinsically high false-positive rate, and many defects in process.

Building source classifiers for transient discovery

Time-domain astronomy: the real-bogus problem

~TBs / night

Image subtraction

~2M / 24h

Real-bogus filtering

~2,000s / 24h

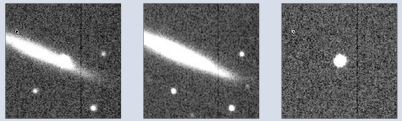
Human inspection

~10s / 24h

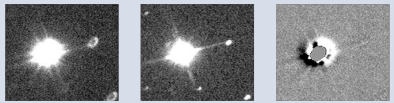
Real-bogus classification

Use ML/DL to automatically identify artifacts in difference images based on extracted properties.

Maximise completeness whilst simultaneously minimising false positives - avoid overwhelming human vetters!



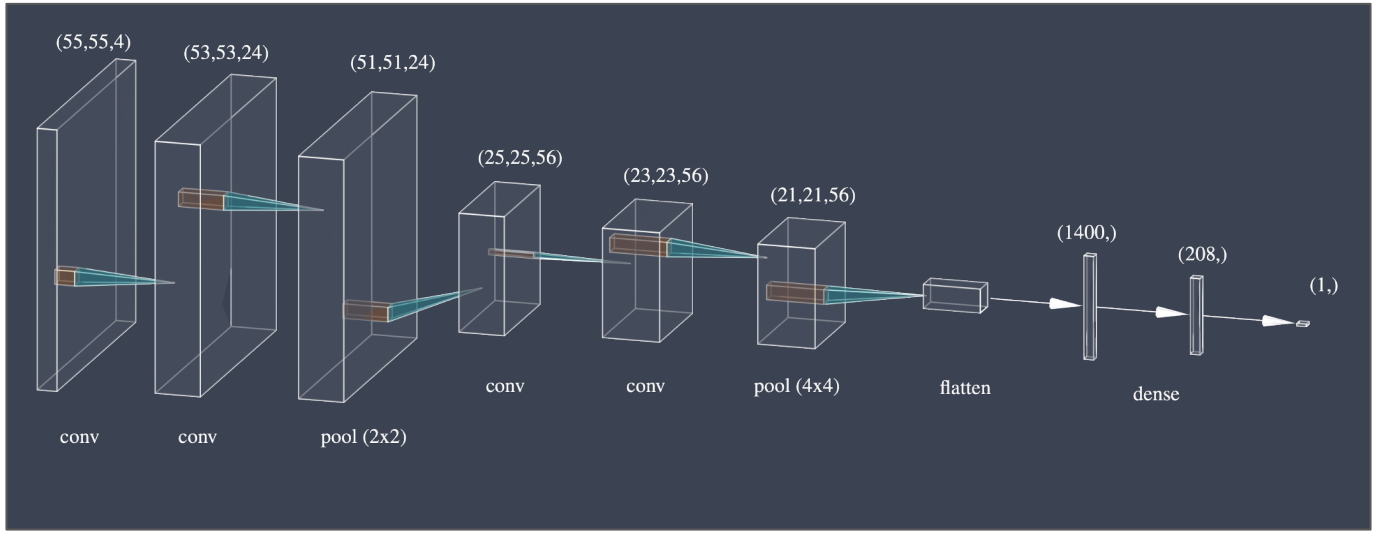
SN2021hiz (real, score 1)



Bad subtraction (bogus, score 0)

see e.g. Bloom+12, Duev+19, **Killestein+21**

A commonality: simplistic architectures do the job



VGG16-likes

Duev+19
Killestein+21
Makhlouf+22
Takahashi+22

Either (or both!) astronomical images are comparatively simple, or RB classification is not a fundamentally taxing task.

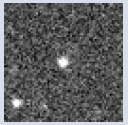
Automated training set

Real:
Minor planets - yield PSF-like detections

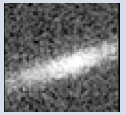
Bogus:
Randomly-sampled difference image-detections, filtered for purity.

Human labelling optional, can be useful for artifacts

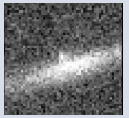
Transient augmentation



Randomly select minor planets in each image



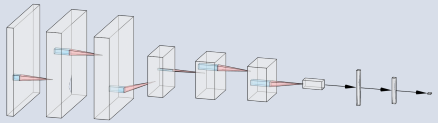
Identify a galaxy near to the MP and extract an offset stamp



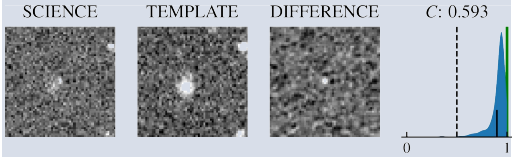
Sum galaxy stamp and minor planet to get a synthetic transient

Realistic PSFs, fast, and configurable injection.

Bayesian CNN



Scores are given as posterior distributions
Classifier uncertainty!



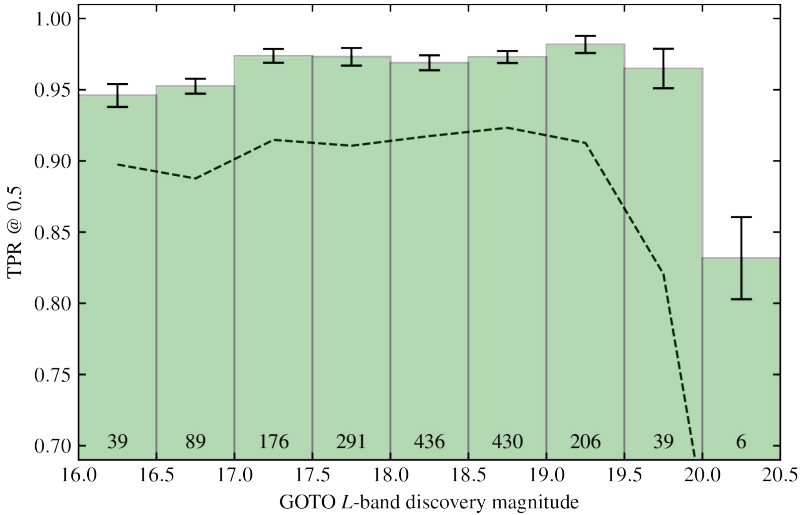
99.1% accuracy
0.5% FNR at 1% FPR

Highly configurable, image-level parallelism allows generating a **400k example dataset in under 24h!**

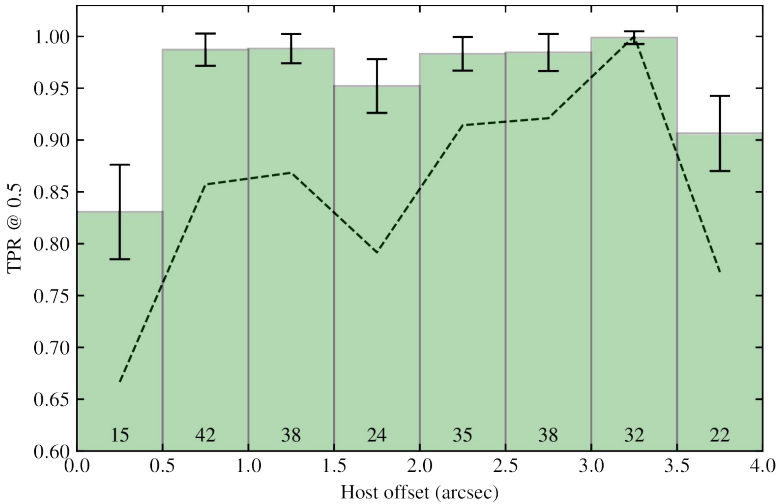
For full implementation details: **Killestein+ 2021** [arXiv: 2102.09892]
All code is publicly available via GitHub - *GOTO-OBS/gotorb*

Transient optimisation: marked performance improvements

Test the classifier on >900 spectroscopically-confirmed transients recovered in the GOTO prototype phase — performance likely to be even better with in-progress instrumentation and pipeline upgrades.



Significant increase in recovery of faint transients - excellent prospects for kilonova recoveries in local Universe.

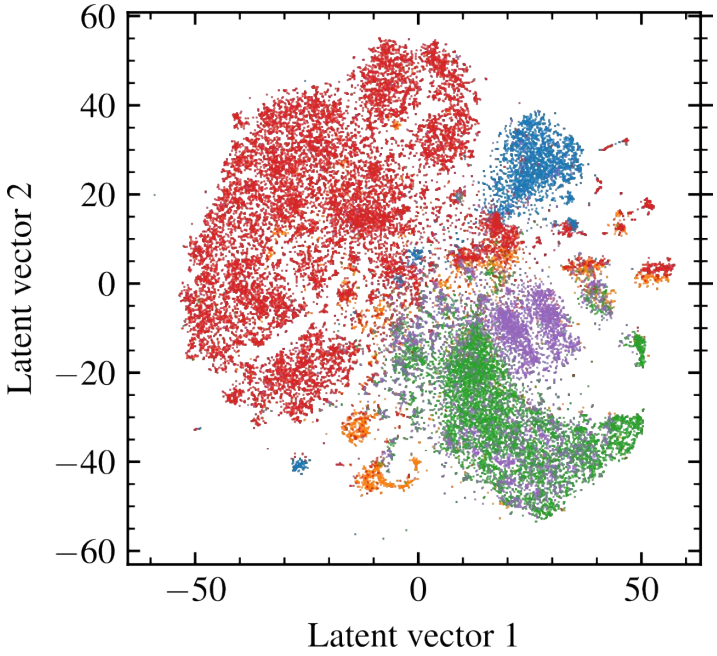


Uniform coverage of host-offset space leads to **boosted recovery of nuclear transients**

Learning rich representations of discovery triplets

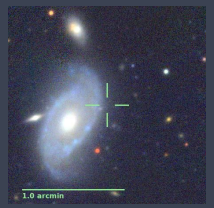
Even without more granular class information than real-bogus, the model intrinsically learns to cluster similar objects in the dense latent space.

Such classifiers could easily be fine-tuned to do multi-class work - if we had the labels!



Contextual classification

Image-only:



+ Always available (discovery image)
- Limited by image quality/resolution
- Apparent type != actual type

Catalog-only:

```
ra,dec,gmag,...
32.042,12.321,19.042,...
31.987,12.021,18.428,...
32.321,12.124,21.421,...
```

Salient information already extracted

+ Far deeper than survey images
- Incomplete (galaxy catalogs)
- Incorrect (misclassifications)
- Inhomogeneous (variable coverage)

How to optimally combine these to get the benefits of both in a ‘sensible’ way?

moriarty: a data-driven contextual classifier

- A dynamic compilation of over 12 large variable star/AGN/galaxy catalogues, backed with de-duplication and voting algorithms to determine the best source properties
- PostgreSQL database with Python API for querying, generating human-readable descriptions, and decision trees for classification.

*named as a nod to Sherlock (Young+2023)

Query times typically 500ms - DB makes heavy use of spatial indexing via **q3c** (Koposov and Bartunov 2006)

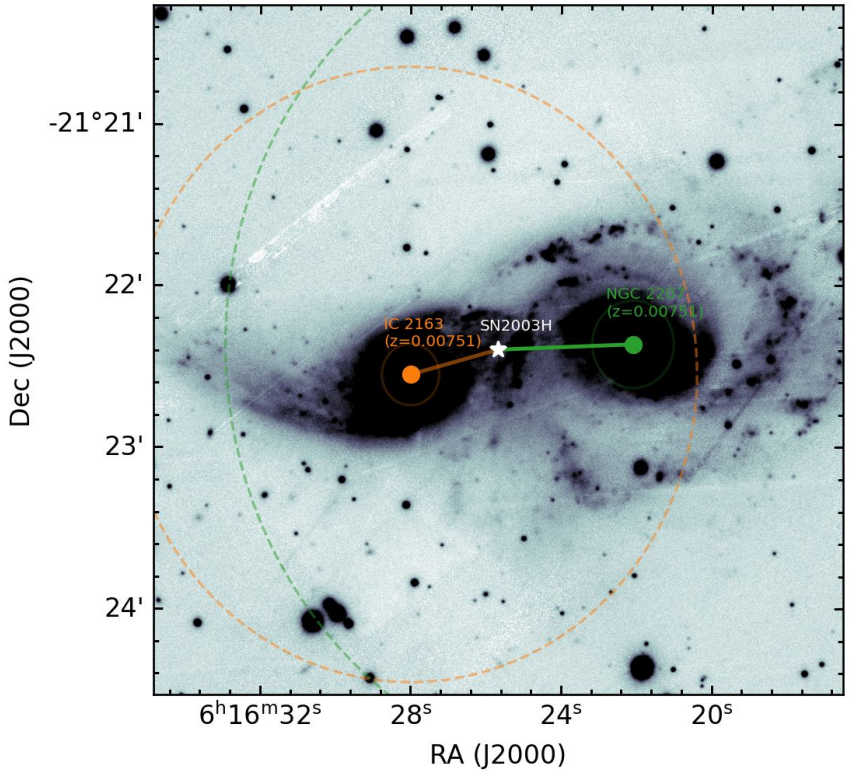
Maximum-likelihood host matching

Cone searches not ideal: sky density of bright galaxies intrinsically higher, so higher false match rate for faint hosts.

Modify the p_{chance} method of Bloom (2006) to work with heterogeneous survey filters (synthetic colours)

$$p_{assoc} = \exp\left(-\pi\theta^2\rho(m)\right)\left(1 - p_{star}\right)$$

Also use probabilistic matches for i.e. source de-duplication and point source associations



Performance and deployment

Finds correct host for 90% of ZTF-BTS (Perley+2021) transients, and significantly outperforms cone search on nearby/gap transients.

Typical query times of <1s - aim to release public API for others to use for host identification

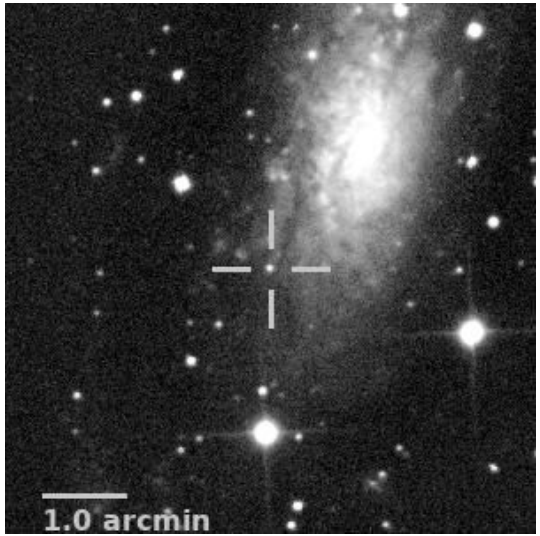
Professor James Moriarty (2 weeks, 3 days ago) GOTO123632.47+111319.71 is likely associated with the B=11.32 mag galaxy NGC 4568 in the gladeplus_galaxies, ps1_stackobjectview_minimal catalogs (65.12" away). Probability of connection 98.829% -- host is at 18.4 Mpc (z=0.0041+/-0.0001), implying a transient absolute magnitude of -15.26 and sky-projected offset of 5.81 kpc. This transient is spatially coincident with the cluster Virgo cluster (20.0 Mpc)

moriarty v2.0 (in prep)

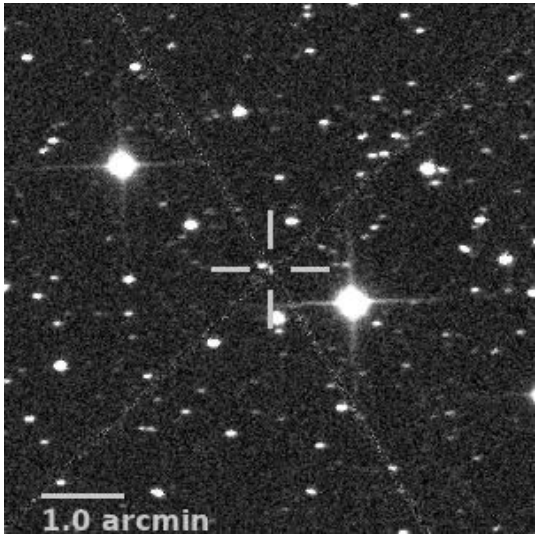
- Full re-write for **Rubin-scale** catalogs
- Updated host association algorithms - moving away from P(chance) to better maximum likelihood approach that neatly handles redshift uncertainties and similar.
- Embeddings for deep learning - combining images and context for higher performance
- Public API to enable brokers to access outputs

But: classifiers far from perfect (or enough)

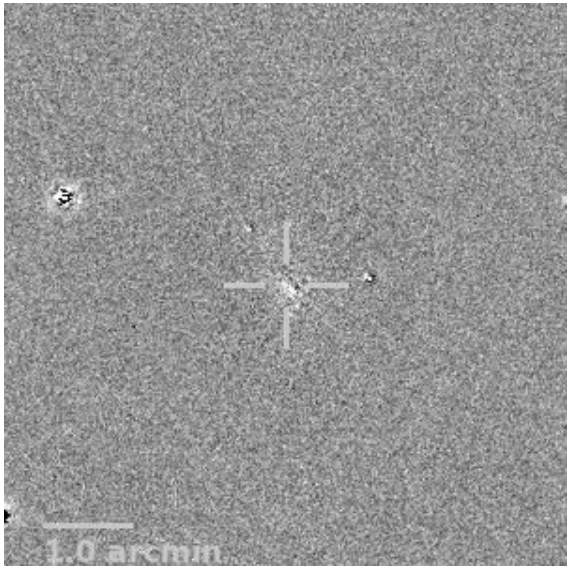
Continued development and fresh labels needed to stay performant



SN2024ggi - missed ingestion by 0.04



Anomalies - ever weirder artefacts unseen in training set

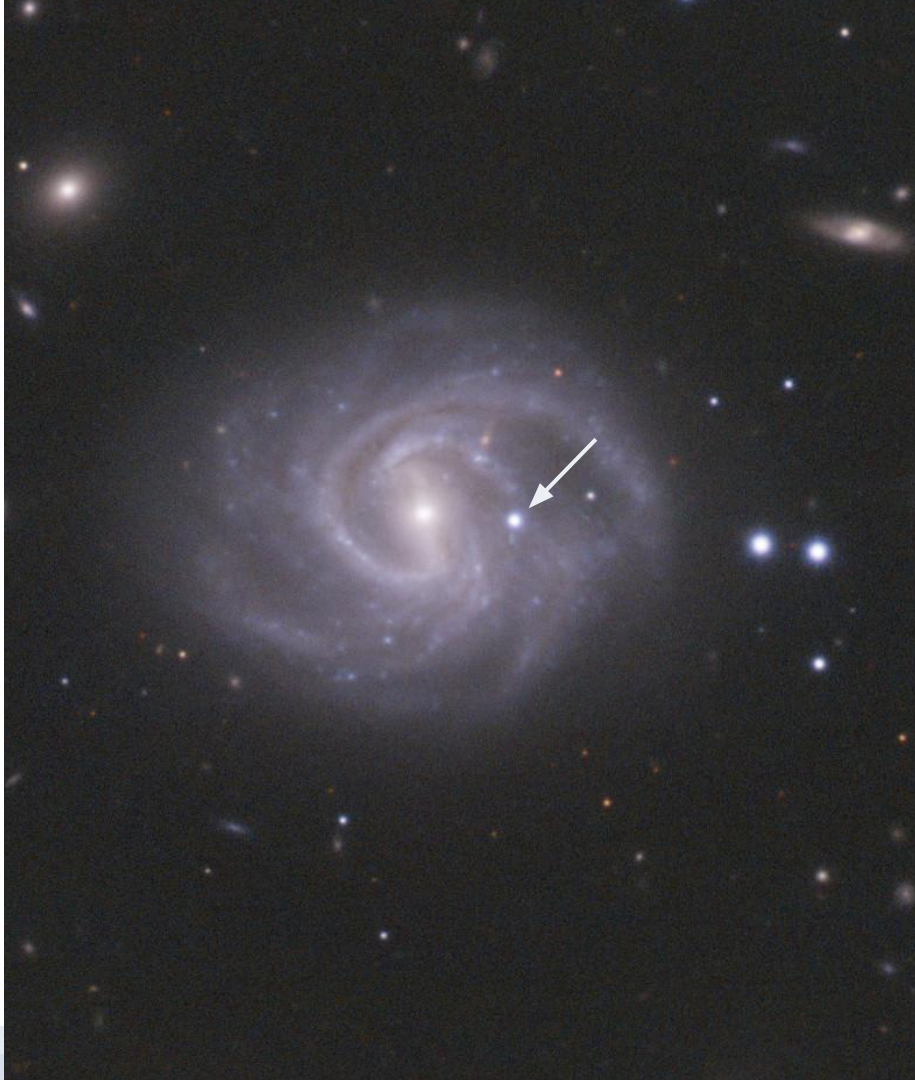


Data drift - changing properties of survey instrument

SN 2024cld

Unveiling the complex mass-loss history of an exotic core-collapse supernova

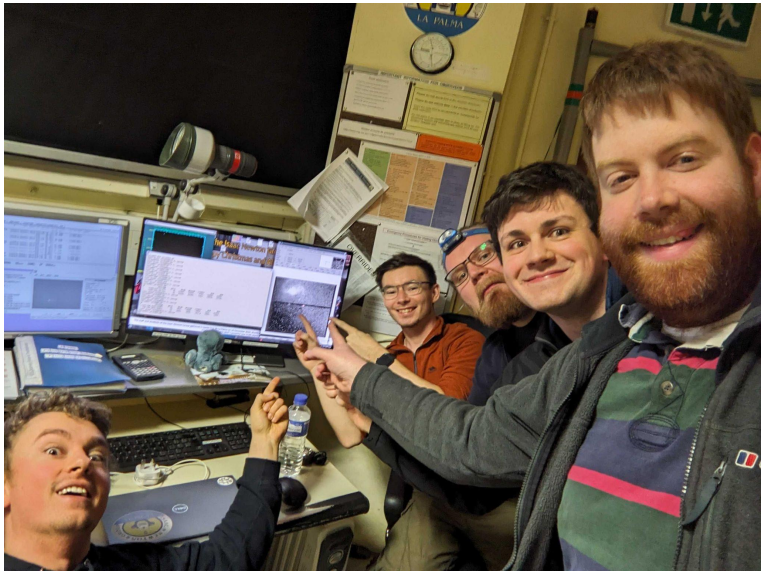
Killestein et al (under review)
arXiv: 2510.27631

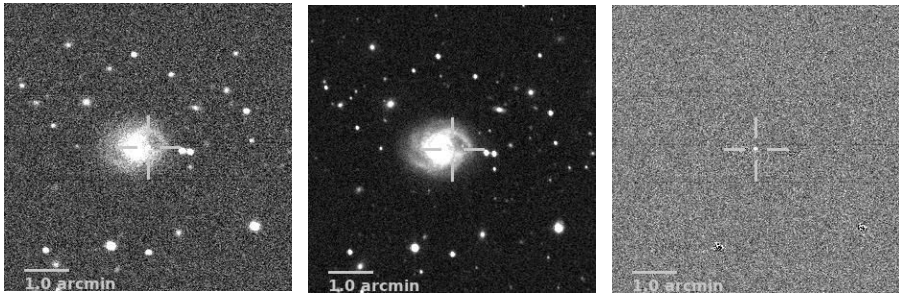


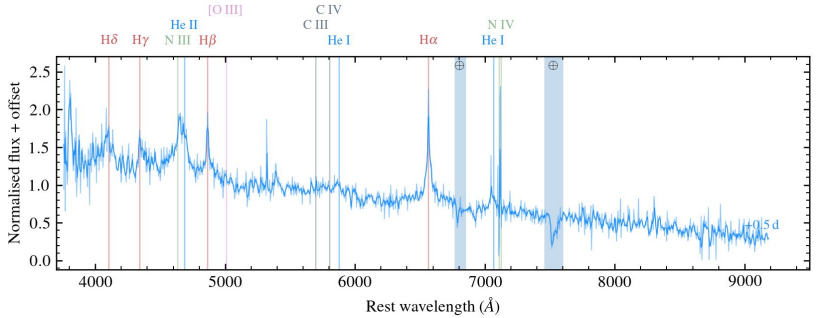
Discovery and classification

Found in **GOTO-FAST** high-cadence survey, with spectrum taken 28 mins after discovery **(12h post-explosion)**.

Flash-ionised SN II - with rarer species like C III and C IV present.

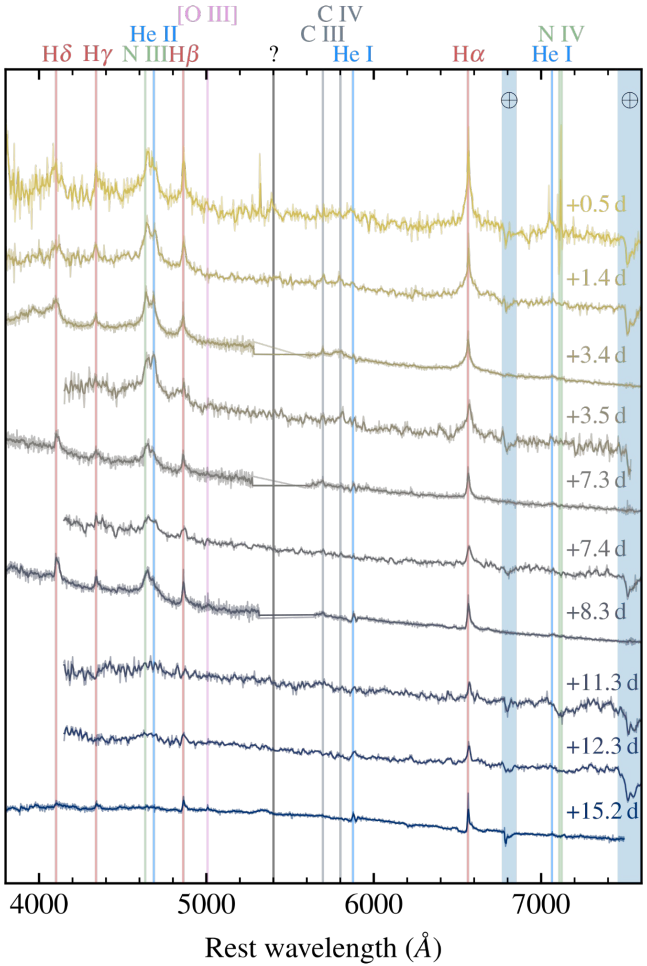




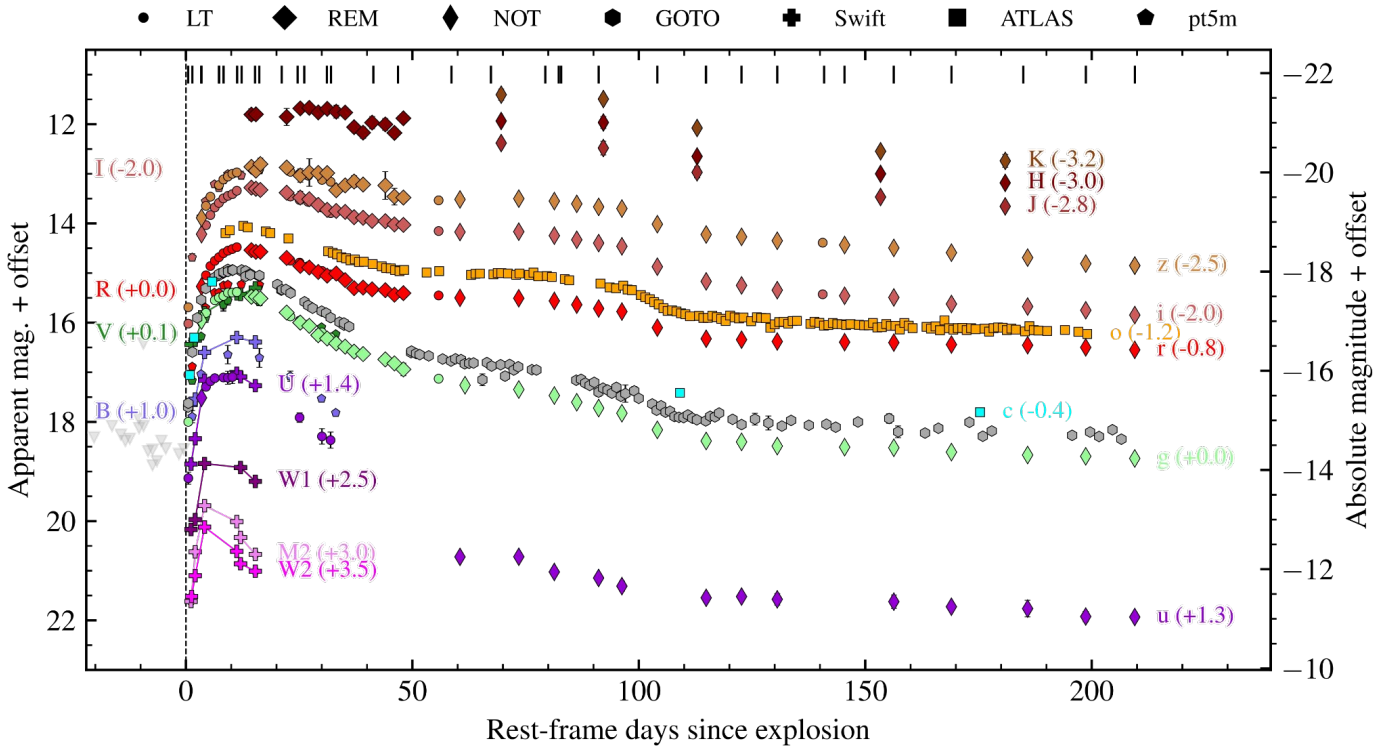


Early flash ionisation -> persistent interaction

- <4d post-explosion: flash ionisation lines of He, C, N. High ionisation state from ejecta-CSM interaction.
- Models suggest mass loss rates 10^{-2} solar mass/yr, in common with most other FI SNe
- >11d post-explosion: narrow H (and He I) emission lines remain.



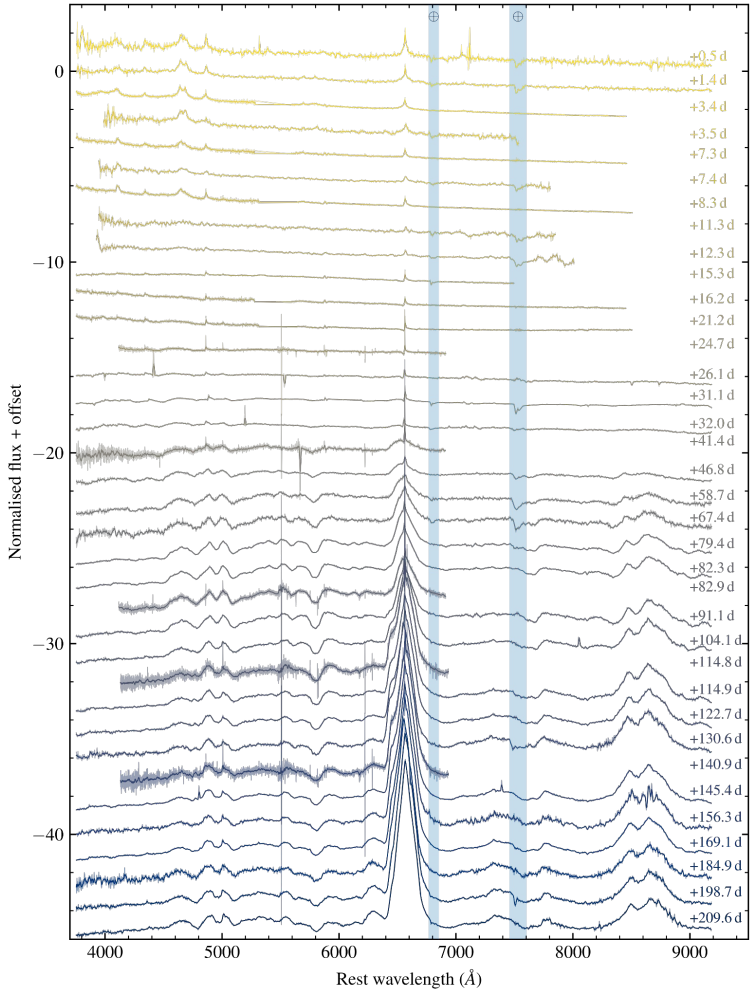
Light curve evolution



- Extended rise to peak of 14d
- Marked double plateau feature powered by interaction
- No evidence for NIR excess as seen in 1998S

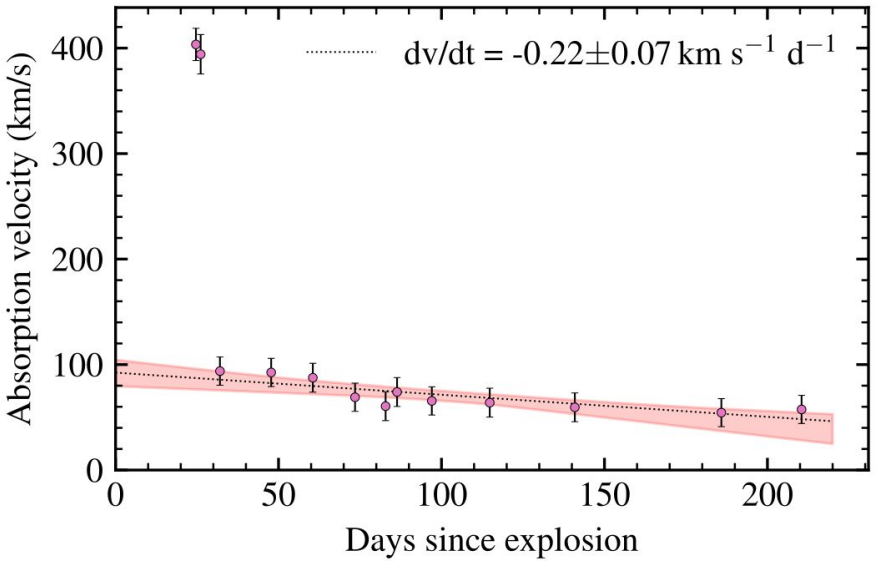
Long-term spectroscopy

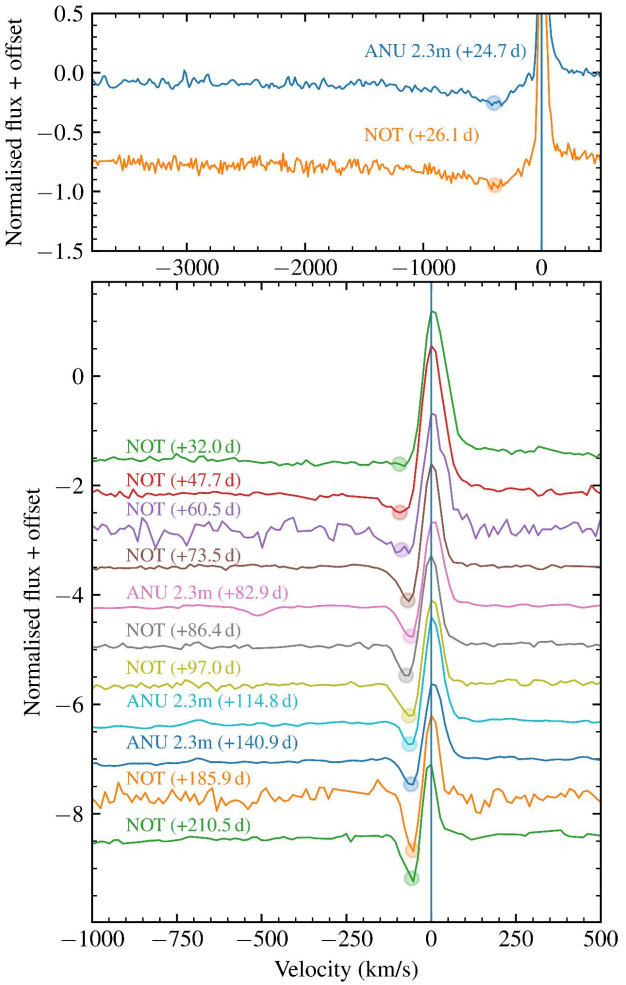
- Long delay to appearance of metal absorptions (1 month!)
- Emergence of strong multi-component Ha profile at late time - asymmetric CSM disk.
- Velocities all comparable to ordinary SNe II, with suppressed [Ca II]
- Persistent narrow emission even in low-res spectra.



High-resolution spectroscopy

Throughout the evolution we see a narrow H alpha P Cygni profile in the host rest frame - **progenitor wind?** Low velocity further evidence for RSG.

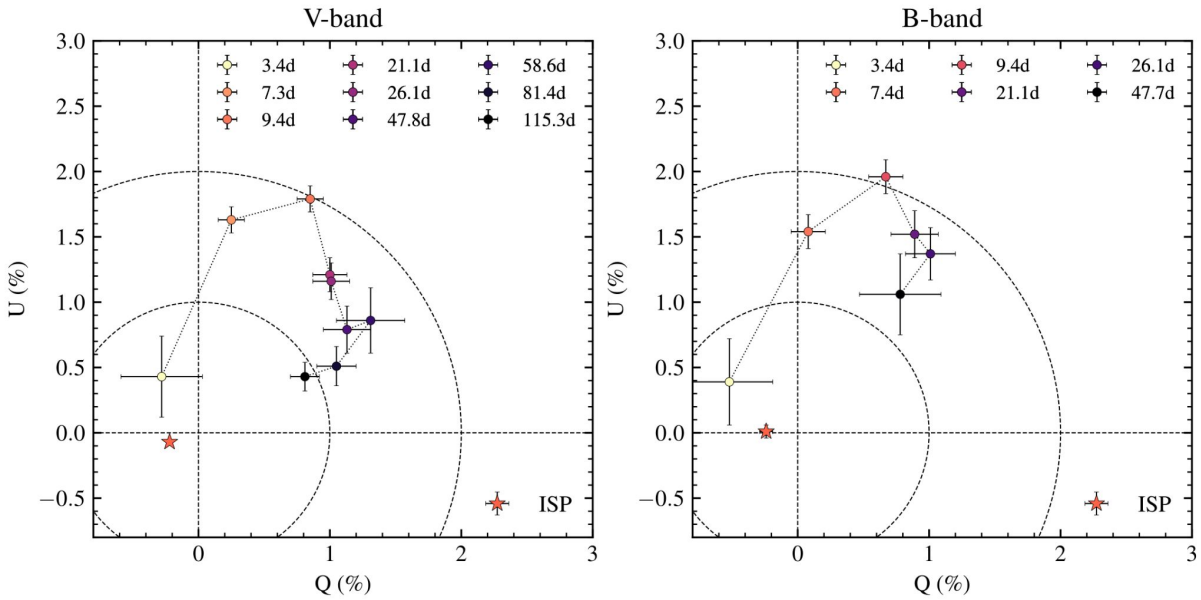




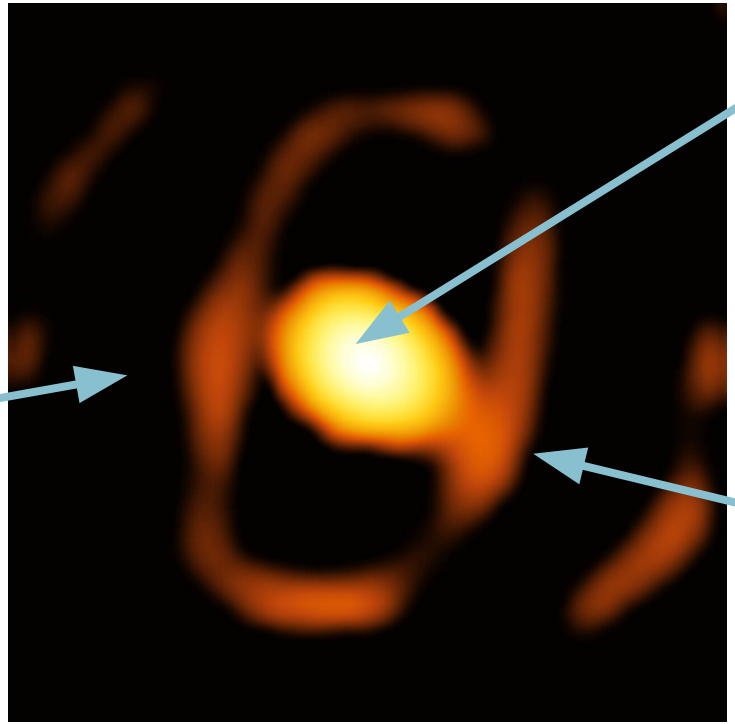
Polarimetry

See marked evolution - initially close to spherical geometry, with a rise to peak p=2% around 10d post-explosion, before slow decline.

Rotation of 40-60 deg: transition between two distinct CSM components?



Tenuous wind-driven CSM ahead of shock along line of sight, visible as very narrow P Cyg at rest



FI from aspherical extended envelope ~2000 solar radii across - powers early rise

Asymmetric CSM shell/torus - drives Ha shoulders seen later and polarisation rotation.

Ohnaka+24: VLTI imaging of WOH G64

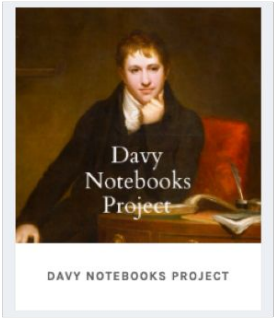
Kilonova Seekers

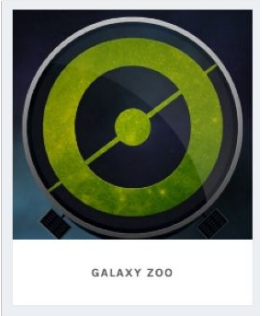
Inviting the public on a real-time journey of scientific discovery

Killestein and Kelsey et al. (2023) - project overview paper
Killestein et al **including volunteers** (2025) - Discovery of GOTO0650

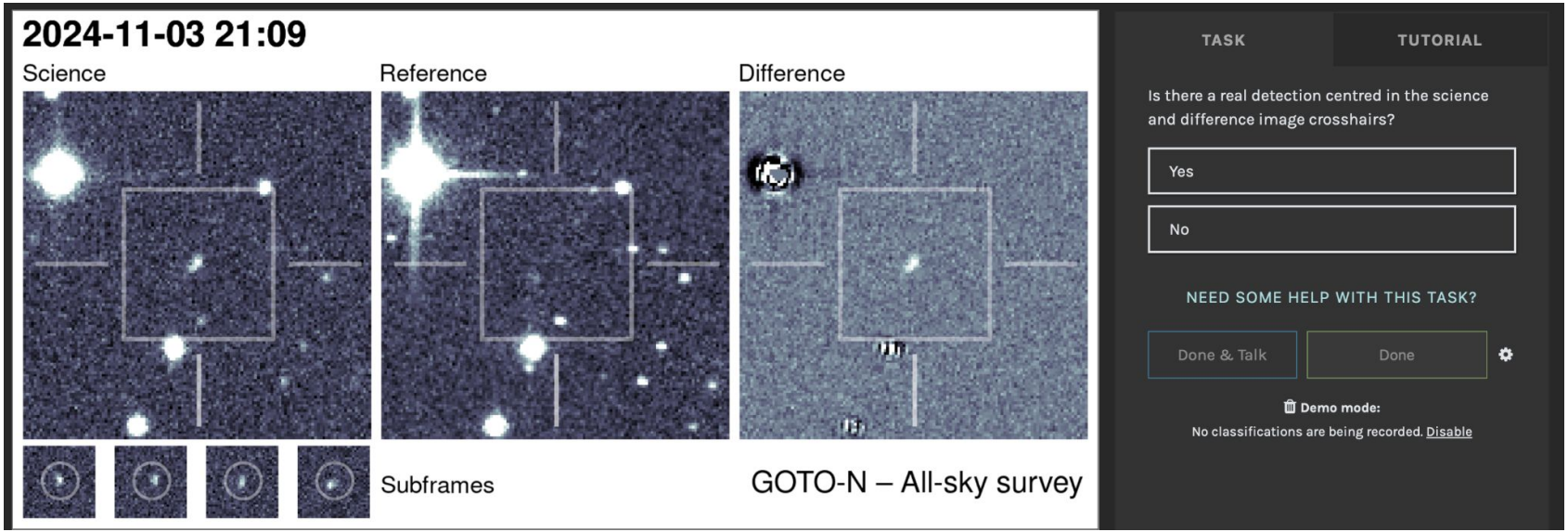
What is citizen science?

- Participation from the public
  - Things computers find difficult
  - Build a training set for machine learning
- Valuable tool for public engagement
  - Engages the public with cutting-edge science
  - Used to make real scientific progress
  - ~ 400 peer-reviewed publications using Zooniverse





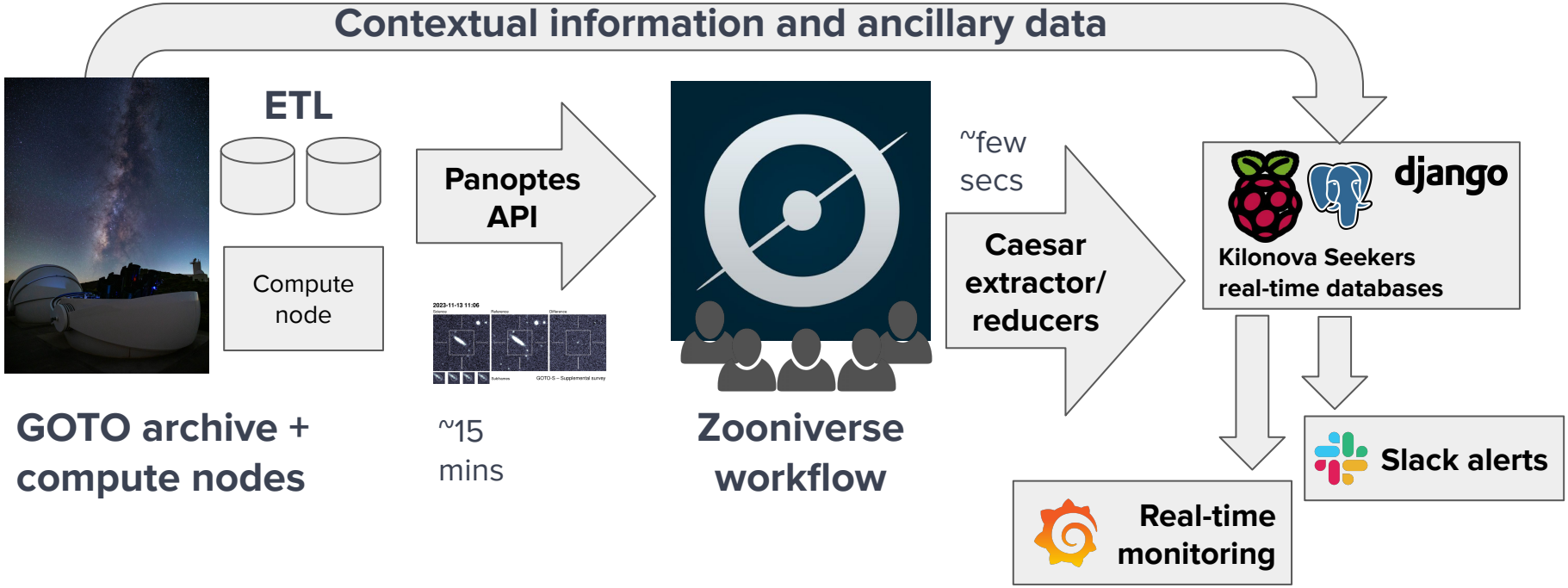
General idea: human real-bogus classification



Invite the public to play ‘spot the difference’ with the Universe with a simple binary yes/no question: aligned well to ML classification, but with **strong real-time elements** to align with GOTO’s mission for **real-time transient follow-up**.

Architecture of Kilonova Seekers

A **real-time, reactive citizen science** project - powered by lightweight infrastructure



Retirement and ‘alerting’ strategy

Alerting

80% agreement

AND

8 votes for given class

Retirement:

15 classifications

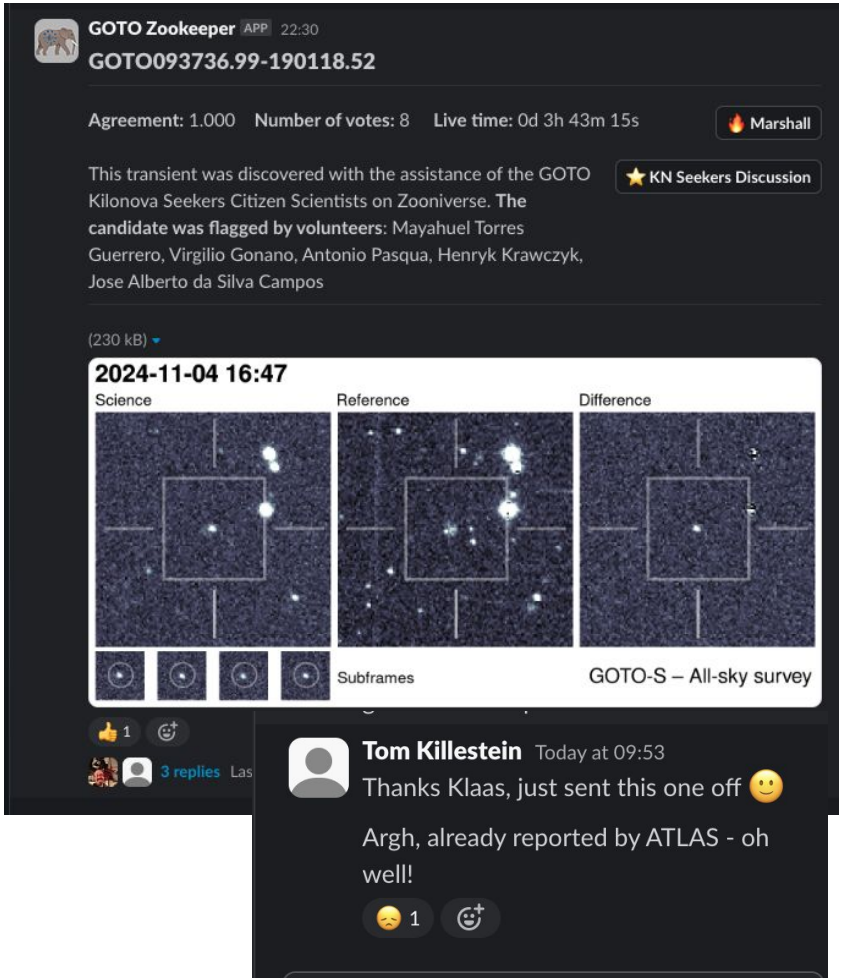
Implemented via Caesar reducers and extractors to push to real-time DB
Focus: obviously real things can quickly trigger alerts (within 1-2h or so)
But: good uncertainty estimation and high-quality labels for the final classifier training

Alerting infrastructure

Slack **“GOTO Zookeeper”** bot posts **alert cards** to a dedicated channel #knseekers-alerts

Acts as interconnect between Zooniverse and GOTO Marshall

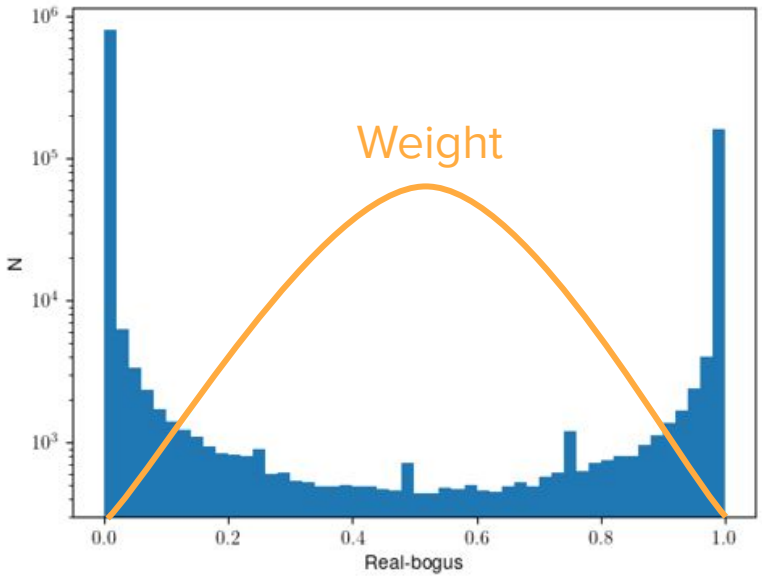
Alerts all in collaboration to objects flagged by volunteers - triage and report to Transient Name Server if interesting



What do we upload?

Balance of fostering rapid discoveries, and filling out parameter space - actively tuned based on interaction with volunteers

Key: probability-density reweighted sampling.
Upweight the uncertain, promote diverse examples



Fast discovery:

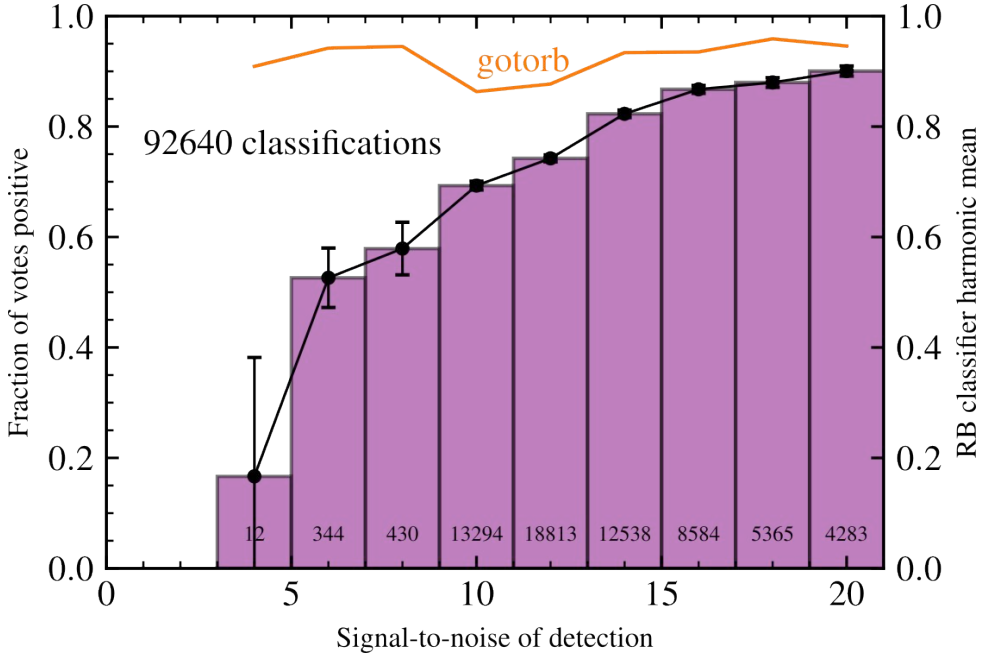
0.7 < RB < 1.0 and **NOT REPORTED ->** weight proportional to **1/PDF(rb)** and 1/10 weight to asteroids (which are otherwise dominant!)

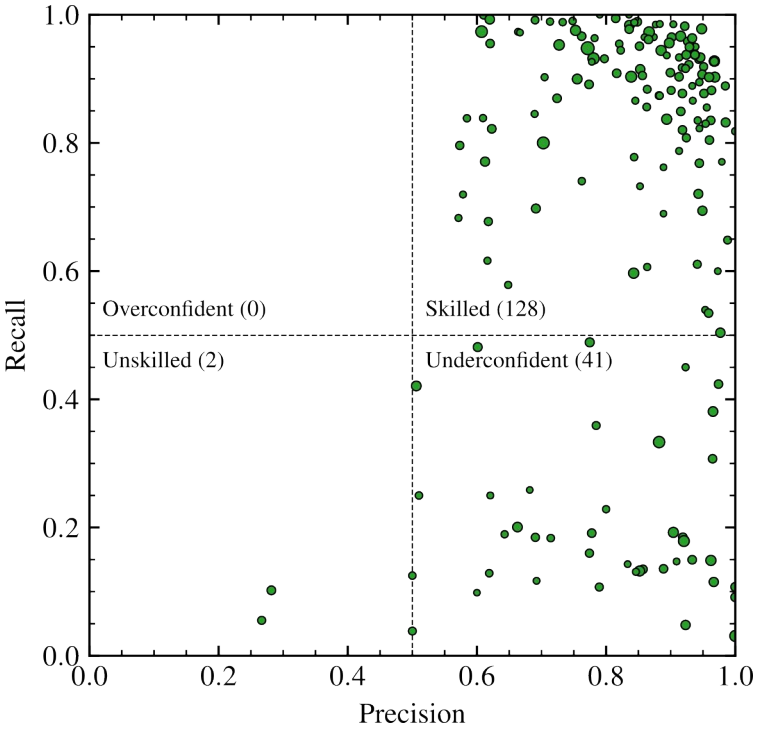
Unbiased:

0 < RB < 1 and weight proportional to **1/PDF(rb)**

Characterising the ‘human’ element

Injected validation sets: asteroids and hand-labelled data from the researchers

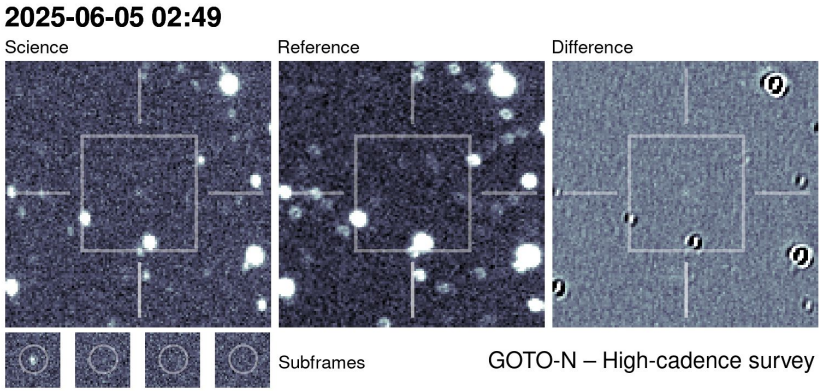




Human feedback - always gold-standard?

Inherent bias in human labelling:

- **Optimism:** Kilonova Seekers volunteers preferentially mark things with a galaxy as real to maximise chances of getting a SN discovery
- Low signal to noise far more likely to be labelled negative even if “real” statistically.



Kilonova Seekers by the numbers

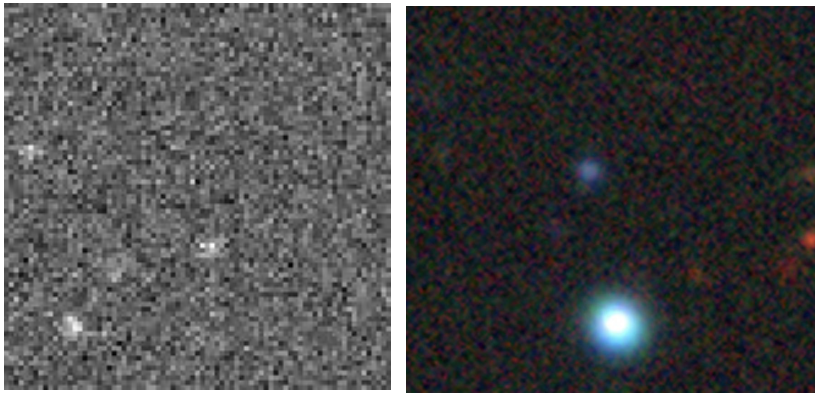
4047 volunteers have contributed to the project

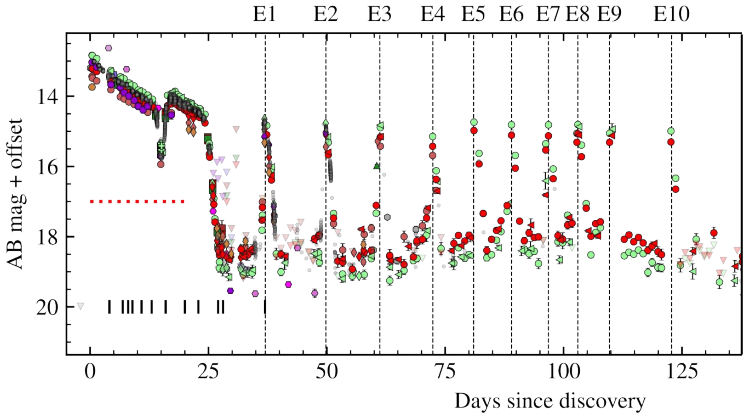
4.5M classifications have been done on GOTO data

228 reports to the IAU Transient Name Server

96 were first discoveries by our volunteers

GOTO0650: striking galactic gold with a dwarf nova





GOTO065054.49+593624.51: Discovery of a bright optical galactic transient

ATel #16842; *T. Killestein (University of Turku), L. Kelsey, G. Ramsay, M. R. Kennedy, A. Kumar, E. Wickens, K. Ackley, M. J. Dyer, J. Lyman, K. Ulaczyk, F. Jimenez-Ibarra, D. Steeghs, D. K. Galloway, V. Dhillon, P. O'Brien, K. Noysena, R. Kotak, R. P. Breton, L. K. Nuttall, E. Palle, D. Pollacco, D Oâ€ Neill, and citizen scientists: Rosemary Billington, Virgilio Gonano, Svetoslav Alexandrov, Antonio Pasqua, Cledison Marcos da Silva, report on behalf of the GOTO collaboration*

Discovered by, followed up, and written up with citizen scientist volunteers - very early discovery of a dwarf nova.

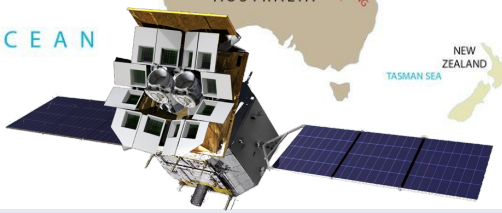
Killestein et al. 2025 in A&A (incl. volunteers)

GOTO0650 around the world!

LCO 0.4m

Other Astronomical Facilities

AAVSO observer

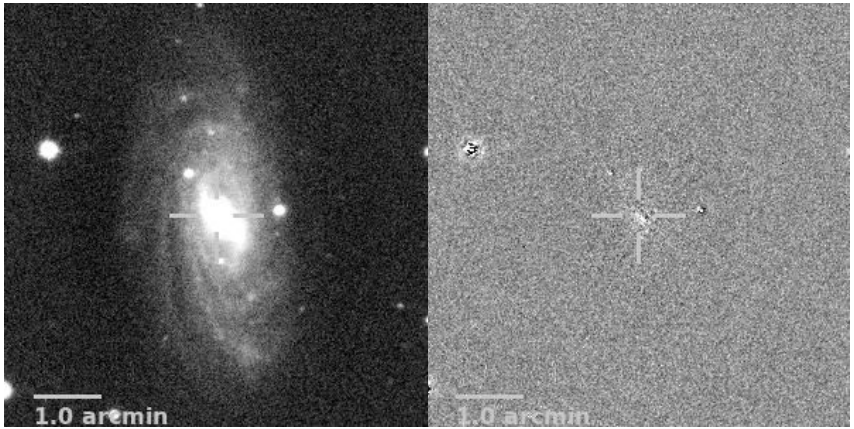


Kilonova Seekers now routinely feeds classifiers

Most recent classifier deploy (10th July 2024) included **18k** volunteer ‘gold-standard’ subjects from uniform subsampling.

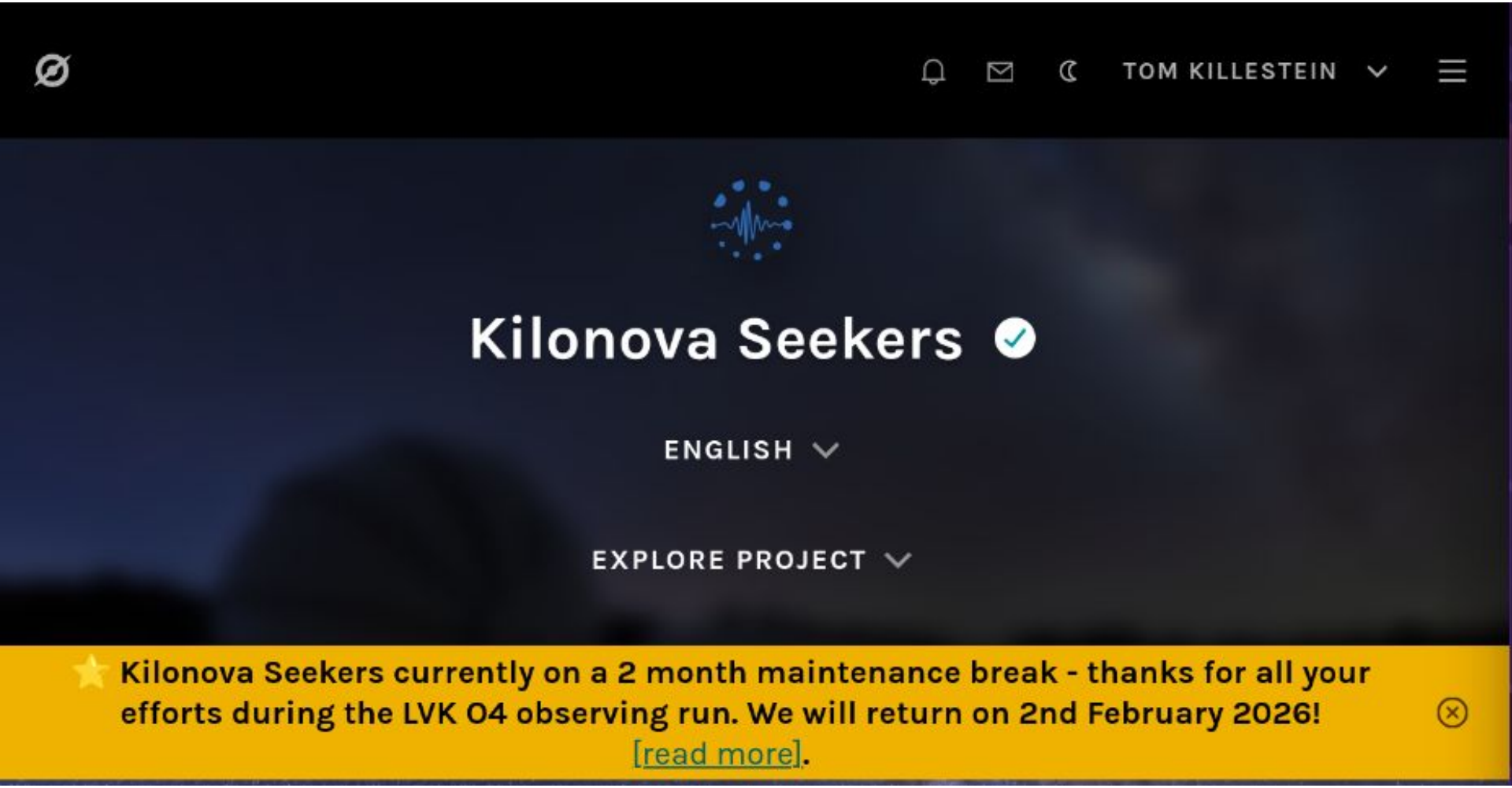
Implemented **data clipping** to mitigate bad or misspecified labels -> 99.6% accuracy + more robustness to satellite trails and other weird artefacts.

Need detailed ablation studies to measure effectiveness - but already seeing improvements in datastream



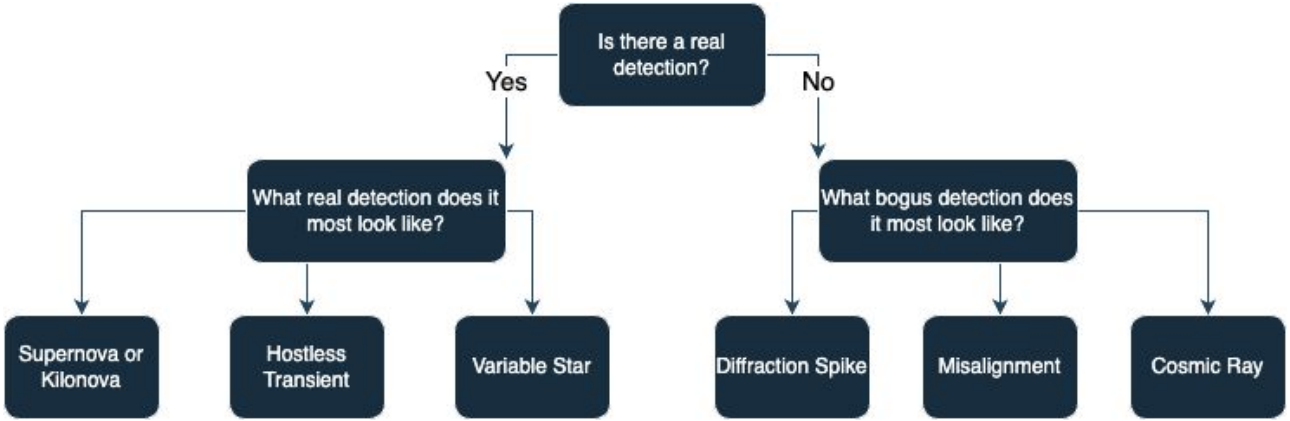
Interesting ‘overcorrection’

Volunteers excited about bright galaxies so we’re seeing more galaxy residuals: provide volunteer feedback


TOM KILLESTEIN
Kilonova Seekers
ENGLISH
EXPLORE PROJECT
Kilonova Seekers currently on a 2 month maintenance break - thanks for all your efforts during the LVK O4 observing run. We will return on 2nd February 2026!
[read more].

Kilonova Seekers multi-class

- Can we do better than simply “yes/no”?
- Real-bogus now good enough that we’re **routinely overwhelmed with real objects** (e.g. varstars!)
- Volunteers already doing this via ‘tag’ system - needs dev effort to support with e.g. contextual catalogs in way that respects source anonymity



Popular Tags:

transient
real-transient
possible-transient
candidate_sn
real
supernova
variable-star
possible-sn
possible
comet
hostless_transient
possible_transient
fermi-alert
real_transient
badsubtraction
satellite
correlatednoise
possible-supernova
misaligned
asteroid

Two major challenges

📈 Scale-up in data volume

Building performant real-bogus classifiers is (largely) solved - how do we do **complex, multi-label classification tasks** in the context of petabyte-scale data streams, with **finite human labelling effort**?

📉 Data drift degrading performance

How do we ensure classification algorithms remain up-to-date against ever-changing data streams: to avoid **performance degradation?**

How can we **track and monitor** data drift in the context of streams?

Key principle:

How can we ~~minimise~~ human involvement?
How can we ***optimise*** human involvement?

thomas.killestein@warwick.ac.uk
https://tkilleste.in