

MMA Subgroup Talk - Michael Coughlin Group

Brian Healy
Postdoctoral Associate, UMN



(OAC-2117977)

Coughlin Group members

Postdocs:

- Brian Healy
- Saleem Muhammed

Grad students:

- Tyler Barna
- William Benoit
- Brendan King
- Rafia Omer
- Andrew Toivonen

Postbac:

- Abigail Gray

- *See also talks by Alec Gunny, Eric Moreno about GW efforts*

Time-domain Astronomy

- Many astronomical objects show brightness variations on short timescales
 - Persistent variables, transients
- To study time-domain astronomy at scale:



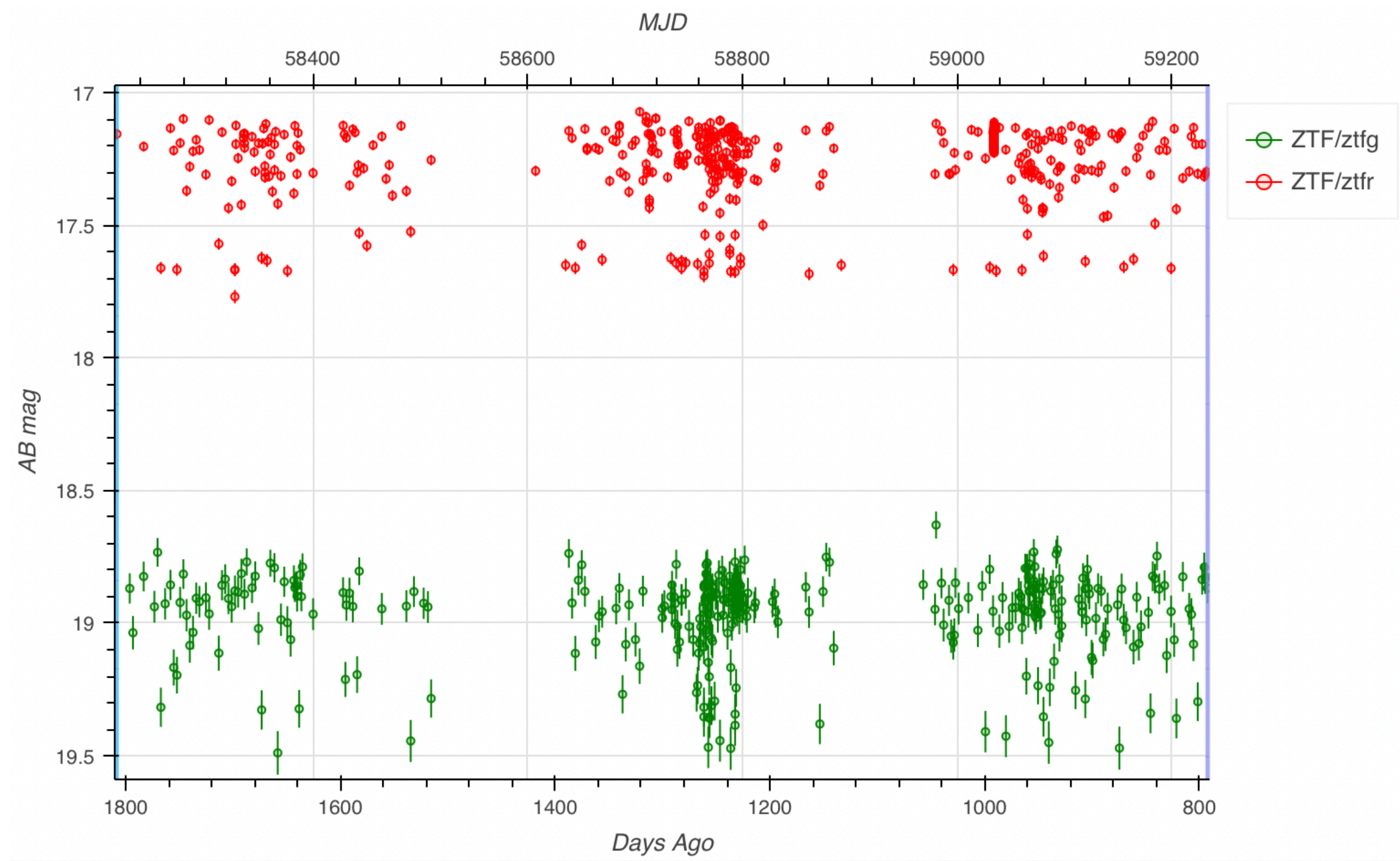
Zwicky Transient Facility (ZTF)

- Northern sky survey @ Palomar Observatory, CA
 - Wide field of view (47 deg²)
 - Fast survey rate (4300 deg²/hr)
 - Observes faint sources in multiple colors; median single visit depth 20.6 mag (5σ , r)
- Full-sky coverage every 2 days
 - DR16: 4.64 billion light curves

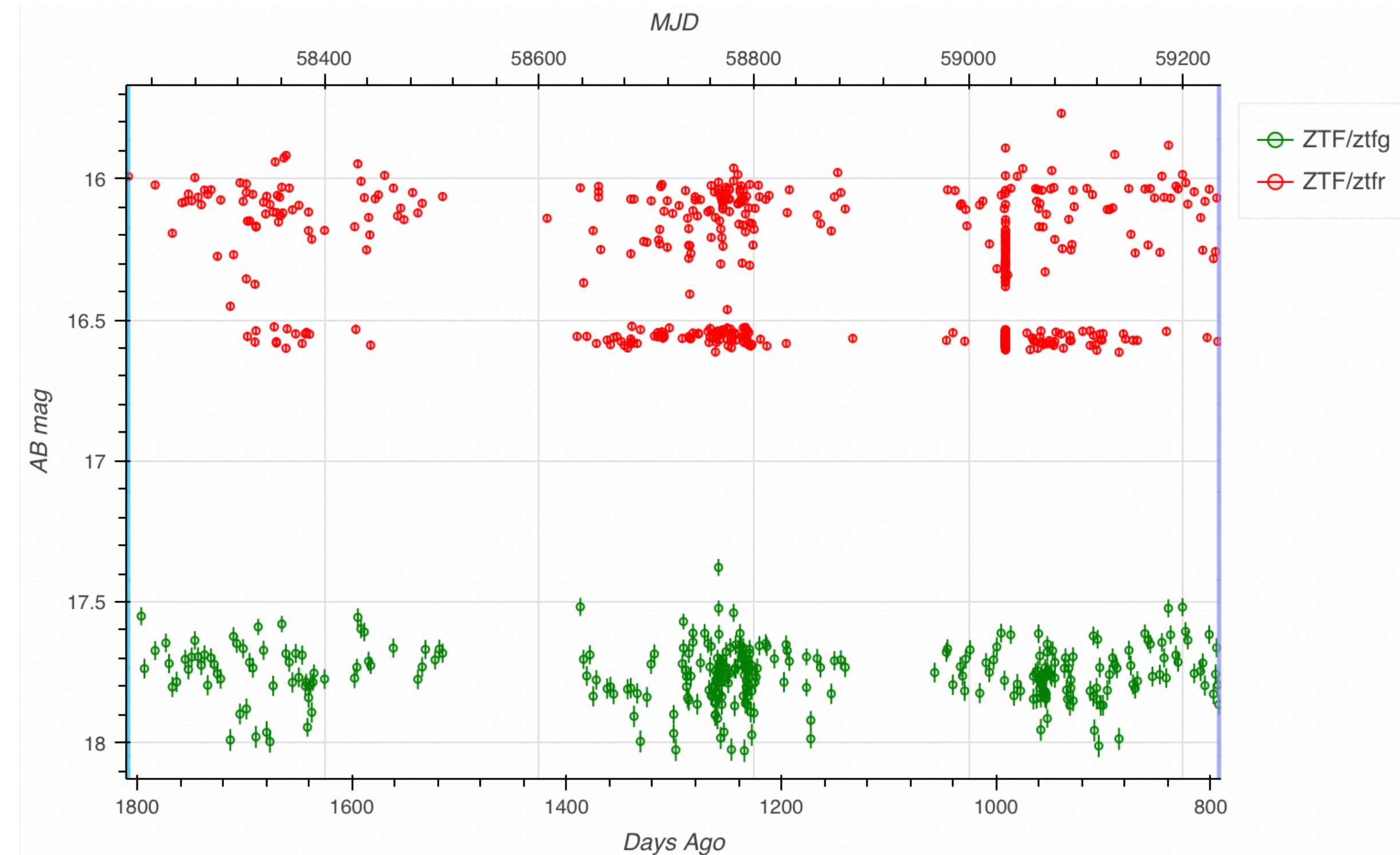


(Palomar Observatory/Caltech)

Example ZTF Data



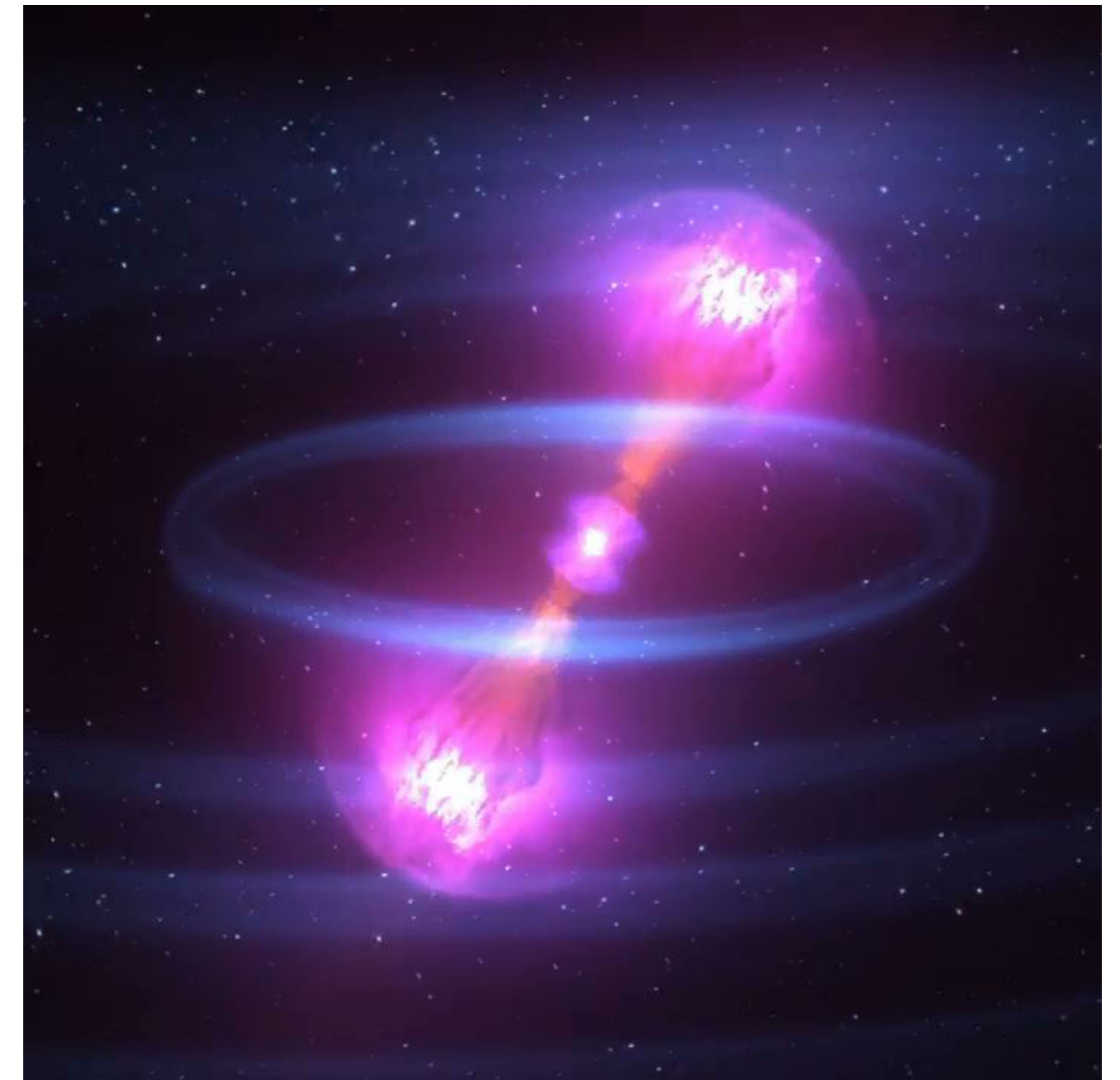
**Astrophysical source
(eclipsing binary)**



**Artifact
(blend)**

Motivating Source Classification

- Source classification benefits:
 - Variable star studies
 - Enable ensemble studies
 - Transient follow-up, esp. kilonovae (neutron star mergers)
 - Exclude sources observed to exhibit past variability
 - Anomaly detection
 - Discover new sources by searching for anomalous outliers



(NASA GSFC/CI Lab)

ZTF Source Classification Project (SCoPe)



The screenshot shows the GitHub repository page for the ZTF Source Classification Project (SCoPe). The repository is public and has 5 stars, 5 watchers, and 17 forks. The main branch is selected, and there are 2 branches and 0 tags. The repository contains a list of files and folders, including .github, .requirements, data, doc, periodfind @ 9413dac, scope-phenomenology @ f95d445, scope, tools, .flake8, .gitignore, .gitmodules, .pre-commit-config.yaml, LICENSE, README.md, combine_preds.py, config.defaults.yaml, get_all_preds.sh, pyproject.toml, and requirements.txt. The repository is described as "SCoPe: ZTF source classification project" and is available at zwickytransientfacility.github.io/scop.... The repository is open-source, Python-based, and has a CI/CD pipeline. It is regularly updated and has documentation. The repository was created during a hackathon.

File/Folder	Description	Last Commit
.github	Update inputs for doc deploy workflow (#294)	2 weeks ago
.requirements	Add tests for inference and active learning sample selection (#291)	2 weeks ago
data	Add variable object examples (#45)	2 years ago
doc	Impute features when using get_features.py (#292)	2 weeks ago
periodfind @ 9413dac	Pin latest periodfind (#311)	2 days ago
scope-phenomenology @ f95d445	Pin scope-phenomenology (#284)	3 weeks ago
scope	Sort light curves to be monotonically increasing in time (#315)	22 minutes ago
tools	Sort light curves to be monotonically increasing in time (#315)	22 minutes ago
.flake8	Initialize repository structure (#1)	2 years ago
.gitignore	Field guide and workflows (#4)	2 years ago
.gitmodules	Update periodfind URL (#299)	last week
.pre-commit-config.yaml	Fix failing tests due to changed repo location (#168)	4 months ago
LICENSE	Initial commit	3 years ago
README.md	Update documentation URL (#259)	last month
combine_preds.py	Inference pipeline (#84)	6 months ago
config.defaults.yaml	Loop over config-specified period algorithms (#303)	last week
get_all_preds.sh	Inference pipeline (#84)	6 months ago
pyproject.toml	DNN model training pipeline (#6)	2 years ago
requirements.txt	Field guide and workflows (#4)	2 years ago

- Open-source
- Python-based
- CI/CD pipeline
- Regularly updated docs
- Hackathon topic!

(van Roestel et al. 2021, Coughlin et al. 2021)

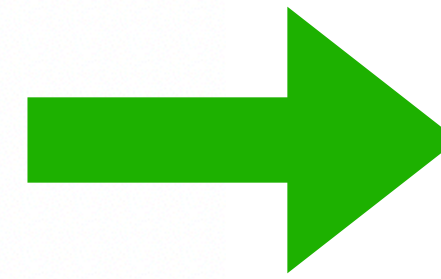
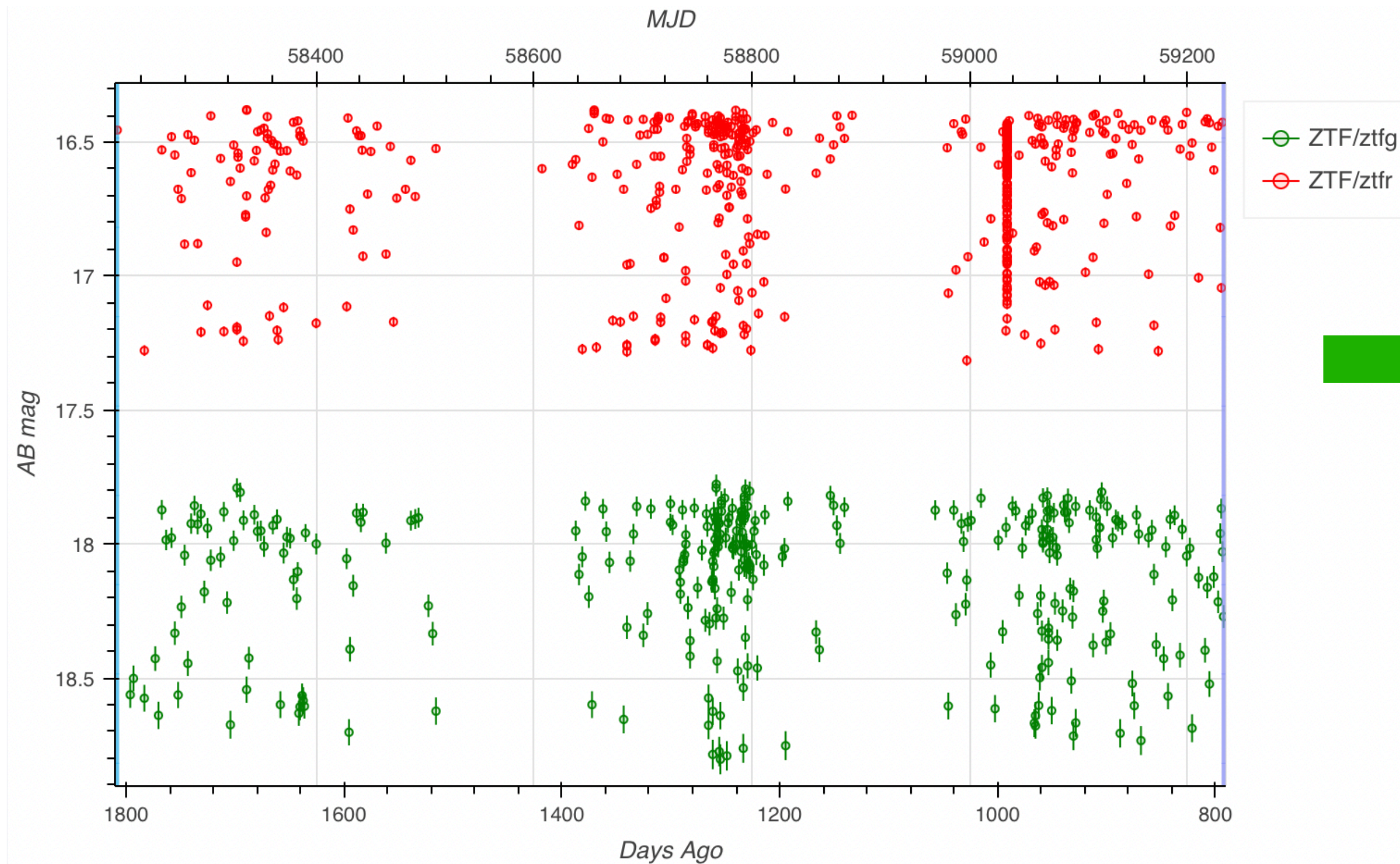
SCoPe Details

- **Supervised, active learning:** training set built up over time (w/human input)
- **Two taxonomies:** ontological (intrinsic), phenomenological (light curve shape)
 - Provides useful information for anomalous sources
 - Avoids complications of overlapping classes

SCoPe Details

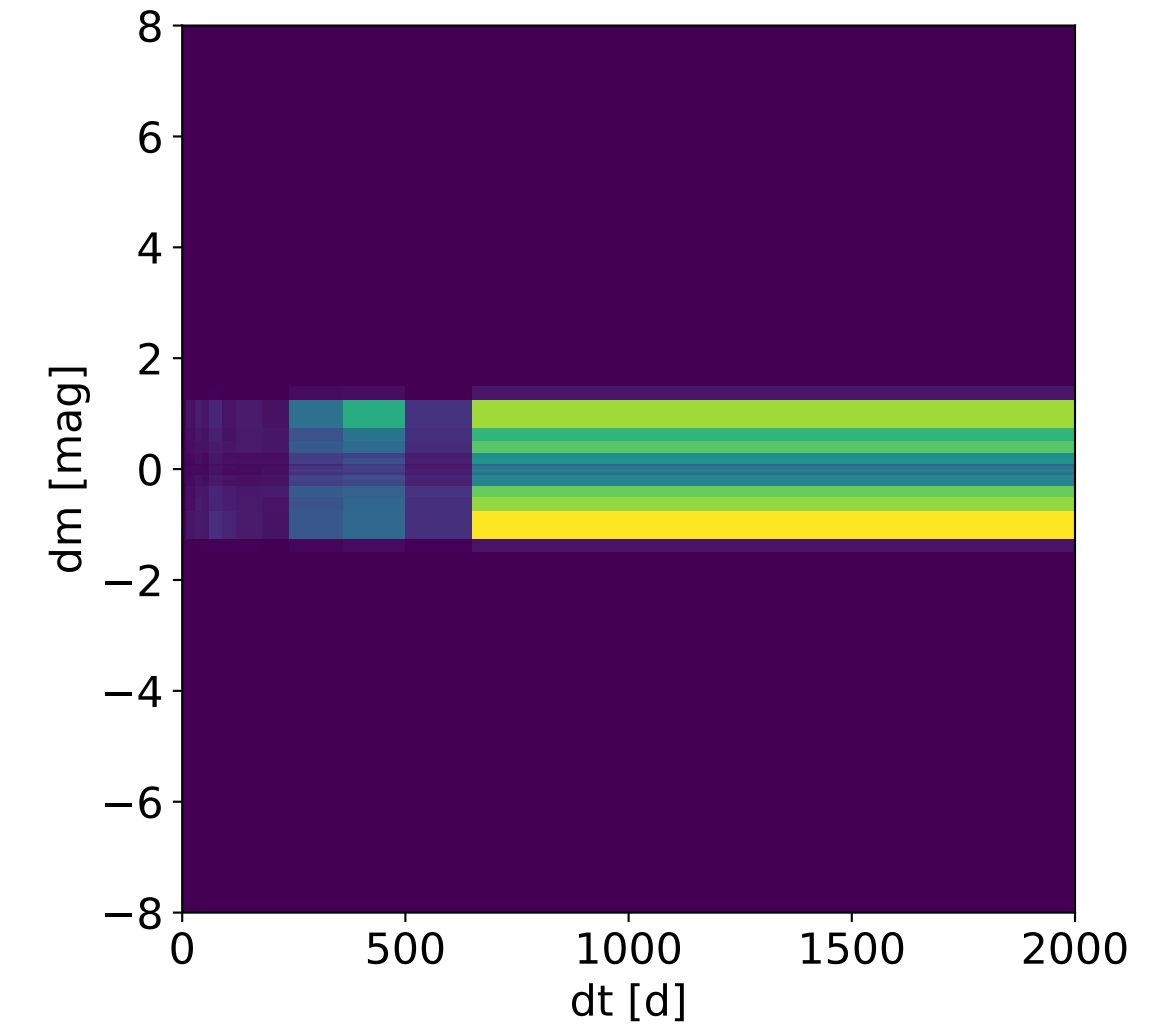
- **Binary classification:** independent predictions for each class
 - Supports multiple labels per source (varying specificity)
 - Flexible to new labels
- Train **convolutional neural network** and **XGBoost** algorithms on each label

Input Features



Summary statistics

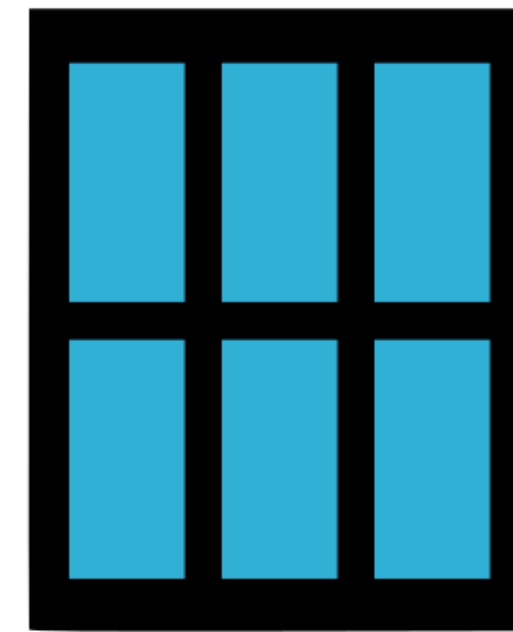
f1_a	18.137365
f1_amp	0.284065
f1_b	-0.000034
f1_phi0	-0.098306
f1_power	0.93193
f1_relamp1	0.082111
f1_relamp2	0.024026
f1_relamp3	0.0
f1_relamp4	0.0
f1_relphi1	-0.075452
f1_relphi2	-0.03388
f1_relphi3	0.0
f1_relphi4	0.0
field	853
i60r	0.316
i70r	0.3674
i80r	0.4464
i90r	0.5692
inv_vonneumannratio	0.71049
iqr	0.25
mean_ztf_alert_braai	0.99176
median	18.014
median_abs_dev	0.101
n	137.0
n_ztf_alerts	406
norm_excess_var	0.000093
norm_peak_to_peak_amp	0.016749
pdot	0.0
period	0.175088
quad	1
ra	355.275604
roms	3.267903
significance	167.419434
skew	67.574478
smallkurt	826.596439
stetson_j	-1.10259
stetson_k	0.815673
sw	0.882065
welch_i	71.02061
wmean	18.050412



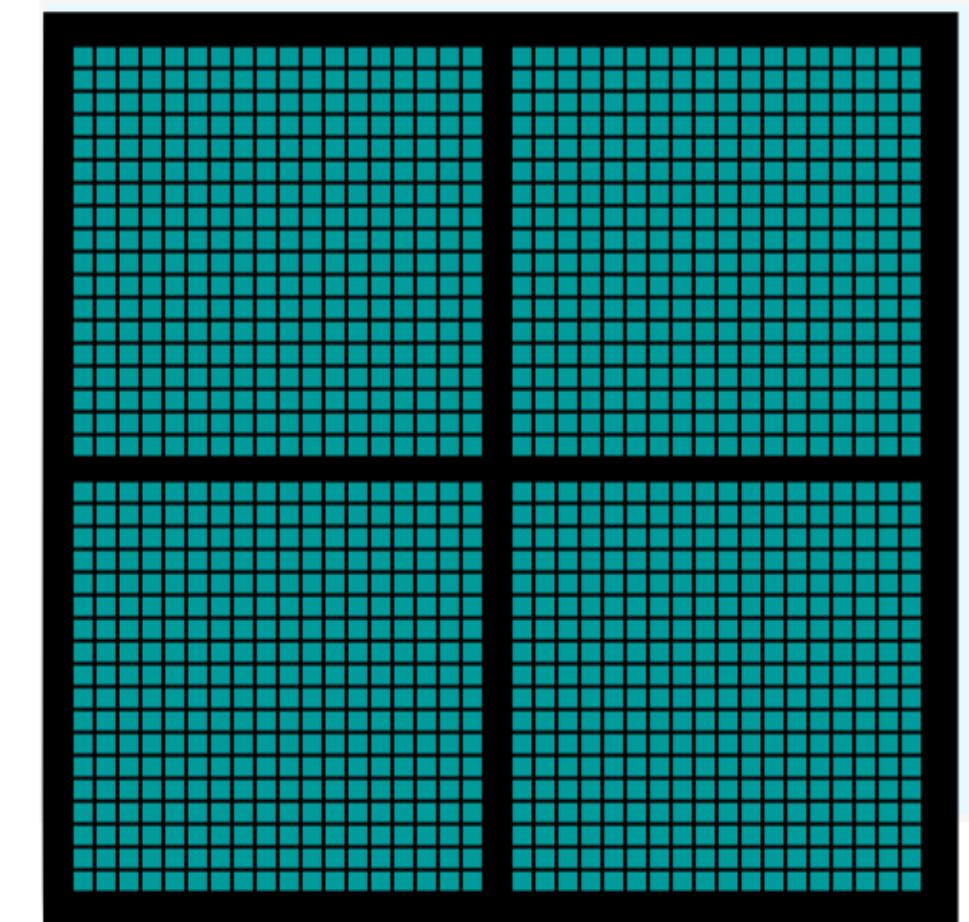
Magnitude-time histograms

Finding Periods

- Lomb-Scargle periodogram takes ~1s per light curve on laptop CPU
 - $\sim 10^9$ LCs: every second counts
- GPU-accelerated algorithms reduce runtime to ~ 0.1 s / LC
- Run in parallel on high-performance compute (HPC) cluster to further accelerate



CPU
Multiple Cores



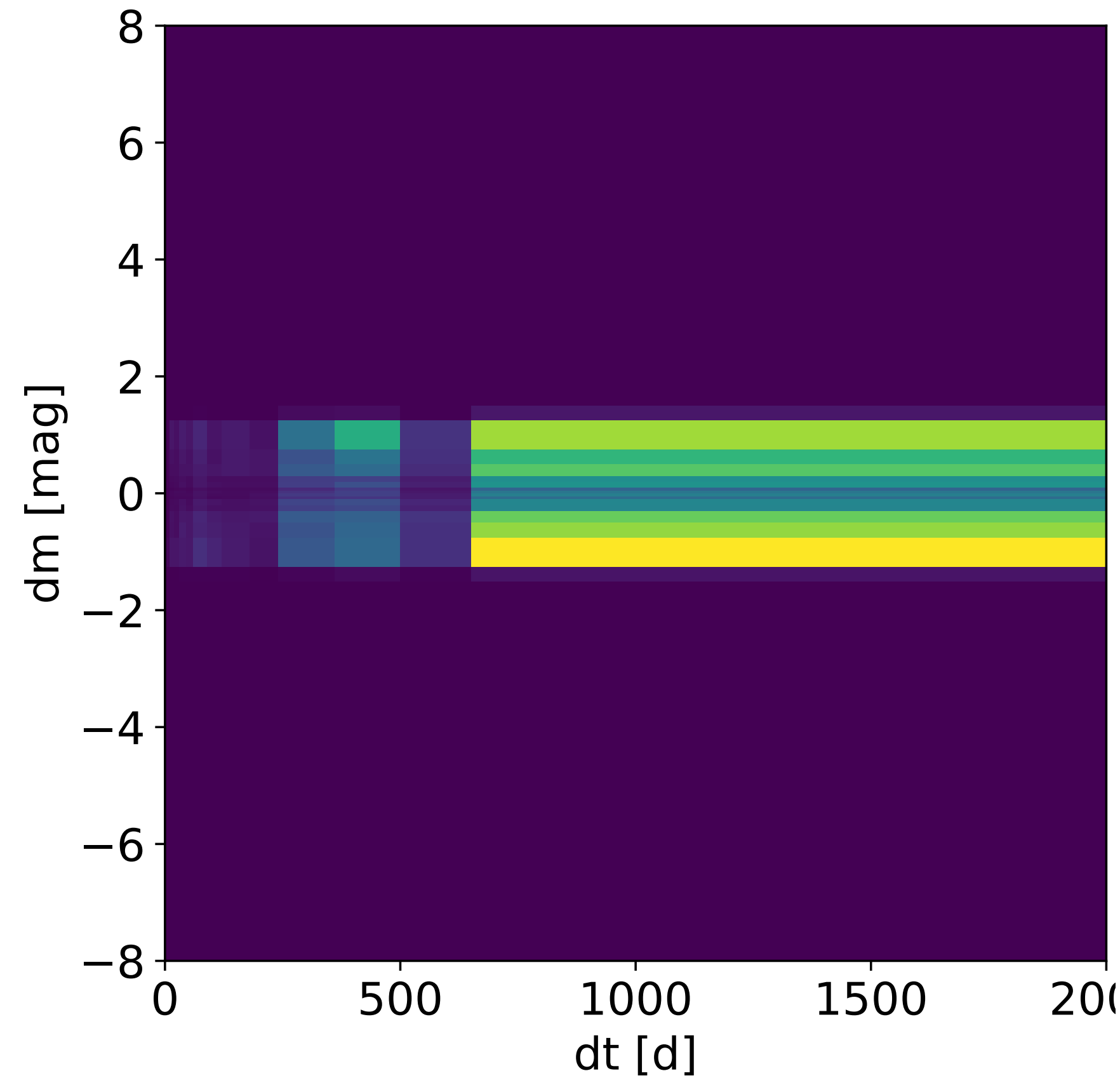
GPU
Thousands of Cores
(cherryservers.com)



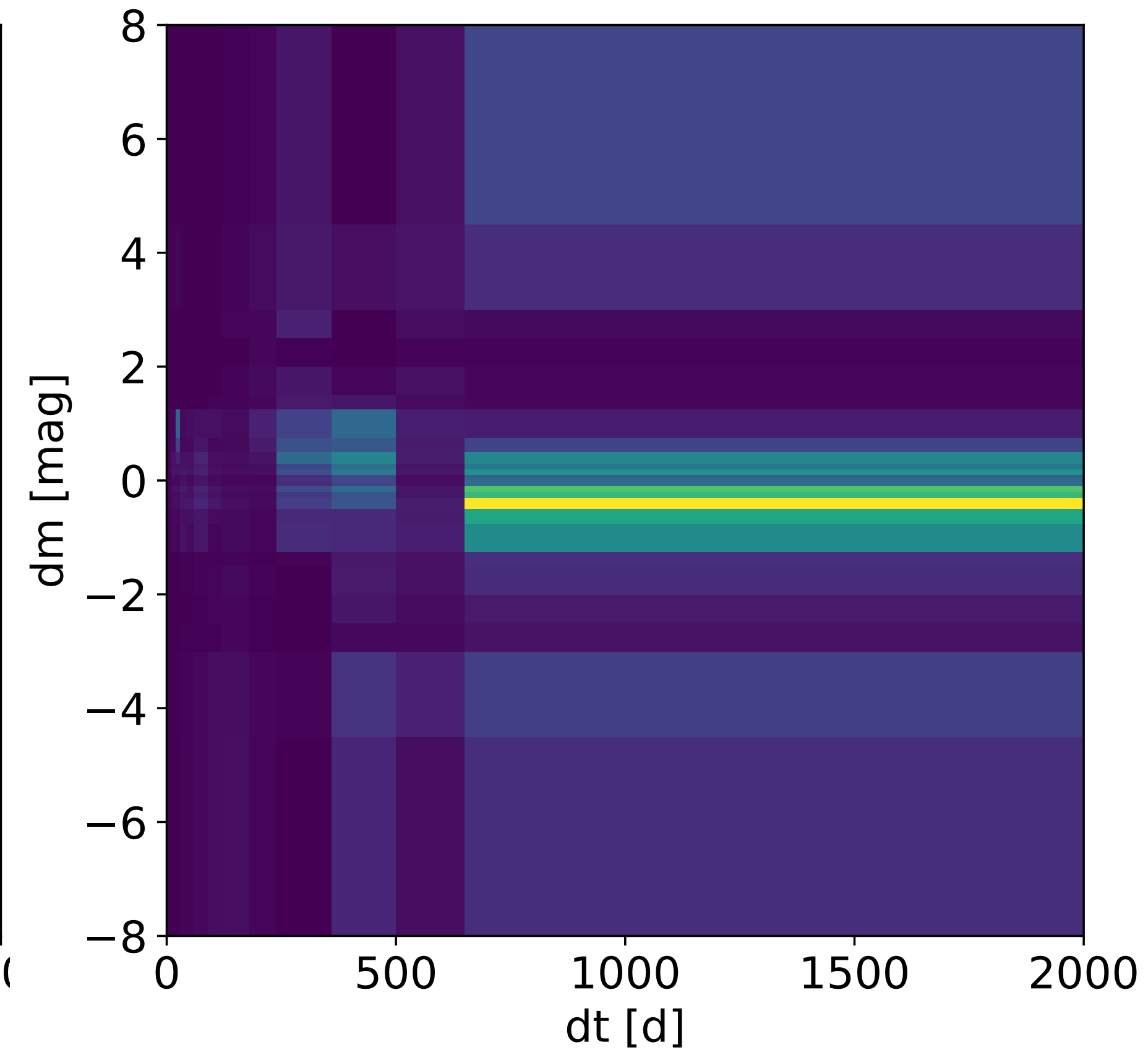
(expansion.sdsc.edu)

Magnitude-Time Histograms (dm dt)

- Map changes in magnitude (brightness), time to 26 bins
- Create standardized “image” as additional feature
- Improves classifier performance (Mahabal et al. 2017)



Eclipsing binary



Flaring

S-CoPe Neural Network Architecture

- Most features pass through fully-connected layers
- dmdt histograms input to convolutional branch
- Dropout layers help avoid overfitting
- One output per classifier (probability)



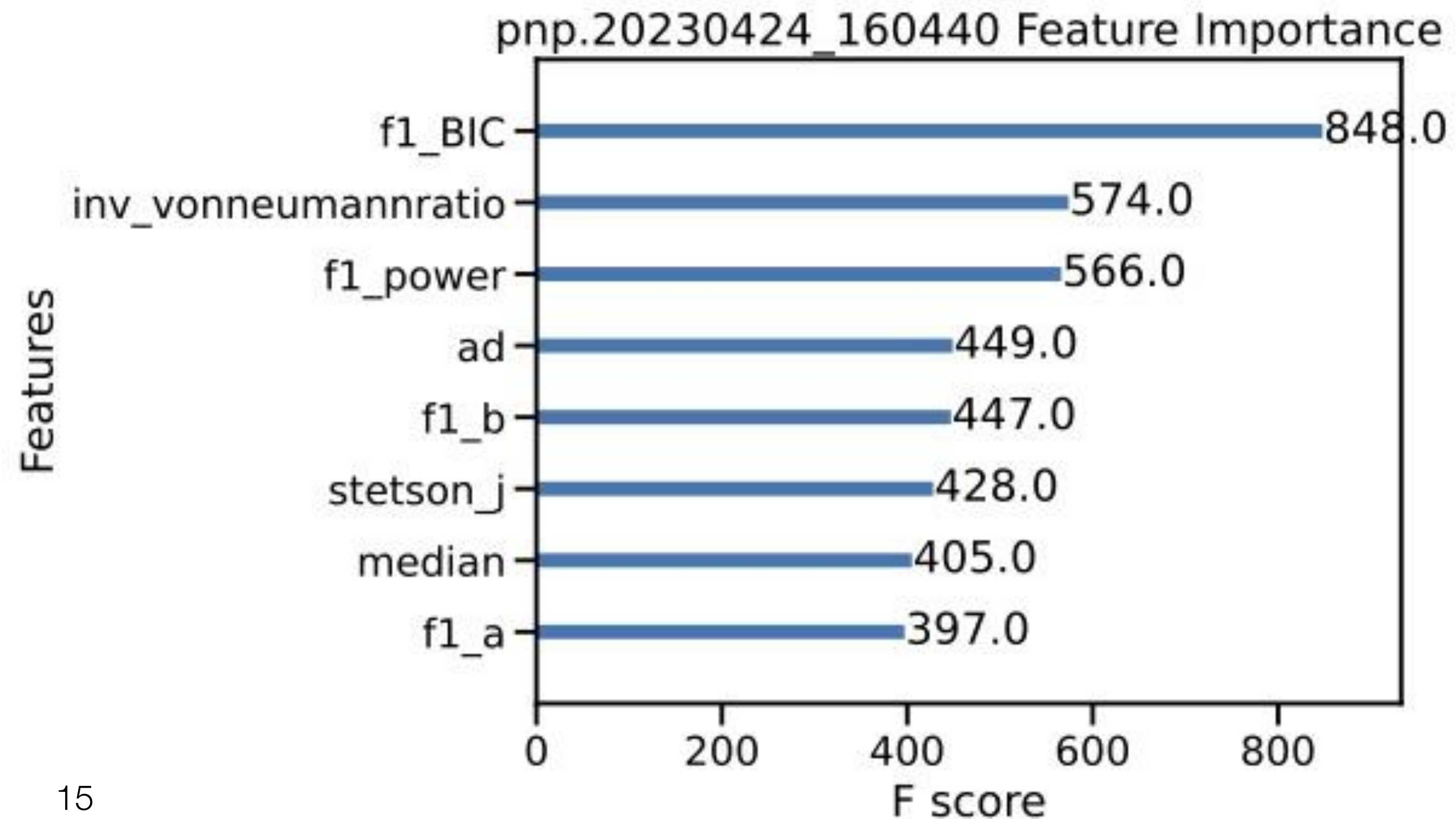
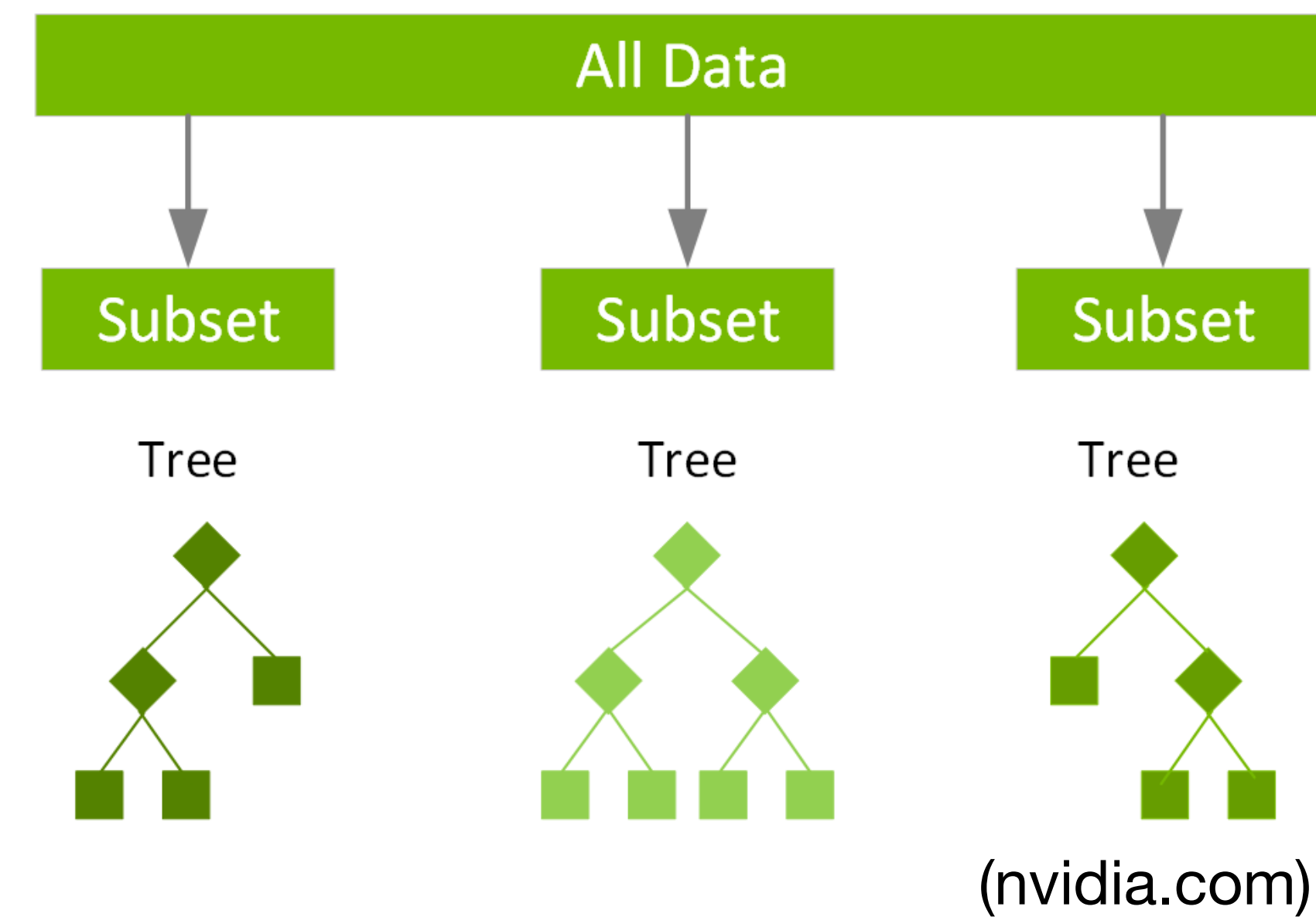
NN Training Details

- ~80,000 sources in training set
 - ~180,000 light curves
- 44 binary classifiers
 - Some labels well-represented, others highly imbalanced



XGBoost

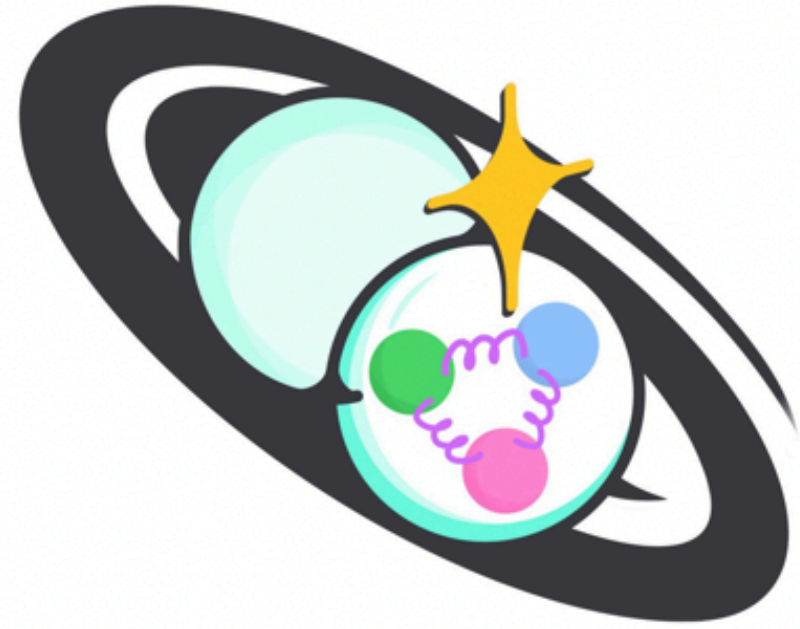
- Combines individual decision trees sequentially, minimizing underfitting
- Feature importance plots generated for each label
- Achieves similar or better performance vs. DNN



Current SCoPe status

- DR16 features generated for 20 ZTF fields on SDSC Expanse GPUs
- Classifications obtained from DNN and XGB algorithms
- Trained models run on sources near transient candidates to assist with follow-up decisions
- Paper in prep to report methods, classifications
 - Catalog of classifications to be made public

Nuclear Multi-Messenger Astronomy (NMMA)



NMMA

NMMA

a pythonic library for probing nuclear physics and cosmology with multimessenger analysis

Please check our official documentation here: <https://nuclear-multimessenger-astronomy.github.io/nmma/>

How to contribute: <https://nuclear-multimessenger-astronomy.github.io/nmma/contributing.html>

A tutorial about how to produce simulations of lightcurves is given in `nmma/tutorials/tutorial-lightcurve_simulation.ipynb`

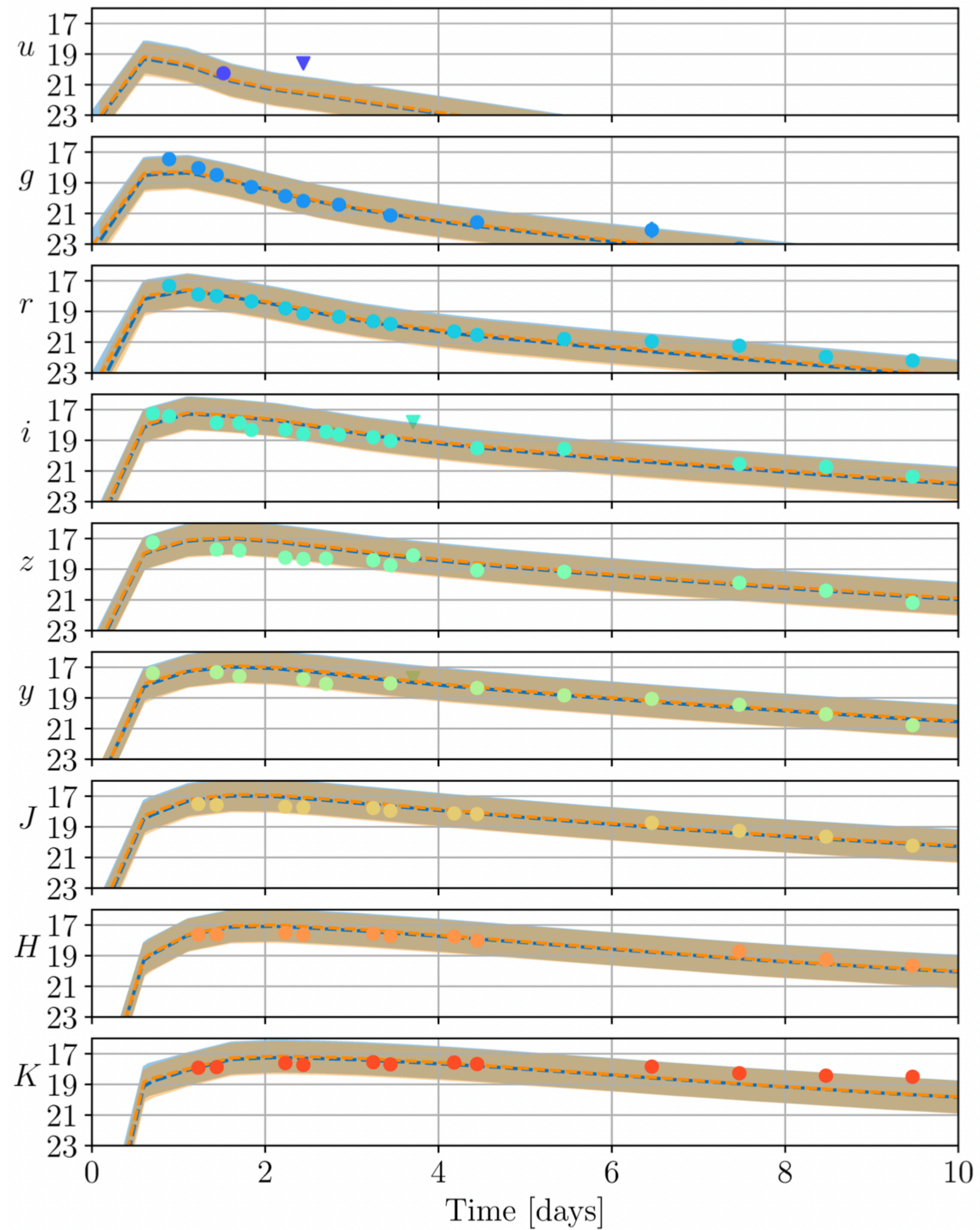


- Open-source
- Python-based
- CI/CD pipeline
- Regularly updated docs

(Pang et al. 2022)

NMMA basics

- MMA pipeline for Bayesian analysis of GW, EM data
- Includes models of KNe, GRB, SNe, etc.
- Workflow: train models on simulated data, download when needed, sample from priors to fit observations and determine event parameters
- Use cluster resources to run at scale



**AT2017gfo
(Pang et al. 2022)**

Training NMMMA models

- Grid of physical parameters produces simulated light curves
 - Want continuous mapping between parameters, light curves
- LCs mapped to eigenvalues using PCA
- Gaussian process/neural network maps between input params, eigenvalues
 - Broad, few-layer NN approximates GP in less compute time

Status/goals for NMMMA

- Updated models being used for analyses of past events
 - New tensorflow training runs faster than GP
- Adding new features as needed, expanding documentation
- Run on GW, EM data as fast as they become available
- Use API service to start automated analyses
- Submit paper \leq 48 hrs after event detected

Kilonova Lightcurves

Data products:

- BNS_ejecta: probability of BNS and $m_{ej} > 10^{-3} M_{\odot}$
- NSBH_ejecta: probability of NSBH and $m_{ej} > 10^{-3} M_{\odot}$
- Peak UVOIR M_{AB} estimate and m_{ej} estimate (90 percent credible interval)

