

IVOA Interop, Victoria – 31<sup>st</sup> May 2018

# ML aspects of ZTF

Matthew J. Graham  
Research Professor  
ZTF Project Scientist

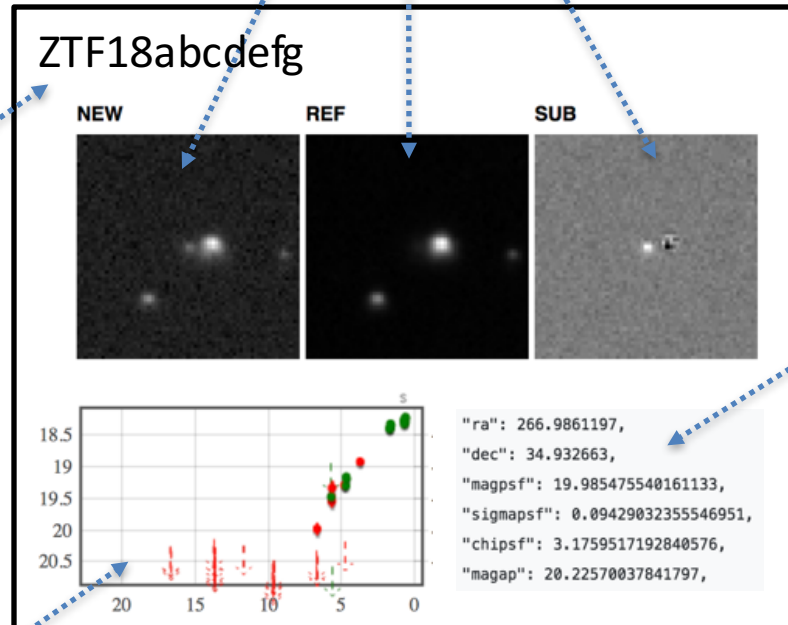
(and Ashish Mahabal, Umaa Rebbaprigada, Adam  
Miller, Yutaro Tachibana)



# Alert structure: AVRO format

63 x 63 pixel 32-bit images

Unique spatially matched alert name



- ZOGY parameters
- Real-bogus score
- Star/galaxy score
- 3 nearest PS1 sources
- Nearest SS object
- Alert history

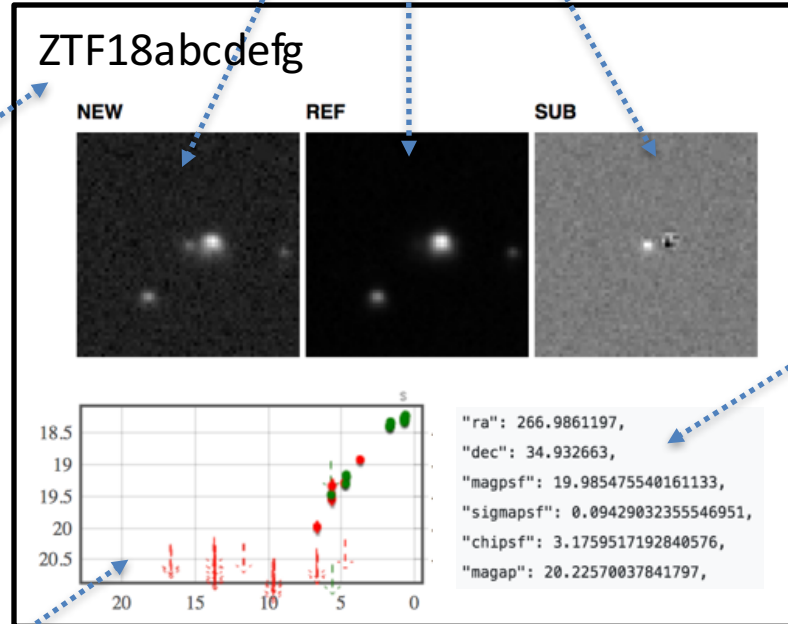
Rolling 30-day window light curve

<https://github.com/ZwickyTransientFacility/ztf-avro-alert>

# Alert structure: AVRO format

63 x 63 pixel 32-bit images

Unique spatially  
matched alert name

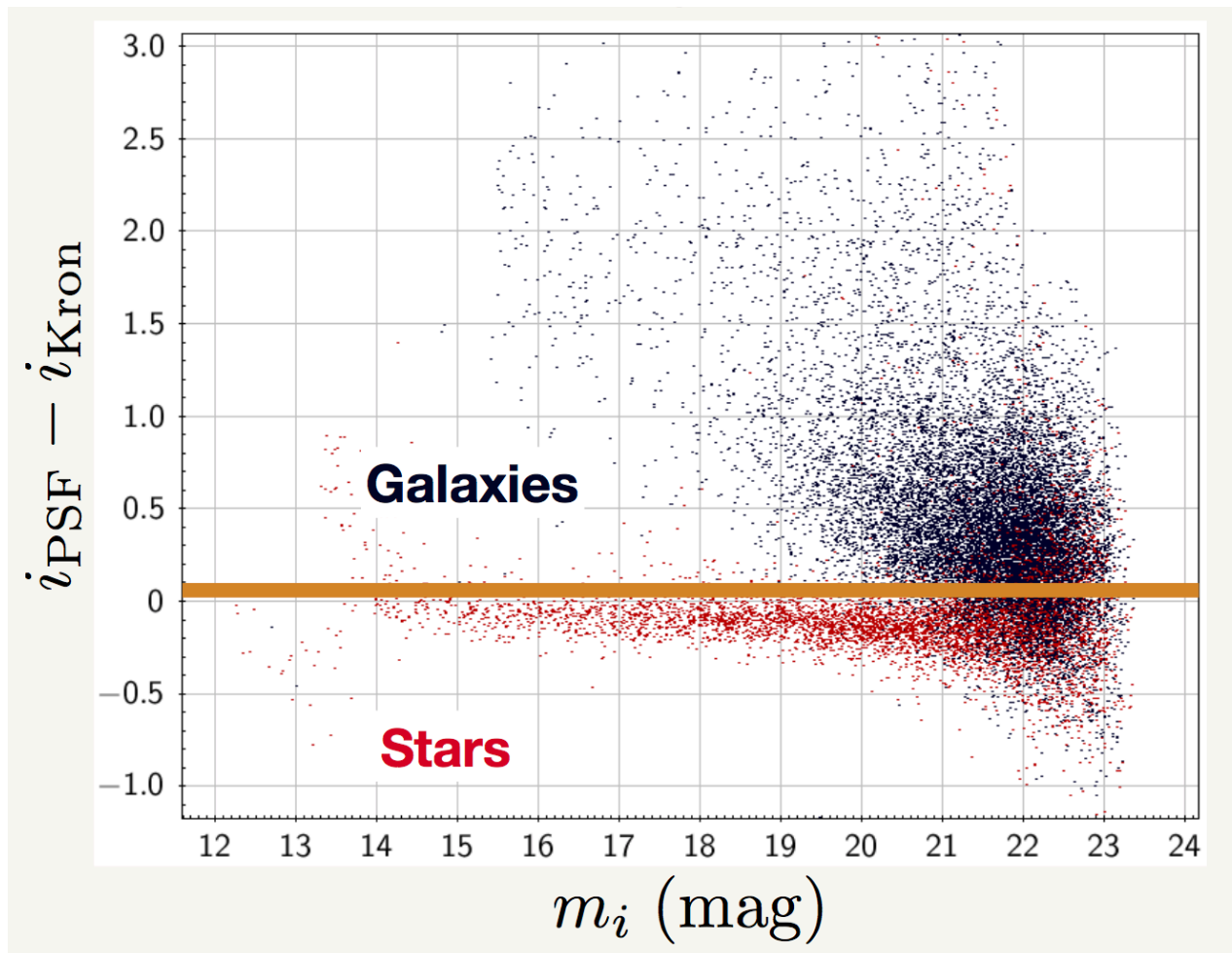


Rolling 30-day window  
light curve

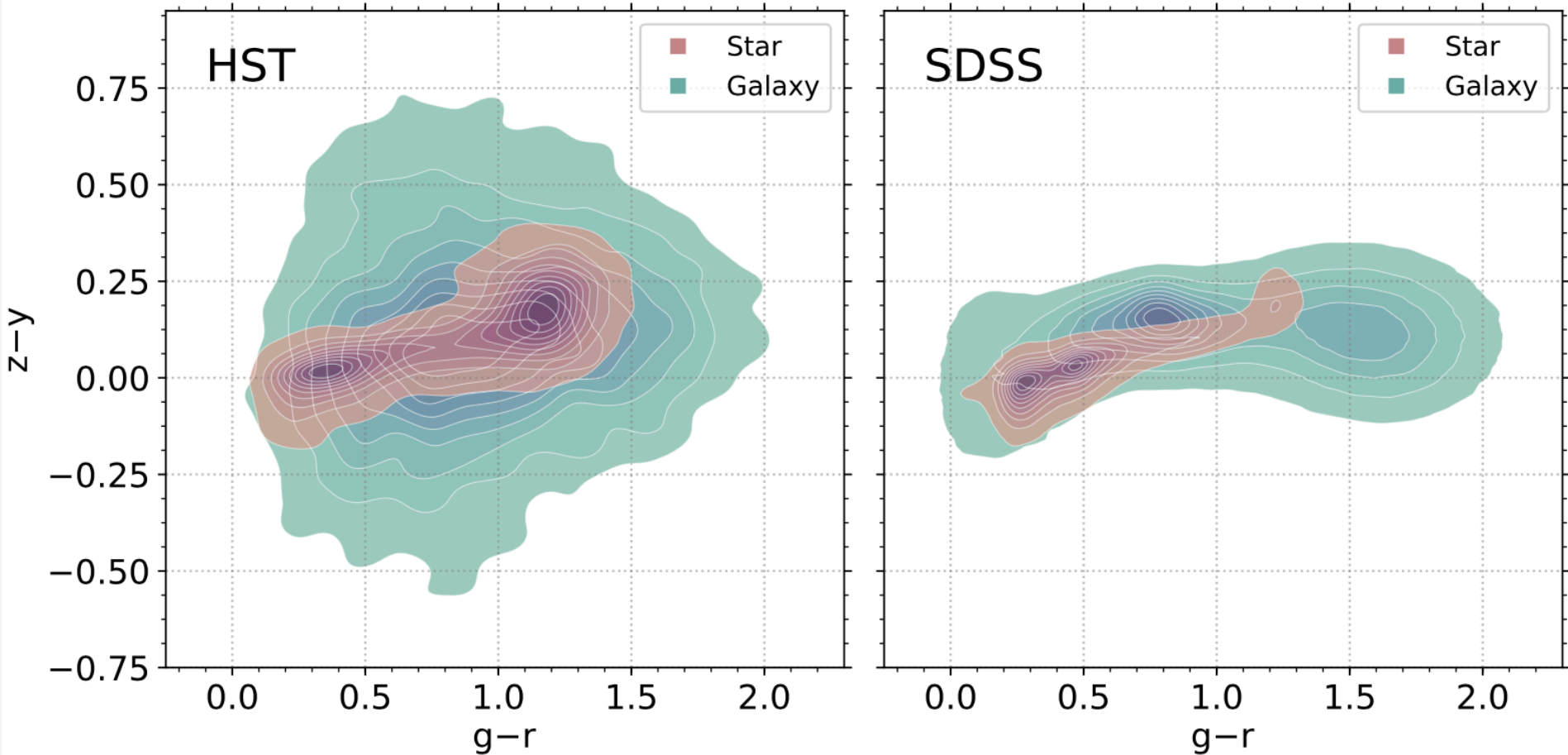
- ZOGY parameters
- Real-bogus score
- Star/galaxy score
- 3 nearest PS1 sources
- Nearest SS object
- Alert history

<https://github.com/ZwickyTransientFacility/ztf-avro-alert>

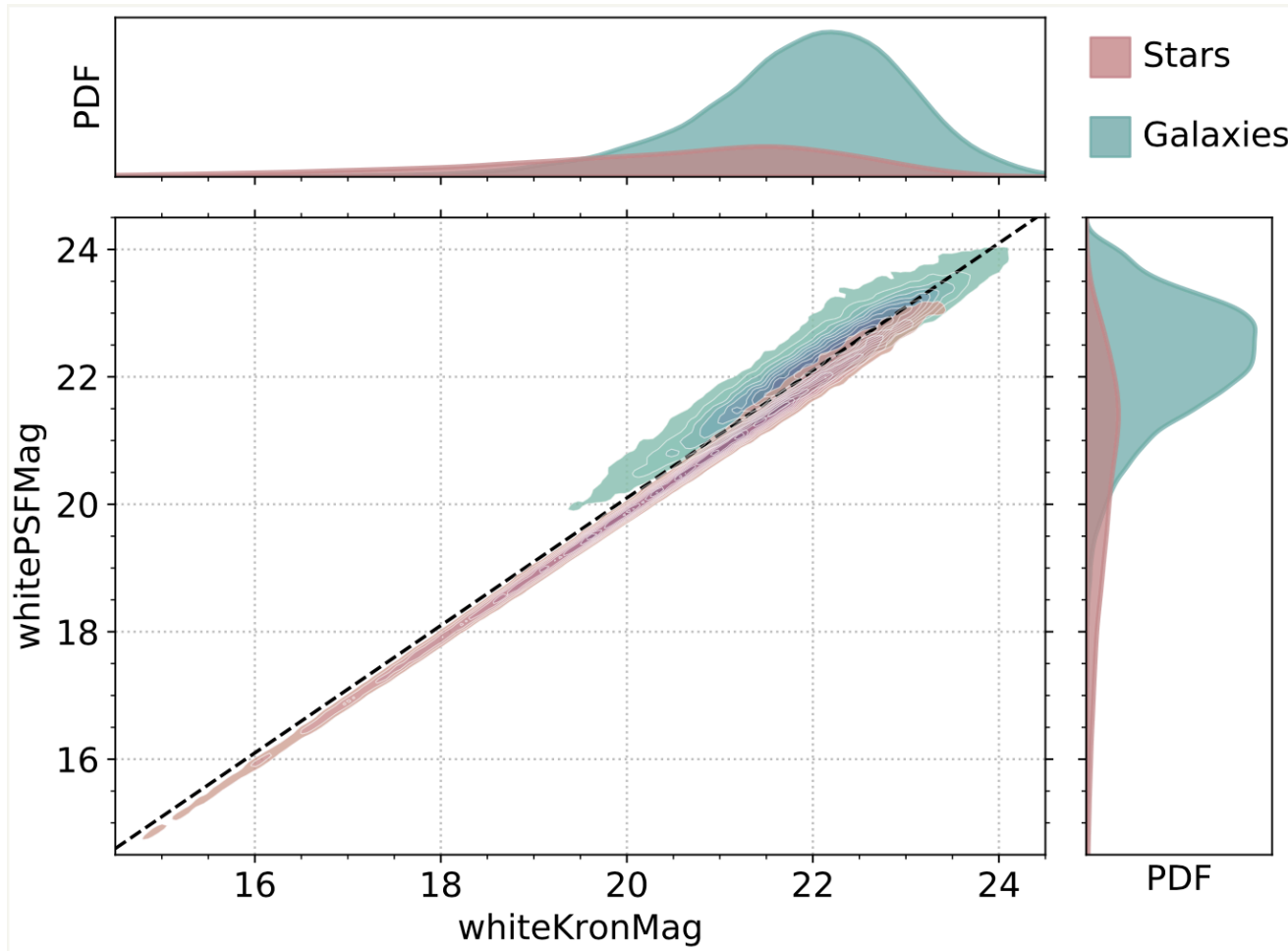
# “Classic” S/G method



# S/G training set



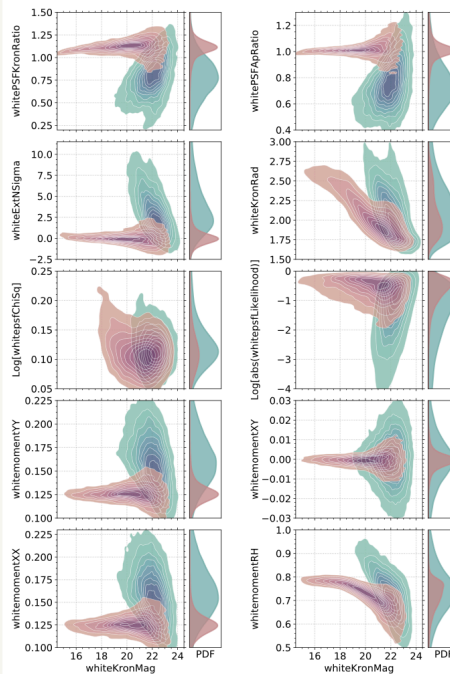
# Simple model



# Creating features

Shape parameters from PS1 stack images

$$\text{white feat} = \frac{\sum_f^{g,r,i,z,y} \text{feat}_f \times \text{SNR}_f^2 \times \delta_f}{\sum_f^{g,r,i,z,y} \text{SNR}_f^2}$$



11 “white” features  
 reduce color dependency  
 no mag measurements

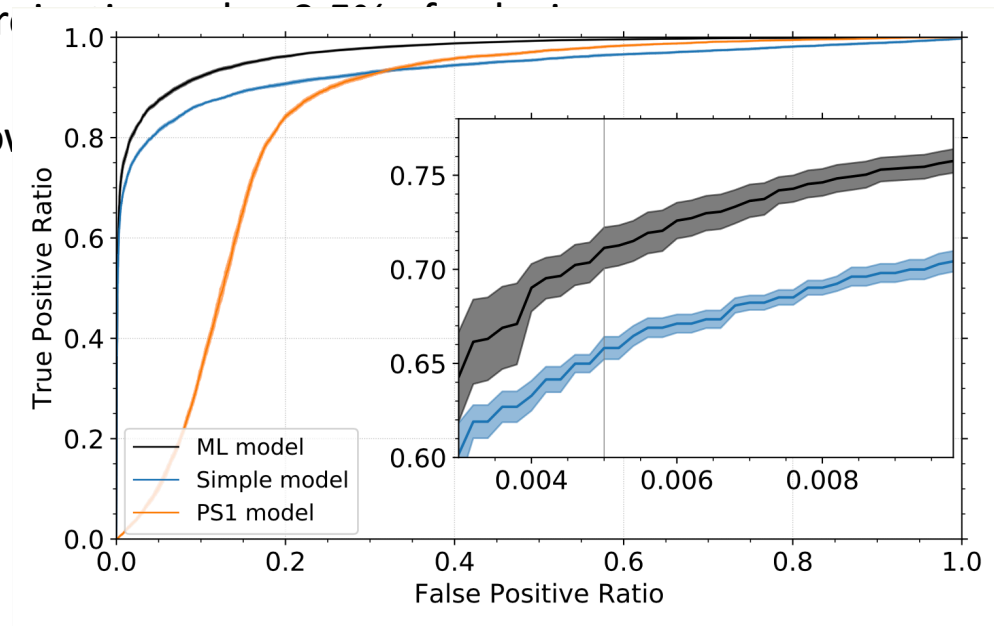
# Results

~1 >> likely star  
 = 0.5 >> absolutely no idea  
 ~0 >> likely galaxy

FPR = 0.005, TPR = 0.7, s/g score = 0.76 (LIGO model optimized)

70% of stars removed while retaining 70% of stars

ML models are superior to low





# Real-bogus: Zooniverse

**Classify**      **Talk**      **zoom++**

ZTF RB\_Project    ABOUT    CLASSIFY    TALK    COLLECT    RECENTS    LAB    ZTF HOME ↗    ZTF IPAC ↗

SCI    REF    DIFF

magpsf:	18.11
sigmapsf:	0.03
classstar:	1.0
sadlstr:	6.0
sgscore:	nan
nmag:	173
nmagname:	[12024]1996YN2
nbnd:	2
rb:	0.91
S/N:	29.49

Is the transient: real or bogus?

- Real
- Bogus
- Skip

Back    Done & Talk    Done

Show the project tutorial

**Metadata**      **Options**      **Tutorial**      **Discuss**

# ZTF Classifier Performance

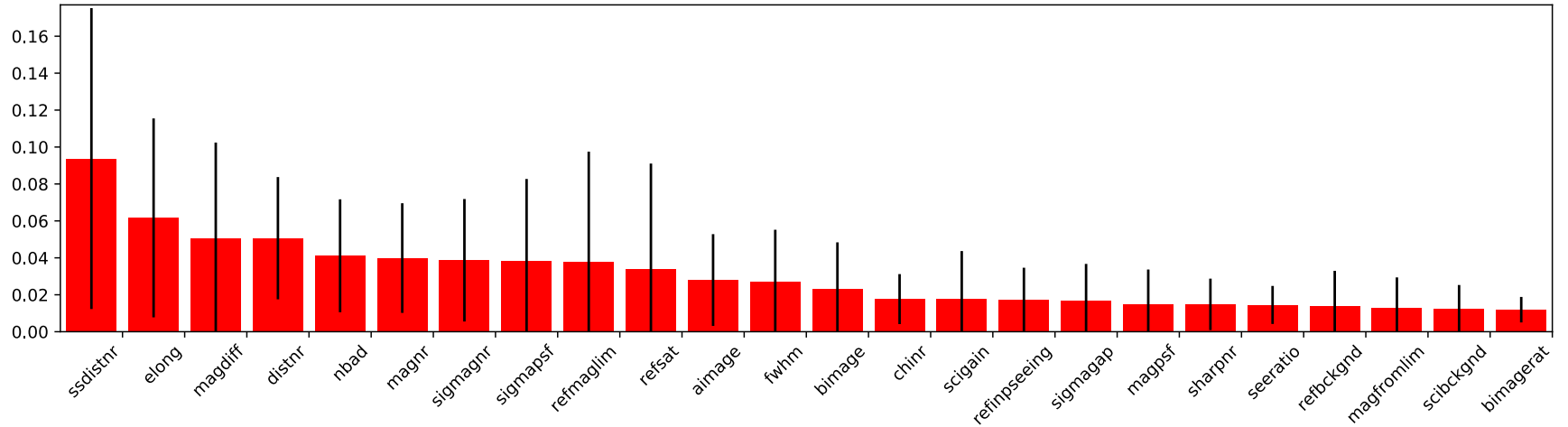
Vers	Date	#Train	# Real	# Bogus	# Feat	FPR	FNR	ACC	FNR at 1% FPR
t1_f1_c1	10 Jan	1620	1316	304	67	30.7	3.8	91.2	36.4
t6_f4_c3	17 Mar	5498	2749	2749	61	11.3	3.8	93.6	27.0
t7_f4_c3	11 April	7767	3361	4406	61	4.8	7.6	94.0	17.1
t8_f5_c3	7 May	14762	5076 (over-sampled to 9686)	9686	59	10.1	2.7	93.6	20.1

f5: removed two features deemed uninformative: infobitsci, infobitsref

Note: I did not balance classes in t7\_f3\_c3, where bogus was the majority class

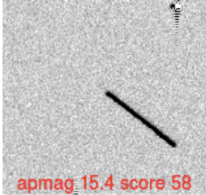
I believe the effect of this is that the all scores are now shifted higher relative to the t7 classifier.

# Feature Importance



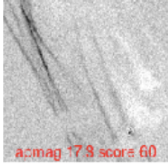
# Streaks

- Based on random forest
- Synthetic data + ZTF images as input



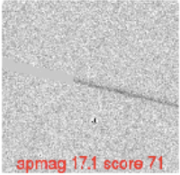
apmag 15.4 score 58

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



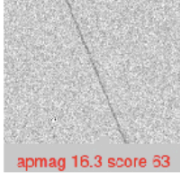
apmag 17.8 score 60

These are ghosts and dementors



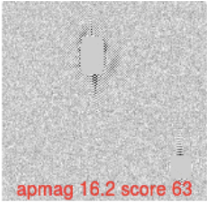
apmag 17.1 score 71

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



apmag 16.3 score 63

Another satellite trail




apmag 16.2 score 63

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](#)

# More research areas

- Deep learning for R/B
- Convert time series to image representation and use CNNs
- Detect transients via CNNs without difference imaging
- Deep learning for streaks
- LSTMs and domain adaptation
- Active learning to minimize follow-up
- Predictive modelling

