

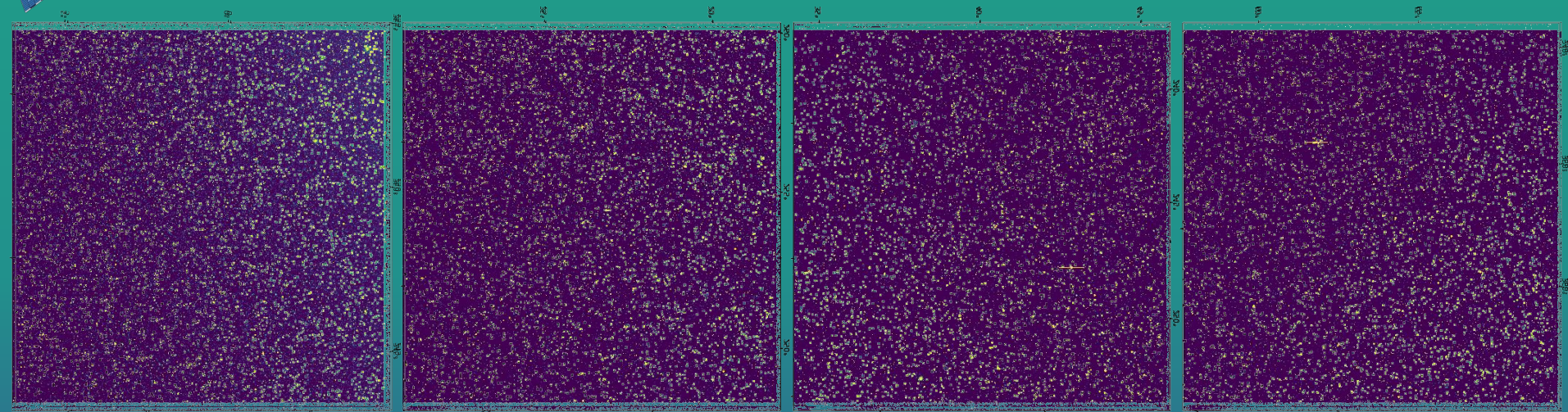
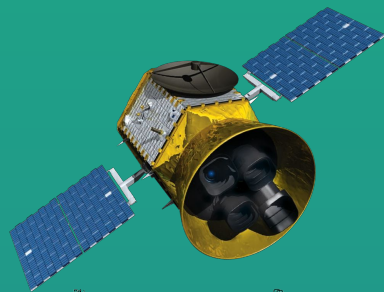
CLASSIFYING TRANSITING EXOPLANET CANDIDATES WITH DEEP LEARNING

Hugh Osborn

CHESS Fellow at University of Bern & MIT

+ Megan Ansdell, Yani Ioannou, Michele Sasdelli,
Jeff Smith, Jon Jenkins, Doug Caldwell, Chedy Raissi, Dan Angerhausen,

THE PROBLEM: FROM RAW DATA TO PLANETS

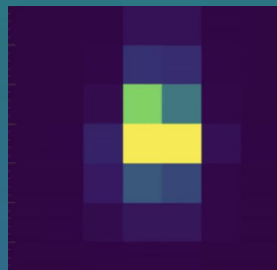


“Postage stamps” for target stars

Typical TESS Raw data

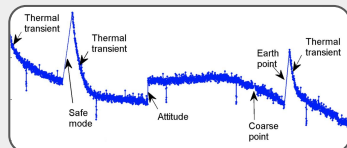
KEPLER & TESS PIPELINES

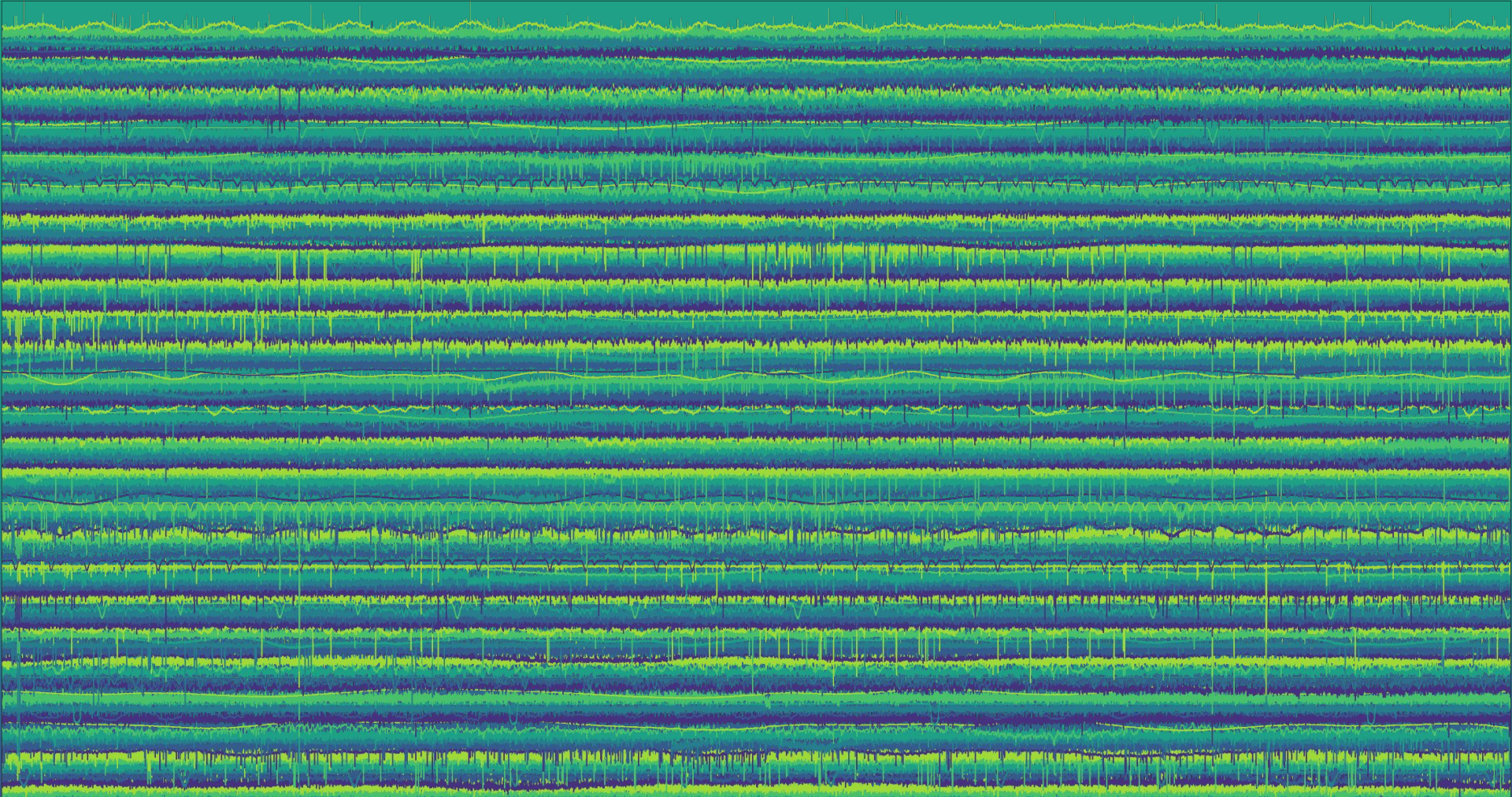
Target Pixel
File (TPF)



Smith+2012, Stumpe+2012

Aperture Photometry & Systematic Correction

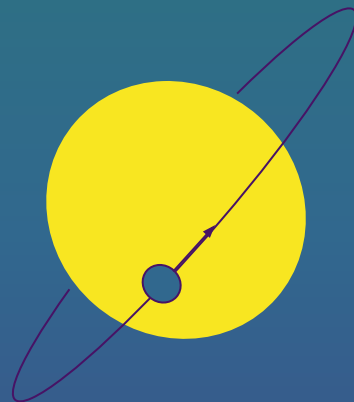
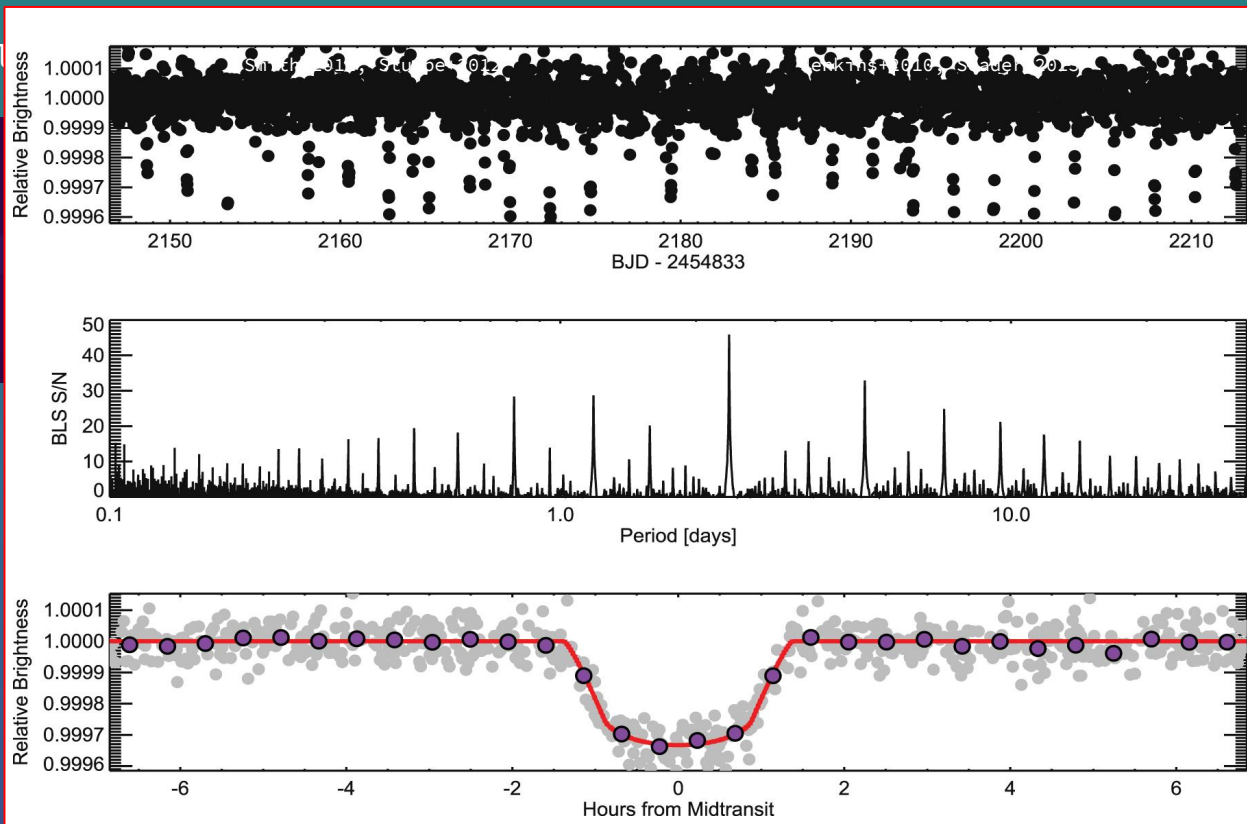
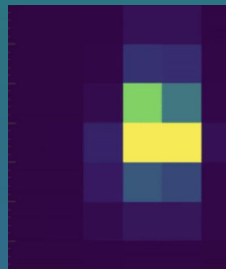




Simulated TESS Data

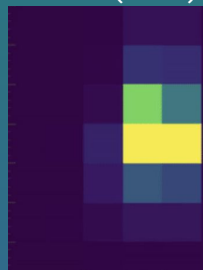
KEPLER & TESS PIPELINES

Target Pixel
File (TPF)



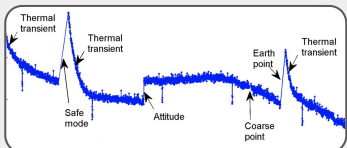
KEPLER & TESS PIPELINES

Target Pixel
File (TPF)



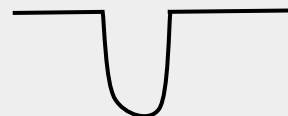
Smith+2012, Stumpe+2012

**Aperture Photometry &
Systematic Correction**



Jenkins+2010, Seader+2013

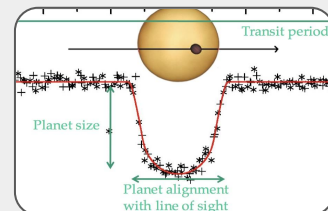
**Transiting Planet
Search (TPS)**



**Threshold Crossing
Event (TCE)**

Wu+2010

Data Validation (DV)



Exoplanet
Catalogues

NASA EXOPLANET ARCHIVE

NASA EXOPLANET SCIENCE INSTITUTE

[Home](#)
[About Us](#)
[Data](#)
[Tools](#)
[Support](#)
[Logos](#)

[Select Columns](#)
[Download Table](#)
[Print Table](#)
[Download Data Files](#)
[View Documentation](#)
[Help/FAQ](#)

	Name	Host	Planet	Distance (light years)	Planet Mass (Earth radii)	Planet Density (g/cm ³)	Planet Age (Gyr)	Planet Type
1	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1b	39.02 ± 0.04	1.25 ± 0.09	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
2	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1c	39.02 ± 0.04	1.28 ± 0.09	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
3	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1d	39.02 ± 0.04	1.34 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
4	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1e	39.02 ± 0.04	1.22 ± 0.09	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
5	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1f	39.02 ± 0.04	1.35 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
6	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1g	39.02 ± 0.04	1.38 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
7	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1h	39.02 ± 0.04	1.40 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
8	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1i	39.02 ± 0.04	1.42 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
9	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1j	39.02 ± 0.04	1.44 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
10	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1k	39.02 ± 0.04	1.46 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
11	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1l	39.02 ± 0.04	1.48 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
12	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1m	39.02 ± 0.04	1.50 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
13	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1n	39.02 ± 0.04	1.52 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
14	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1o	39.02 ± 0.04	1.54 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
15	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1p	39.02 ± 0.04	1.56 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
16	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1q	39.02 ± 0.04	1.58 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
17	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1r	39.02 ± 0.04	1.60 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
18	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1s	39.02 ± 0.04	1.62 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
19	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1t	39.02 ± 0.04	1.64 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
20	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1u	39.02 ± 0.04	1.66 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
21	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1v	39.02 ± 0.04	1.68 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
22	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1w	39.02 ± 0.04	1.70 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
23	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1x	39.02 ± 0.04	1.72 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
24	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1y	39.02 ± 0.04	1.74 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
25	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1z	39.02 ± 0.04	1.76 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
26	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1aa	39.02 ± 0.04	1.78 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
27	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ab	39.02 ± 0.04	1.80 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
28	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ac	39.02 ± 0.04	1.82 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
29	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ad	39.02 ± 0.04	1.84 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
30	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ae	39.02 ± 0.04	1.86 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
31	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1af	39.02 ± 0.04	1.88 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
32	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ag	39.02 ± 0.04	1.90 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
33	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ah	39.02 ± 0.04	1.92 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
34	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ai	39.02 ± 0.04	1.94 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
35	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1aj	39.02 ± 0.04	1.96 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
36	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ak	39.02 ± 0.04	1.98 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
37	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1al	39.02 ± 0.04	2.00 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
38	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1am	39.02 ± 0.04	2.02 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
39	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1an	39.02 ± 0.04	2.04 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
40	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ao	39.02 ± 0.04	2.06 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
41	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ap	39.02 ± 0.04	2.08 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
42	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1aq	39.02 ± 0.04	2.10 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
43	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ar	39.02 ± 0.04	2.12 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
44	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1as	39.02 ± 0.04	2.14 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
45	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1at	39.02 ± 0.04	2.16 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
46	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1au	39.02 ± 0.04	2.18 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
47	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1av	39.02 ± 0.04	2.20 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
48	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1aw	39.02 ± 0.04	2.22 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
49	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ax	39.02 ± 0.04	2.24 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
50	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ay	39.02 ± 0.04	2.26 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
51	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1az	39.02 ± 0.04	2.28 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
52	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ba	39.02 ± 0.04	2.30 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
53	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bb	39.02 ± 0.04	2.32 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
54	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bc	39.02 ± 0.04	2.34 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
55	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bd	39.02 ± 0.04	2.36 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
56	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1be	39.02 ± 0.04	2.38 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
57	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bf	39.02 ± 0.04	2.40 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
58	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bg	39.02 ± 0.04	2.42 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
59	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bh	39.02 ± 0.04	2.44 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
60	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bi	39.02 ± 0.04	2.46 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
61	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bj	39.02 ± 0.04	2.48 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
62	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bk	39.02 ± 0.04	2.50 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
63	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bl	39.02 ± 0.04	2.52 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
64	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bm	39.02 ± 0.04	2.54 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
65	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bn	39.02 ± 0.04	2.56 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
66	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bo	39.02 ± 0.04	2.58 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
67	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bp	39.02 ± 0.04	2.60 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
68	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bq	39.02 ± 0.04	2.62 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
69	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1br	39.02 ± 0.04	2.64 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
70	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bs	39.02 ± 0.04	2.66 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
71	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bt	39.02 ± 0.04	2.68 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
72	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bu	39.02 ± 0.04	2.70 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
73	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bv	39.02 ± 0.04	2.72 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
74	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bw	39.02 ± 0.04	2.74 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
75	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bx	39.02 ± 0.04	2.76 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
76	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1by	39.02 ± 0.04	2.78 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
77	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1bz	39.02 ± 0.04	2.80 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
78	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ca	39.02 ± 0.04	2.82 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
79	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cb	39.02 ± 0.04	2.84 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
80	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cc	39.02 ± 0.04	2.86 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
81	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cd	39.02 ± 0.04	2.88 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
82	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ce	39.02 ± 0.04	2.90 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
83	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cf	39.02 ± 0.04	2.92 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
84	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cg	39.02 ± 0.04	2.94 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
85	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ch	39.02 ± 0.04	2.96 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
86	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ci	39.02 ± 0.04	2.98 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
87	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cj	39.02 ± 0.04	3.00 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
88	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ck	39.02 ± 0.04	3.02 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
89	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cl	39.02 ± 0.04	3.04 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
90	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cm	39.02 ± 0.04	3.06 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
91	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cn	39.02 ± 0.04	3.08 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
92	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1co	39.02 ± 0.04	3.10 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
93	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cp	39.02 ± 0.04	3.12 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
94	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cq	39.02 ± 0.04	3.14 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
95	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cr	39.02 ± 0.04	3.16 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
96	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cs	39.02 ± 0.04	3.18 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
97	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1ct	39.02 ± 0.04	3.20 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
98	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cu	39.02 ± 0.04	3.22 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
99	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cv	39.02 ± 0.04	3.24 ± 0.10	5.3 ± 0.7	3.6 ± 0.4	Super-Earth
100	TRAPPIST-1	TRAPPIST-1	TRAPPIST-1cw	39.0				

MANUAL VETTING



Used for Kepler on all Quarters (later used as labels for machine learning)

Current TESS team: 21 vetters. >200 human hours per sector

MANUAL VETTING



Can a machine do better?

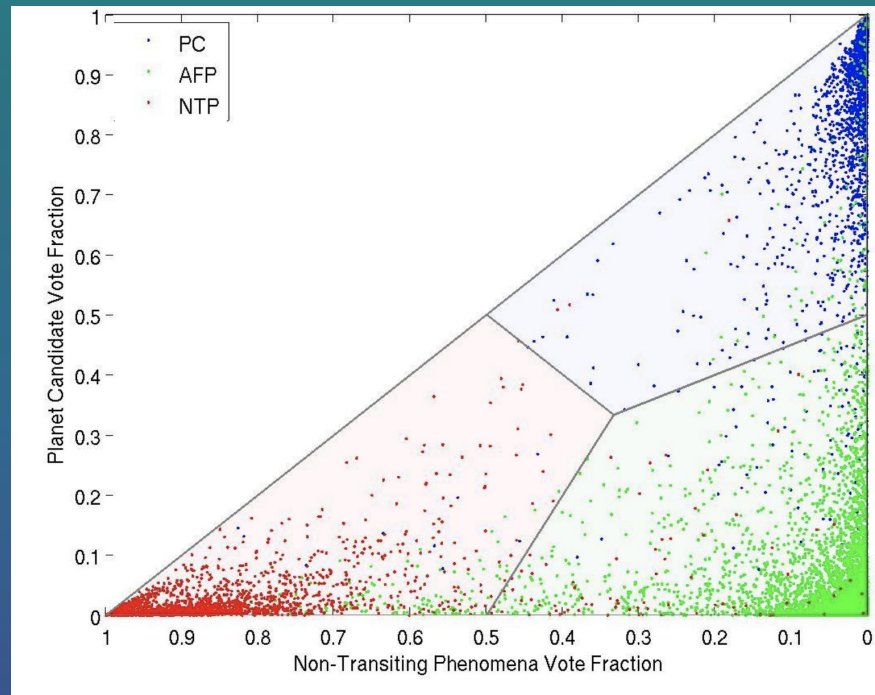
AUTOVETTER - RANDOM FORESTS

The Kepler team also produced a random forest – MacAuliff et al, (2015)

Used 230 features calculated from candidate lightcurve, model fits, etc.

3 output classes: planet, astrophysical dip (e.g. EBs) & non-transiting phenomena

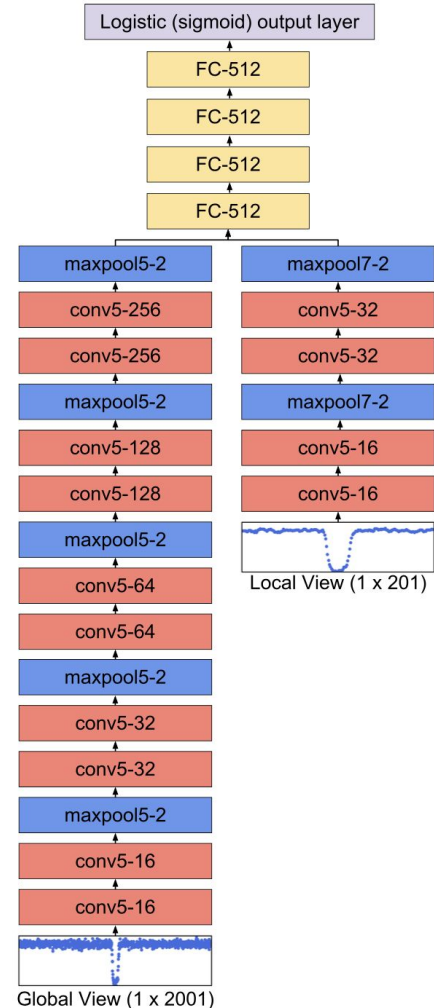
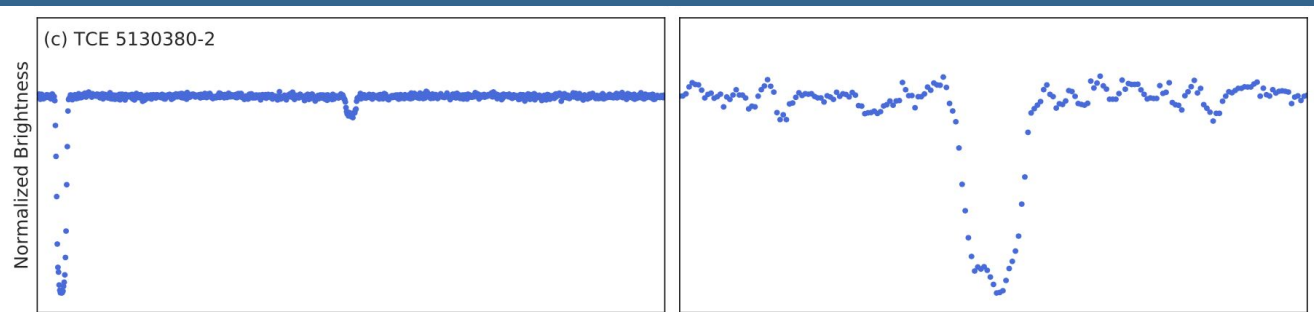
94.15% precision & 97.2% average precision (on human-labelled data)



SHALLUE ET AL 2018

Astronet - Shallue & Vanderburg (2018)

- Deep Convolutional Neural Net
- Inputs are “local” and “global” transit view of each candidate (TCE)
- Two disjoint 1D convolutional columns + 4 fully connected layers
- Output is binary classifier in the range $[0,1]$



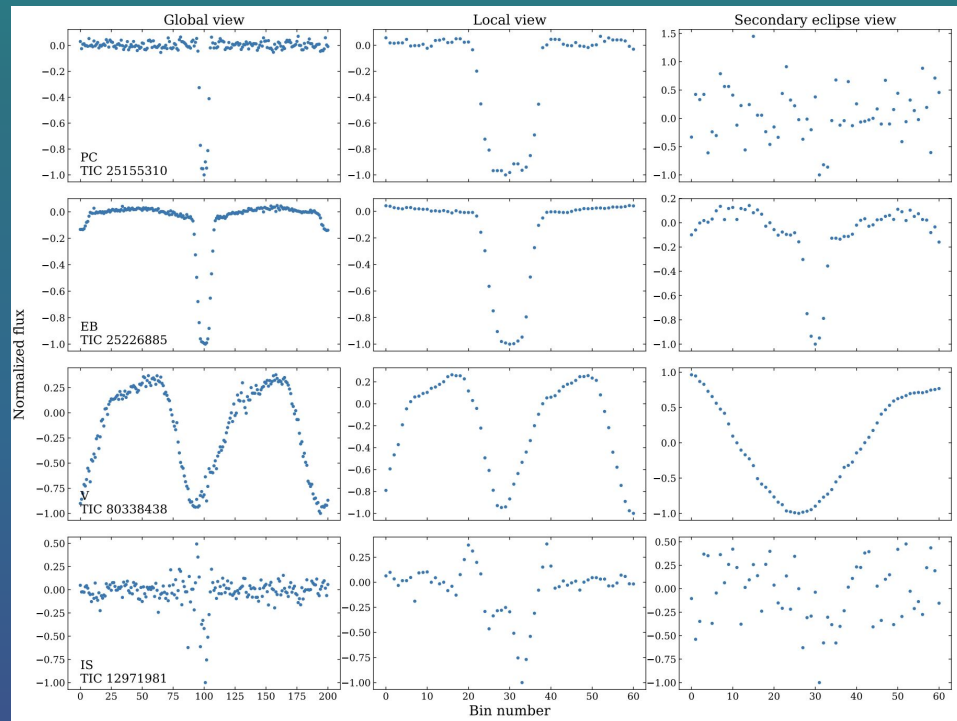
DEVELOPMENTS ON ASTRONET

Application to K2 data
(Dattilo et al 2019).

Application to TESS vetting:
Yu et al (2019)

Included secondary eclipse
region as an input.

Currently used in TESS
vetting at MIT.



2018 NASA FDL-EXOPLANET TEAM

Megan Ansdell
[Exoplaneteer]
UC Berkeley



Hugh Osborn
[Exoplaneteer]
LAM, Marseille

Yani Ioannou
[Deep Learning Expert]
University of Cambridge

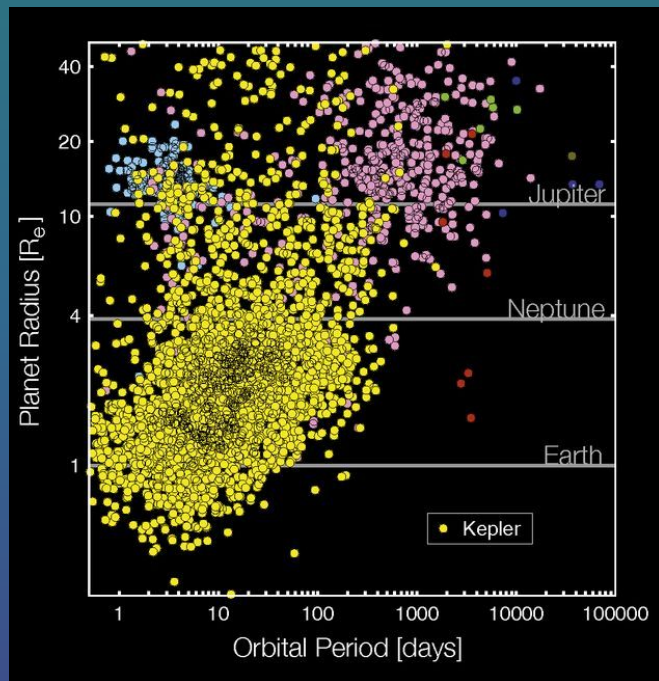
Michele Sasdelli
[Deep Learning Expert]
University of Adelaide

2018 FDL Exoplanet Team Mentors:

- *Science Expertise* → **J. Smith, D. Caldwell, J. Jenkins** (NASA Ames / SETI Institute), **D. Angerhausen** (University of Bern / CSH)
- *Machine Learning* → **C. Raissi** (INRIA), Yarin Gal (Oxford)
- *Compute Power* → **M. Mascaro** (Google Cloud)

KEPLER INPUT DATA & LABELS

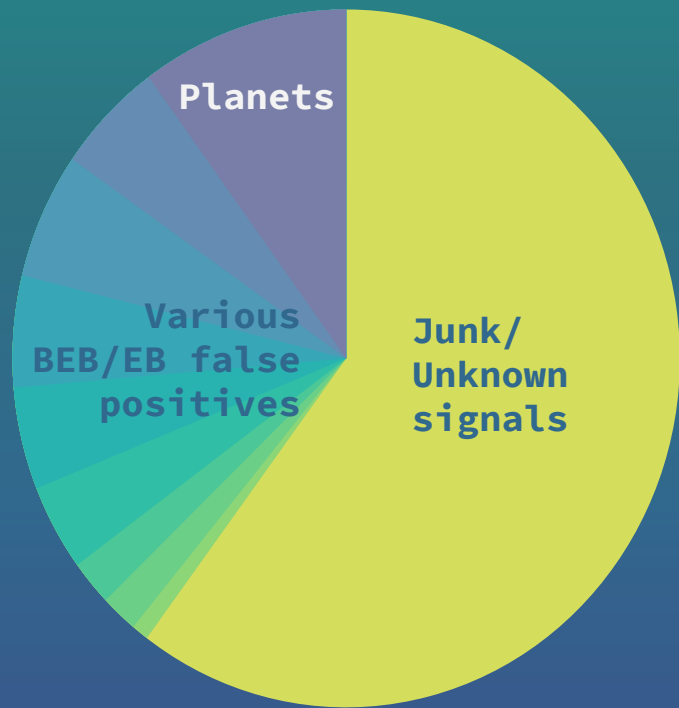
- 16,000 Threshold Crossing Events (TCEs) from Kepler DR24
- Labelled by human vetters
- ~25% planets & ~75% false positives
- Preprocessed the data following Shallue & Vanderburg:
 - Detrended lightcurve
 - Phase-folded onto TCE period
 - Binned to global & local view



Ansdeell, Ioannou, Osborn,
Saselli, et al. (2018)

TESS INPUT DATA & LABELS

- 4 Simulated sectors - i.e. we know the exact ground-truth
- Pixel-level signal injection, processed with the TESS pipeline
- ~16,000 candidates, ~14% planets
- Preprocessed the data following Shallue & Vanderburg



ASSESSING MODEL PERFORMANCE

Definitions

Ground Truth _{PL}	UNK		False Positives	
	PL	False Negatives	True Positives	False Negatives
	EBs		False Positives	
		UNK	PL Predicted	EBs

ASSESSING MODEL PERFORMANCE

Precision

* also known as accuracy

Ground Truth	UNK			
	PL			
	EBs			
		UNK	PL	EBs

Predicted

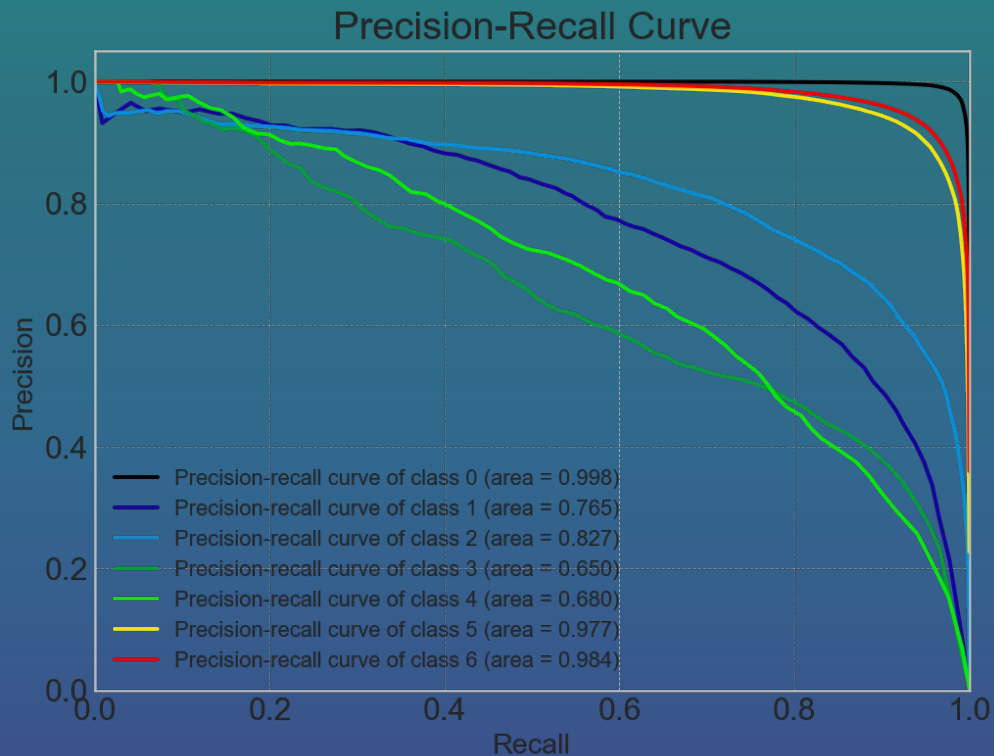
ASSESSING MODEL PERFORMANCE

Recall

UNK			
Ground Truth _{PL}			
EBs			
	UNK	PL	EBs
	Predicted		

ASSESSING MODEL PERFORMANCE

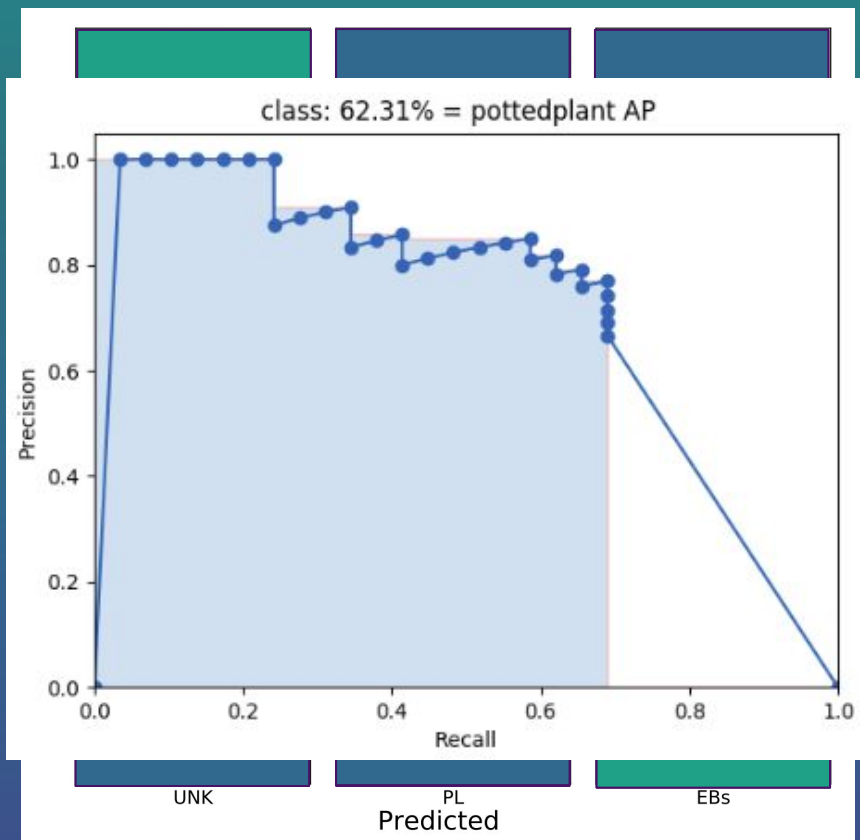
Precision-Recall Curve



ASSESSING MODEL PERFORMANCE

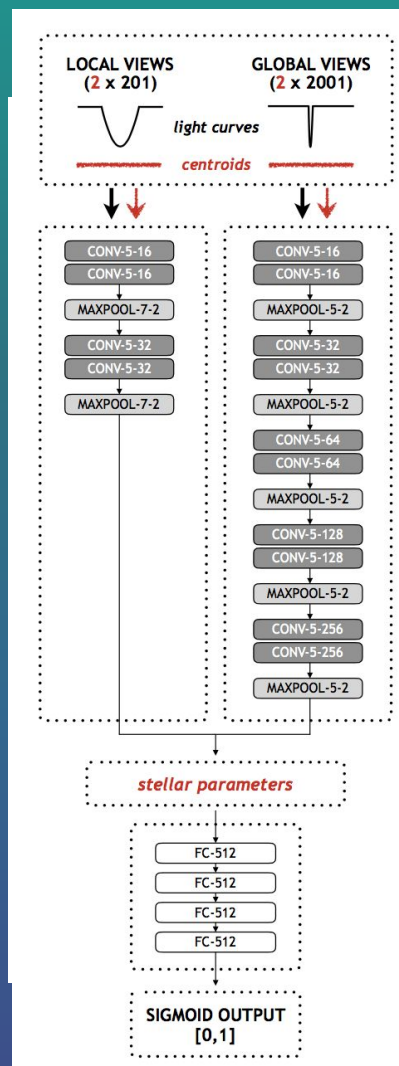
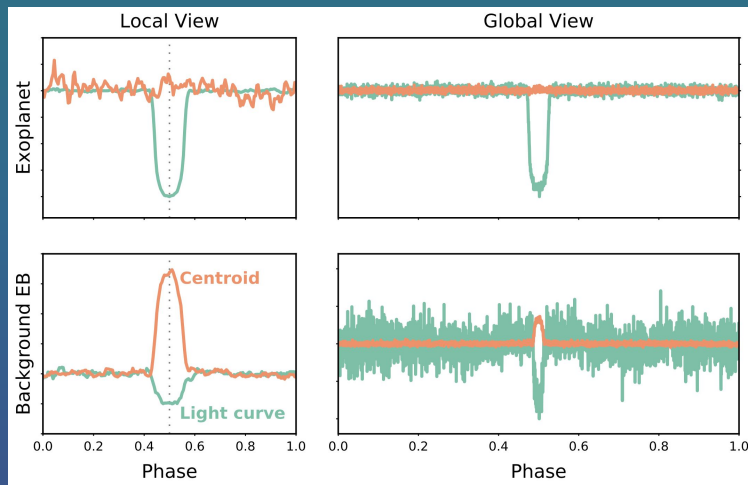
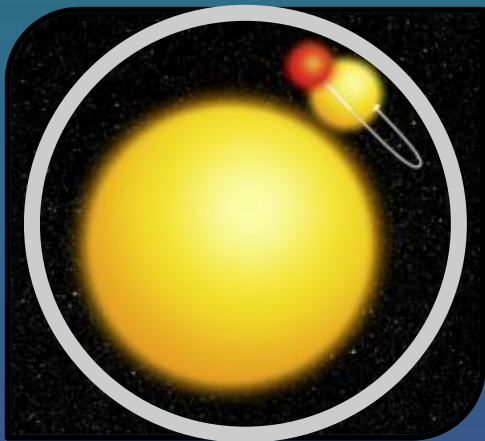
Average Precision

- Weighted average of precision for all classes.
- Functionally similar to Area Under Curve (AUC) for a multi-class classifier - i.e. probability a random positive sample is correctly predicted at any P-R threshold



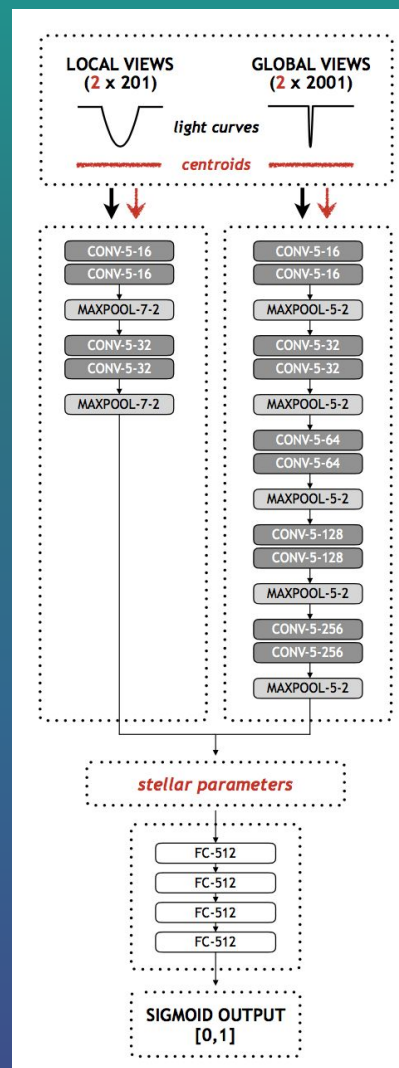
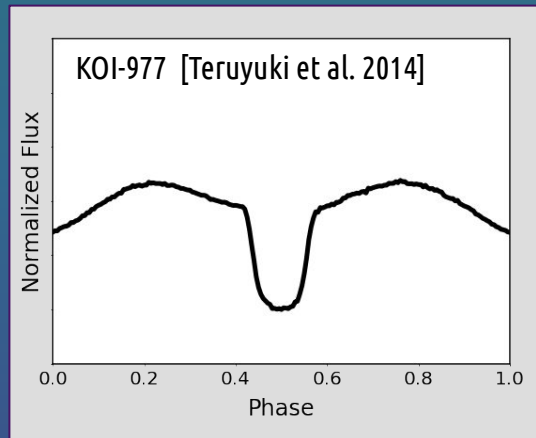
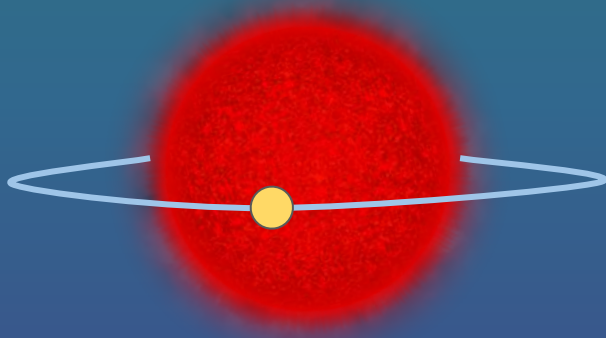
DOMAIN KNOWLEDGE - CENTROIDS

- Position of centre of light over time
- Important for identifying background EBs



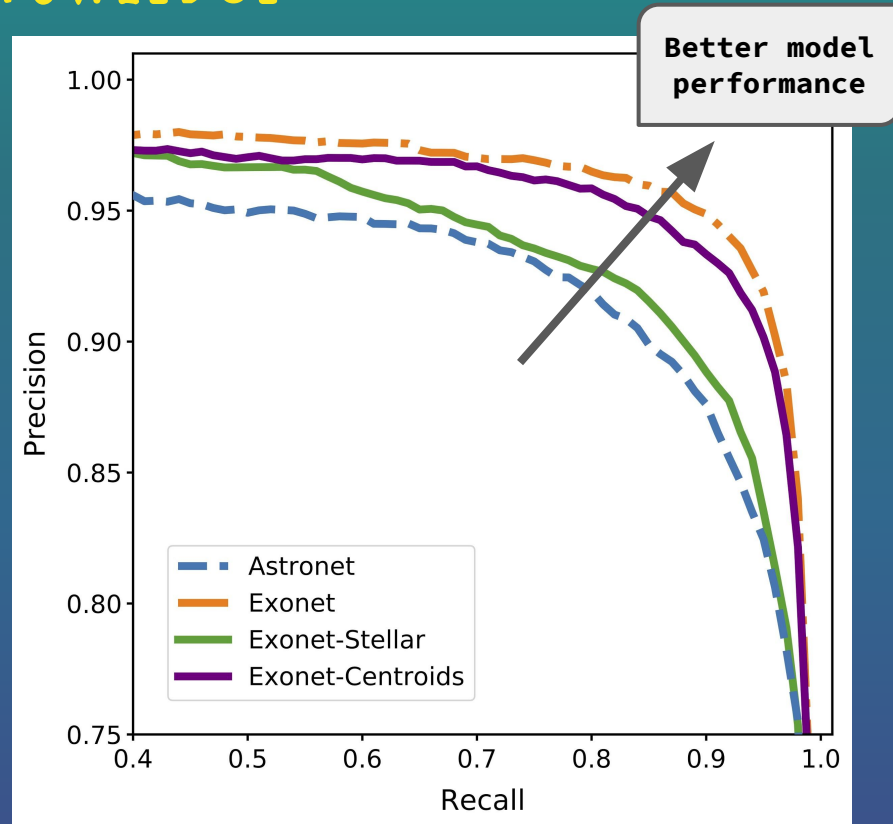
DOMAIN KNOWLEDGE - STELLAR PROPERTIES

- From stellar properties catalog: mass, radius, density, $\log g$, metallicity
- Important for identifying, e.g., giant star binaries



PERFORMANCE WITH DOMAIN KNOWLEDGE

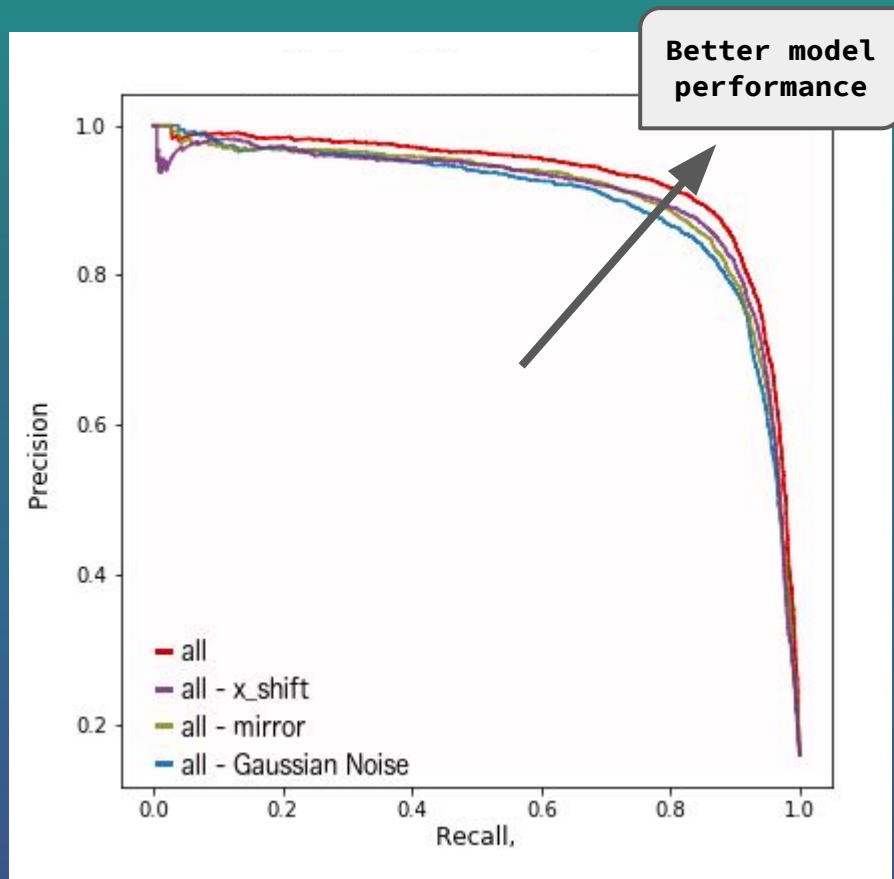
- Centroids & Stellar info both improve performance
- Also helped by cross validation & model ensembling



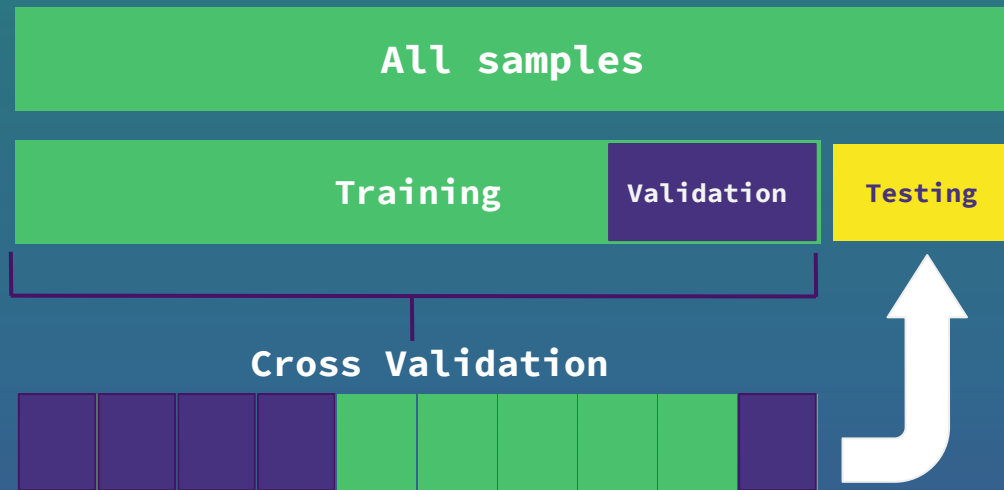
DATA AUGMENTATION

- Modify input data to create “new” data for the neural net, preventing overfitting

	Avg. Precision
Exonet: no augmentation	85.2%
Exonet - Gaussian	89.6%
Exonet - xmirror	90.4%
Exonet - xshift	90.5%
Exonet - all	92.7%



ENSEMBLING & CROSS-VALIDATION



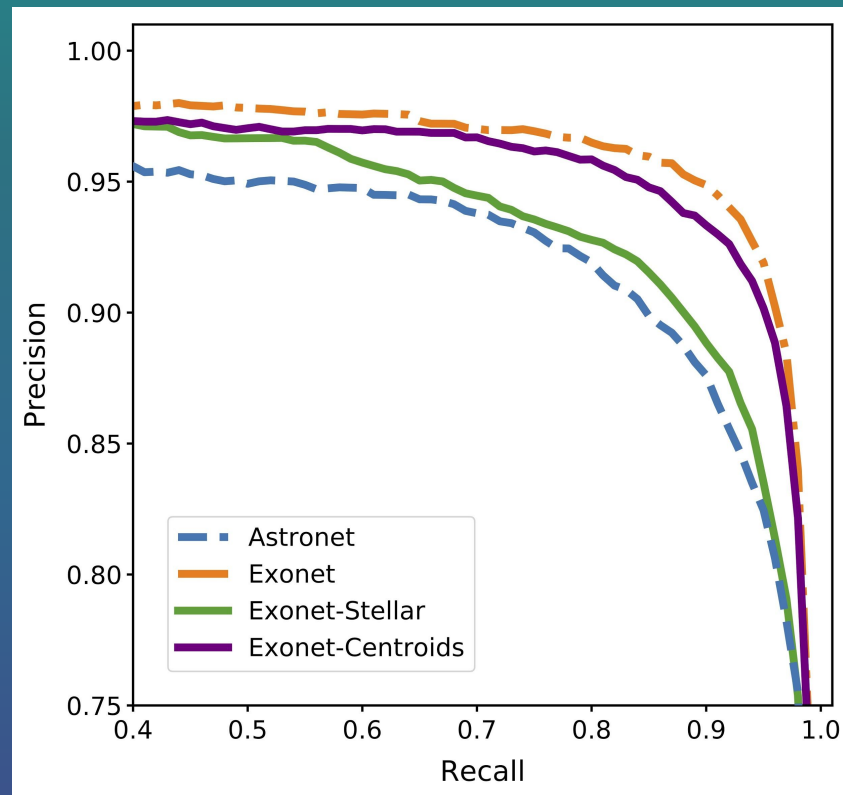
Ensembling / “bagging”:
Taking average of models
applied to test data

*always need test set

Multiple Validation sets = multiple trained models

KEPLER PERFORMANCE

- Thanks to domain knowledge, augmentation, ensembling, etc - Exonet-Kepler improves on Astronet, and is the **best classifier** of Kepler candidates yet.



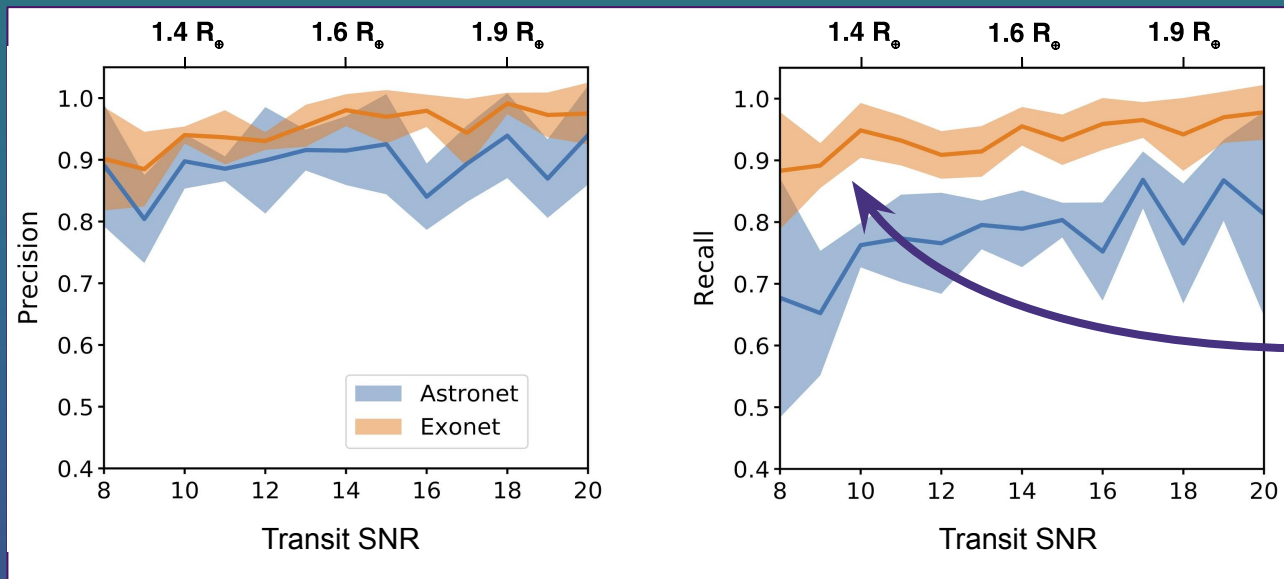
KEPLER PERFORMANCE

- Thanks to domain knowledge, augmentation, ensembling, etc - Exonet-Kepler improves on Astronet, and is the **best classifier** of Kepler candidates yet.

	Planet Precision	Avg. Precision
Autovetter	94.15%	97.19%
Astronet	95.8%	95.5%
Exonet	97.5%	98.0%

KEPLER PERFORMANCE

Improved Performance for Lowest SNR Transits

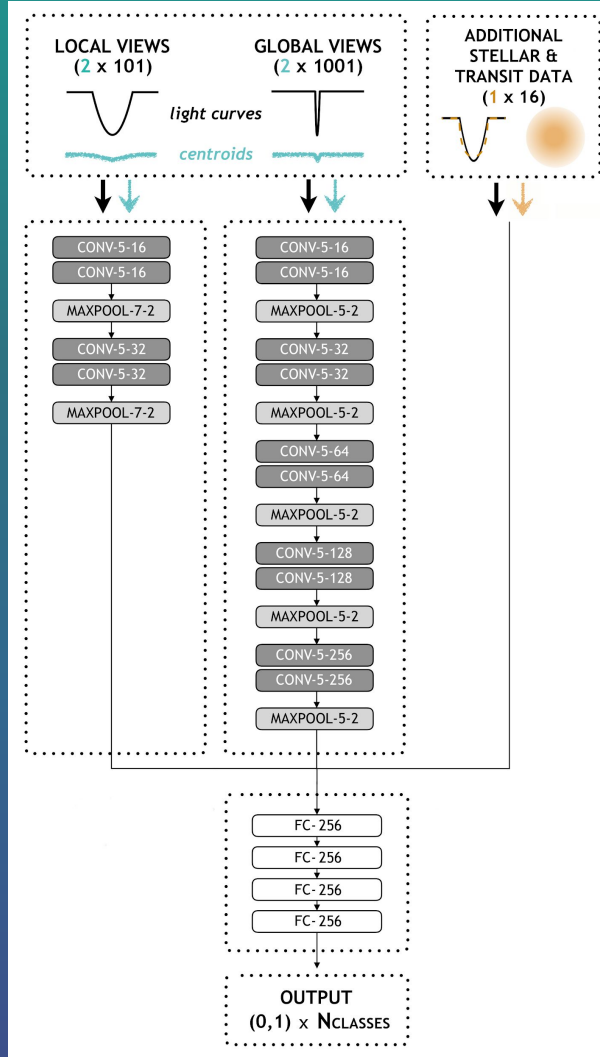


15-20% gains in recall for Earth-sized planets

CLASSIFYING TESS DATA

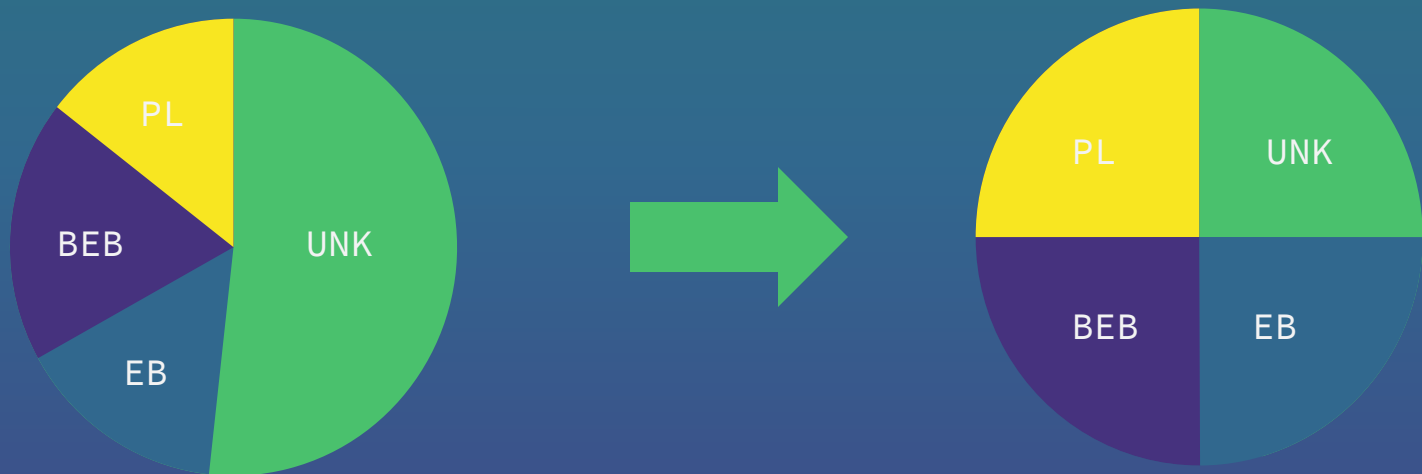
Slightly modified from Kepler -> TESS

- Added additional transit-derived information
- Reduced bins from 2001 to 1001
- Used multi-class modelling



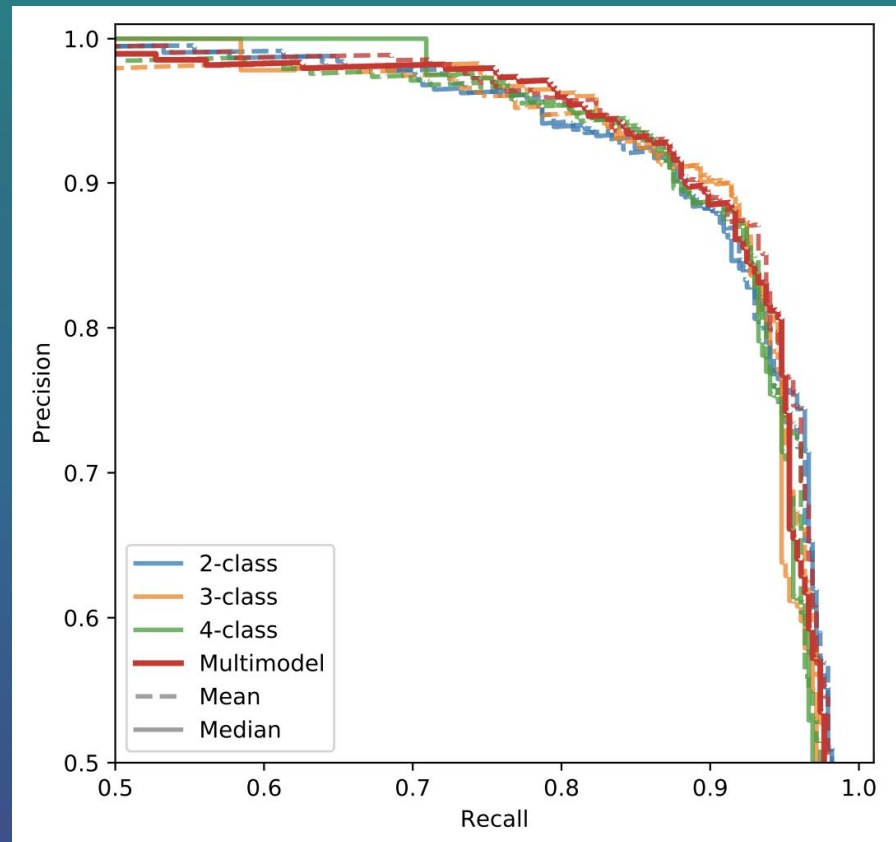
BALANCED BATCH SAMPLING

- Models tend to predict the majority class in unbalanced data
- Re-balancing means that each epoch sees same number of samples from each – helps training

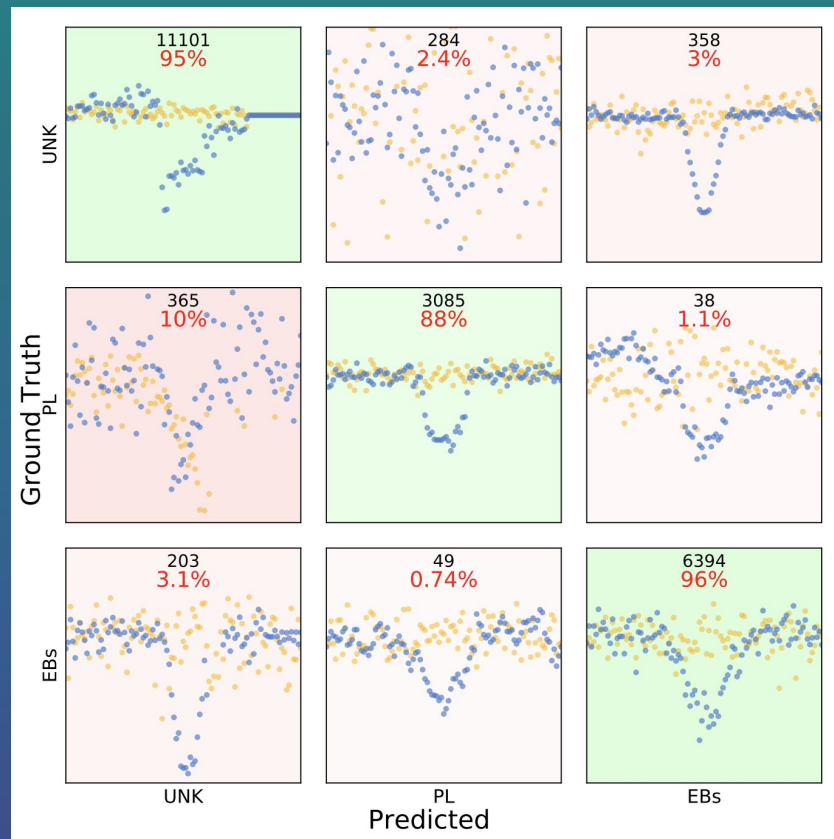
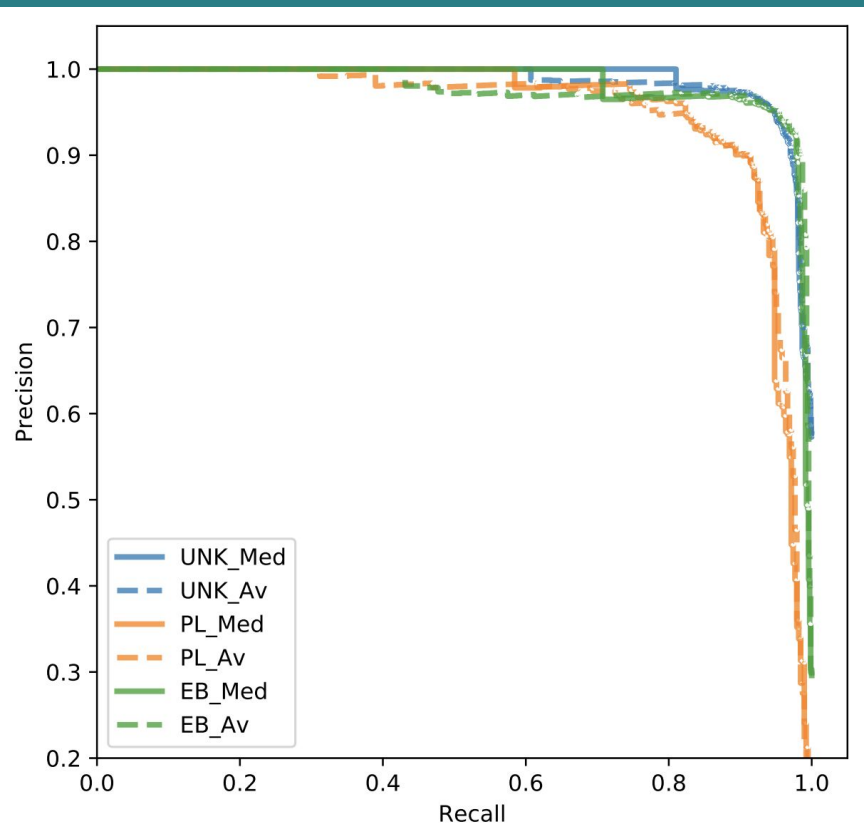


PERFORMANCE ON TESS SIMULATIONS

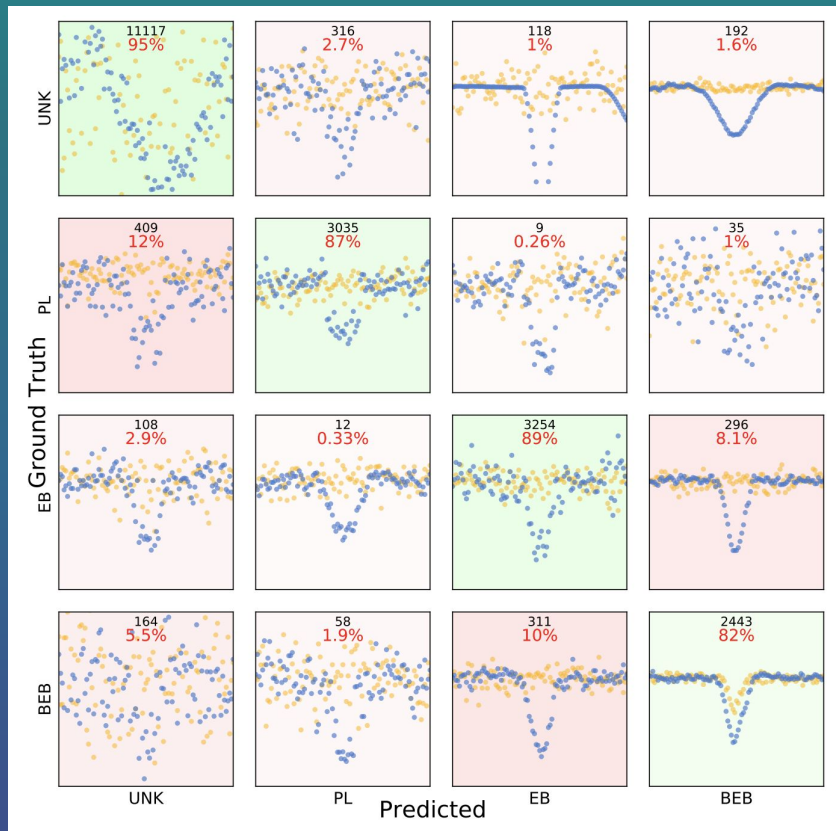
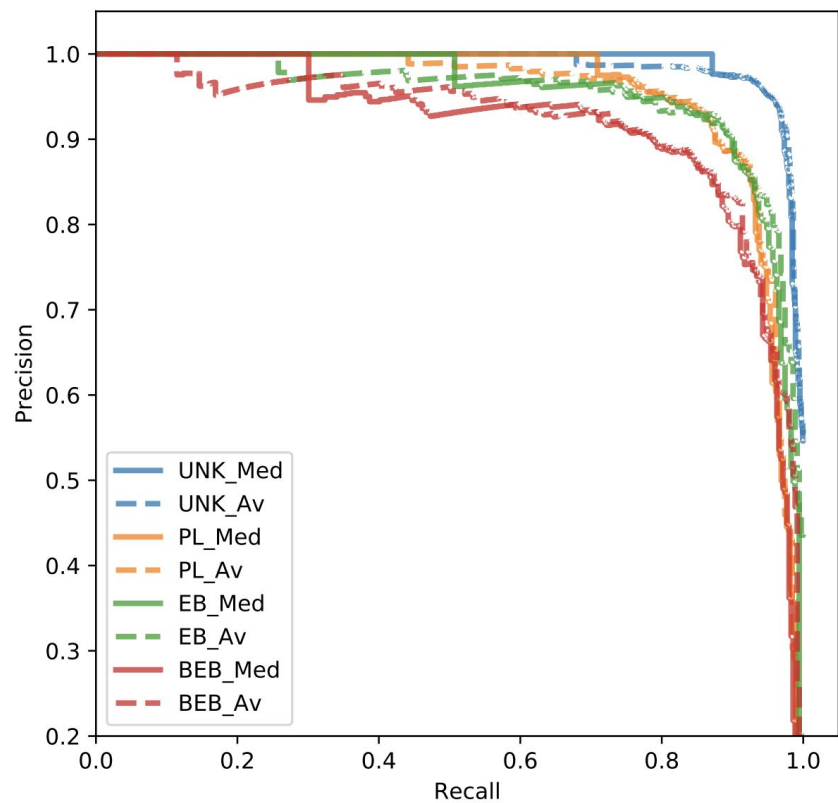
		Planet Precision	Planet Recall	Av. Precision
Binary	Planet	91.8	87.8	95.2
	Not Planet	97.6	98.5	99.4
3-class	Planets	<u>90.4</u>	<u>90.1</u>	<u>95.6</u>
	EBs	95.1	95.1	96.9
	Unknown	94.8	94.9	97.7
4-class	Planets	89.1	88.8	94.4
	EBs	87.4	91.7	94.7
	BEBs	88.5	81.7	91.7
	Unknown	94.6	95.5	97.8



3-CLASS MODEL

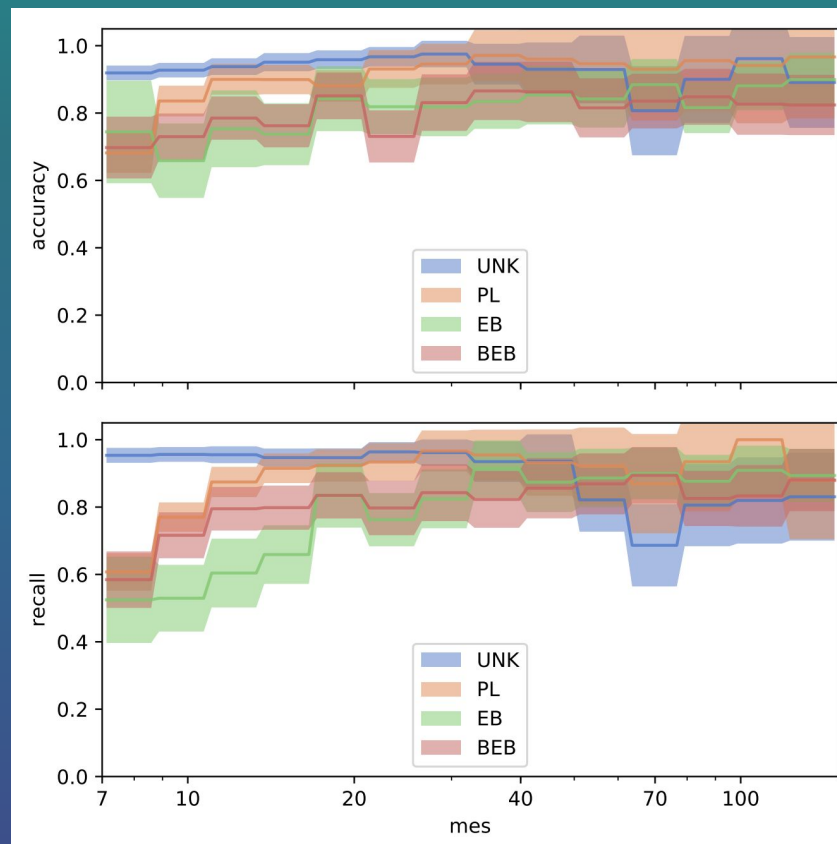


4-CLASS MODEL



PERFORMANCE AS A FUNCTION OF SNR

- Recall deteriorates at low SNR
- 70% precision/accuracy in $7 < \text{SNR} < 8.5$ range
- “Unknown” consistently accurate – model has learnt systematic features

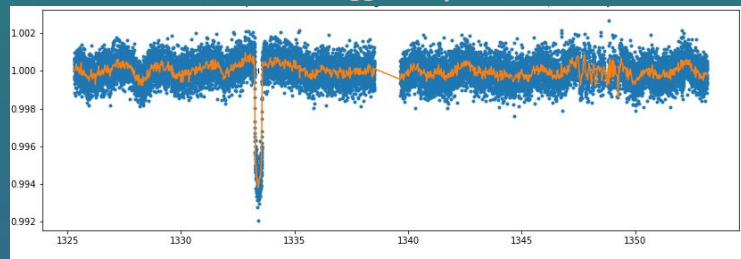


COMPARISON WITH ANSDELL ET AL, 2019

	Average Precision on planets
Kepler	98.5%
TESS	95.6%

← Why?

A Monotransit flagged as periodic in real TESS data.



- Labels: Human vetting vs. Simulated ground truth
- Minimum transits: Kepler ≥ 3 - vs - TESS ≥ 2
- “Near misses” - 196 “false positives” are planets
 - 44% from monotransits
 - 25% from period confusion
- Including “near misses” - planet precision from 90.3% to 95.1%

APPLICATION TO REAL TESS DATA

Far faster than other TESS vetting methods!

- ~60 minutes to pre-process lightcurves
- 5 minutes to predict with trained model on one GPU

But real data \neq simulated data

- Different noise characteristics
- Do injections match reality?
- No “ground truth” to make comparisons

TESS has 2 candidate pipelines producing candidates. Overlap is not perfect.

APPLICATION TO REAL TESS DATA

All TOIs in Sectors 1-5

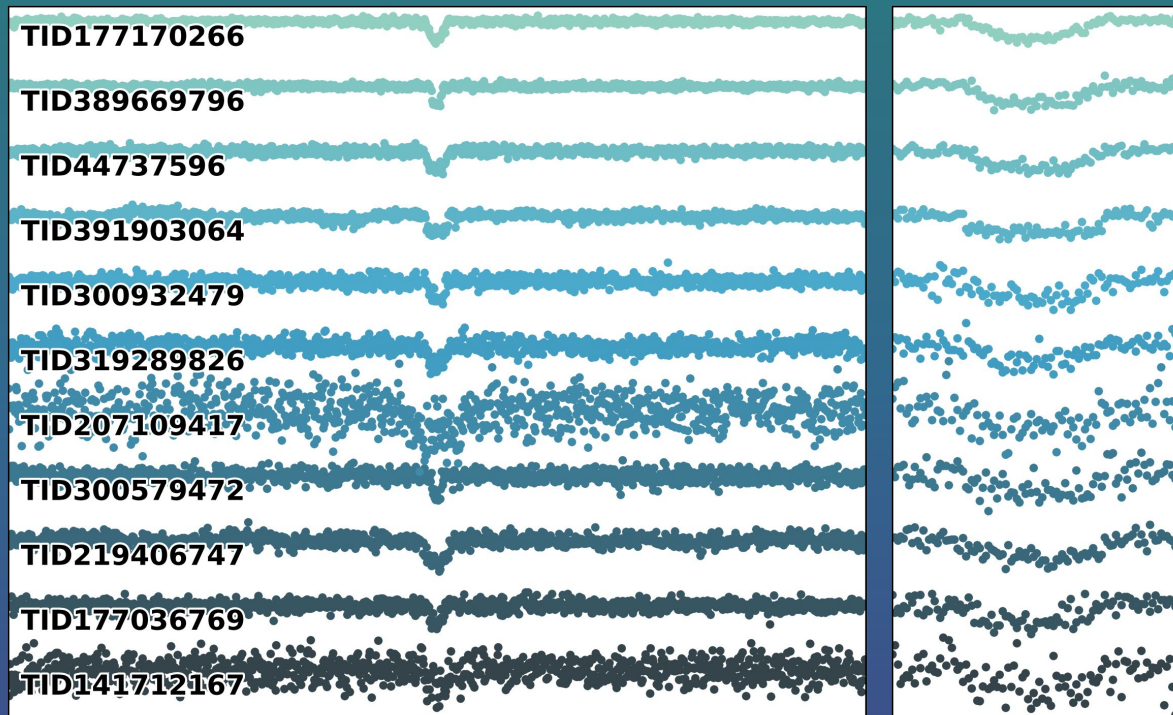


NEW PREDICTED PLANETS

>100 new candidates from model predictions

Problems:

- Many giant binaries in predicted sample
- Some targets share the same period & epoch - reflections from a bright binary



CONCLUSION

- Machine Learning using “domain knowledge” enables fast & more accurate classification of transiting planet candidate vetting.
- Kepler-ExoNet is the best-performing model yet tested, with a precision on Kepler candidates of 97.5%
- TESS-ExoNet also performs well, achieving 90–95% precision on simulated training set.
- However, models trained on simulations do **not** perform as well on real data!
- We have identified promising new candidates missed by manual vetters.

THANKS!
ANY QUESTIONS?

Hugh Osborn

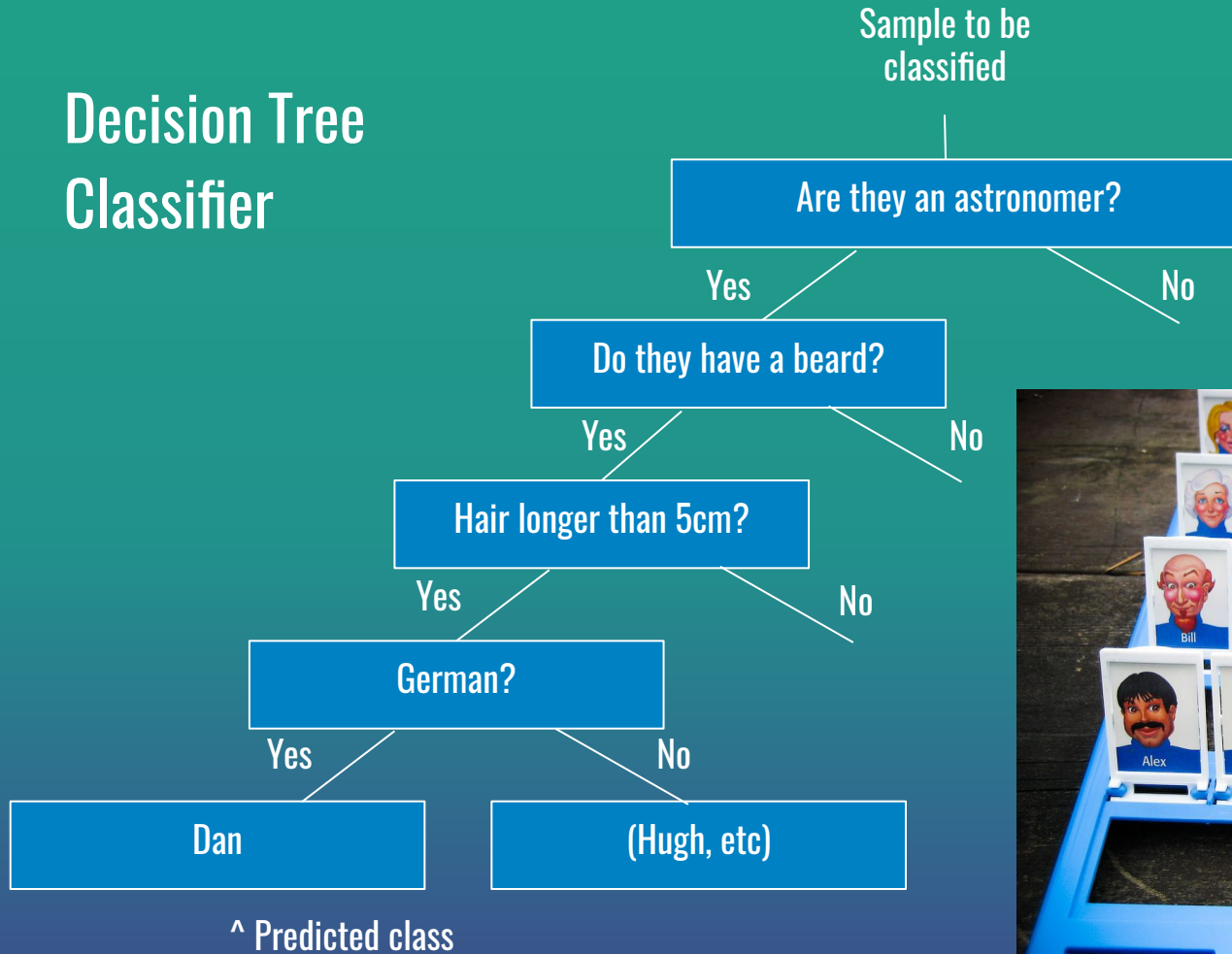
CLASSIFICATION WITH MACHINE LEARNING

Classes >

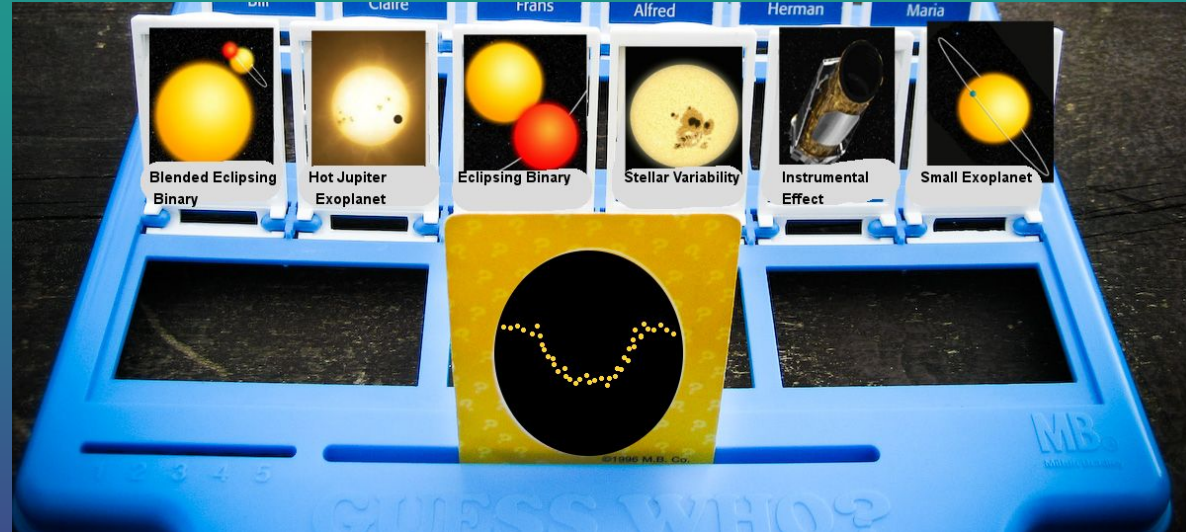
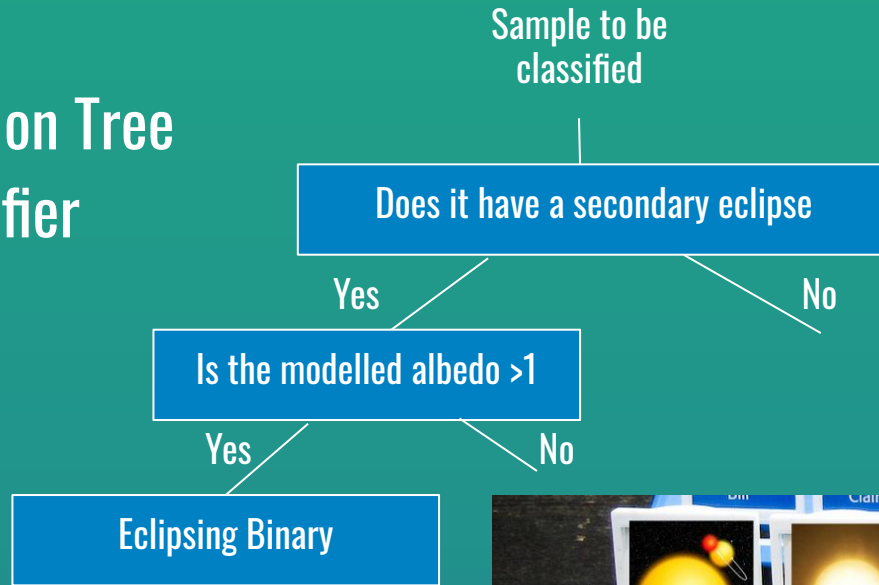
Samples ^

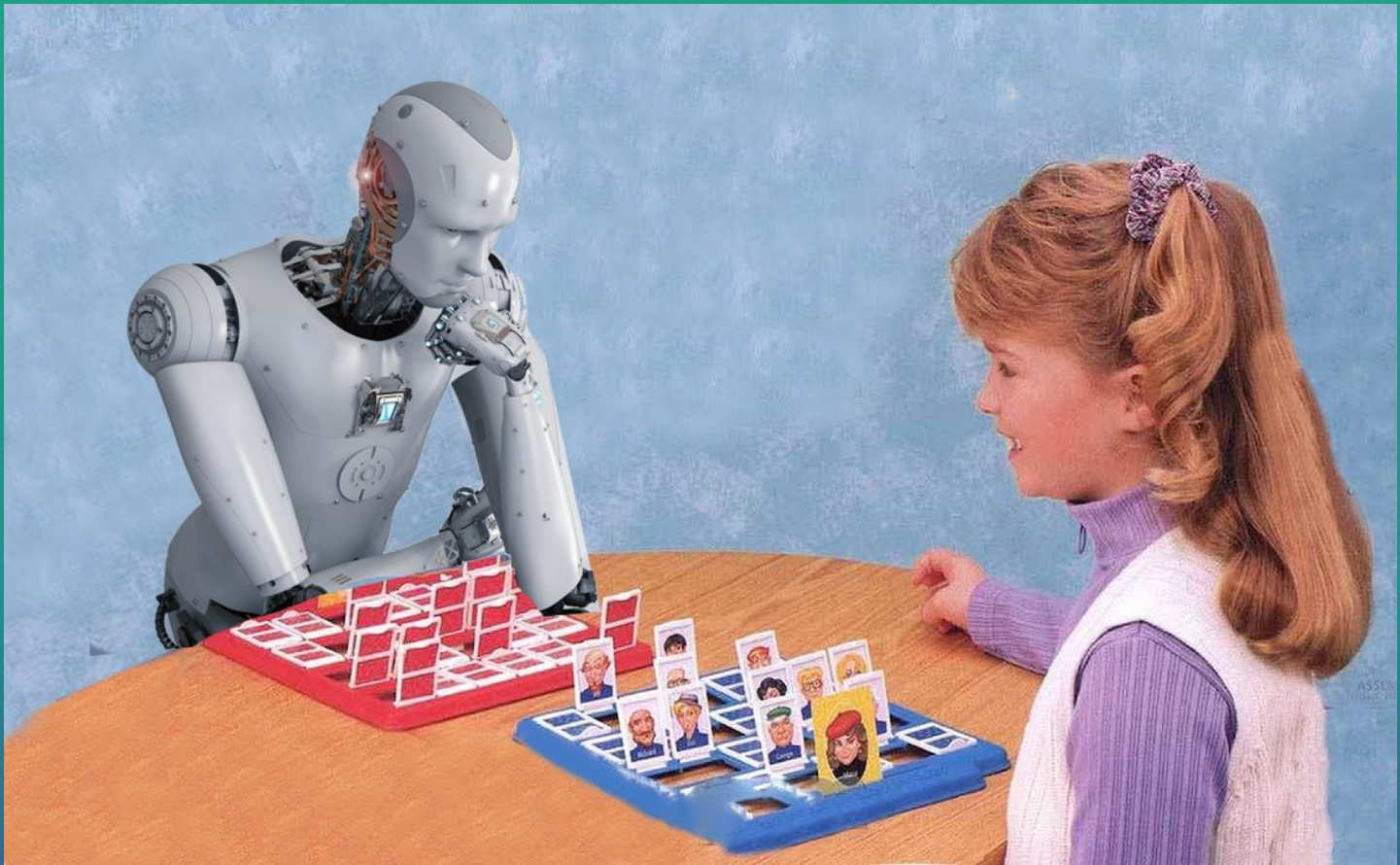


Decision Tree Classifier



Decision Tree Classifier





How can we classify with minimal human processing?
With Machine Learning

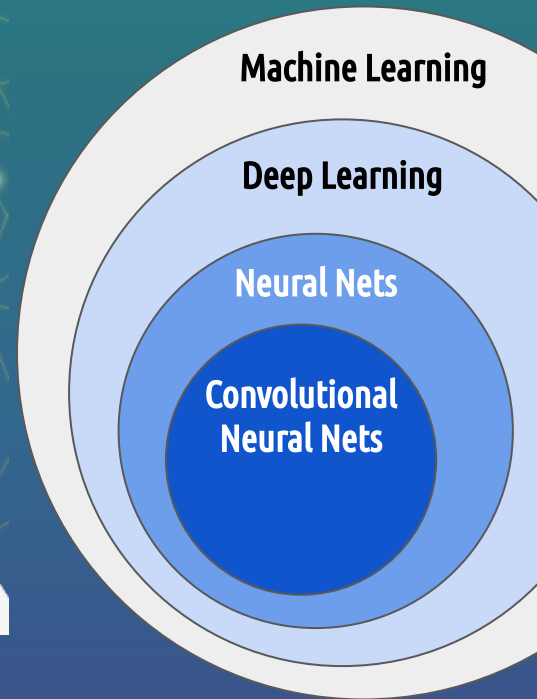
MACHINE LEARNING



Translation



Self-driving cars

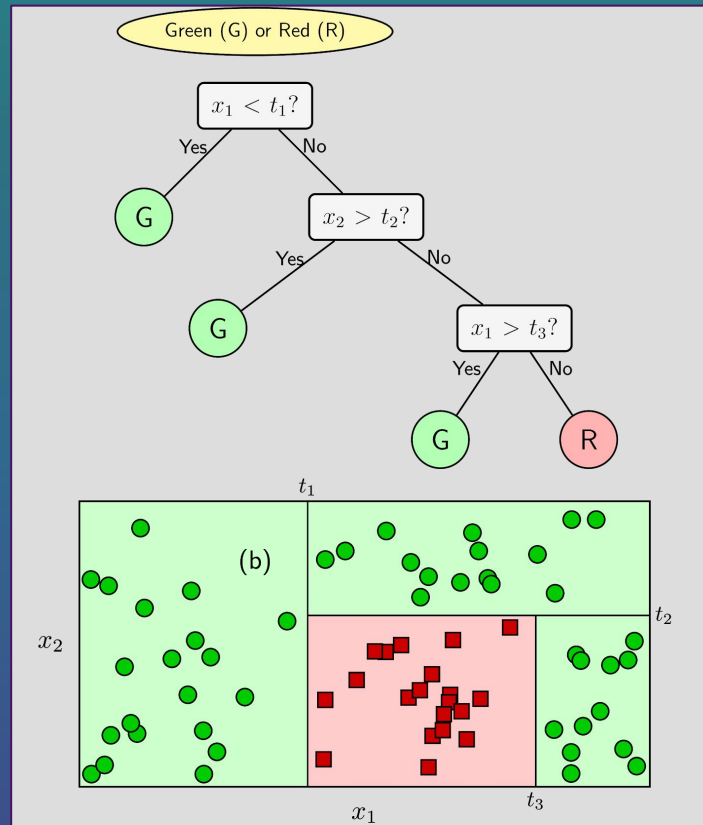


DECISION TREES

- Decision trees are the simplest form of machine learning
- The thresholds and position of each decision node are varied until error is minimised.

Problems:

- Decision thresholds are linear (eg 1D)
- Requires input of 'features' derived from data



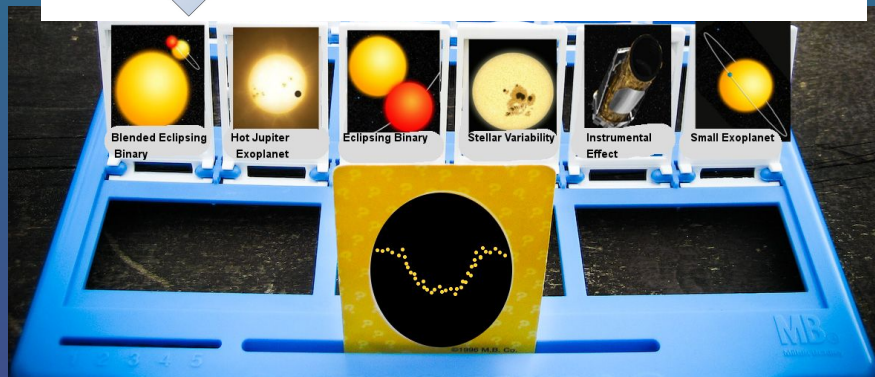
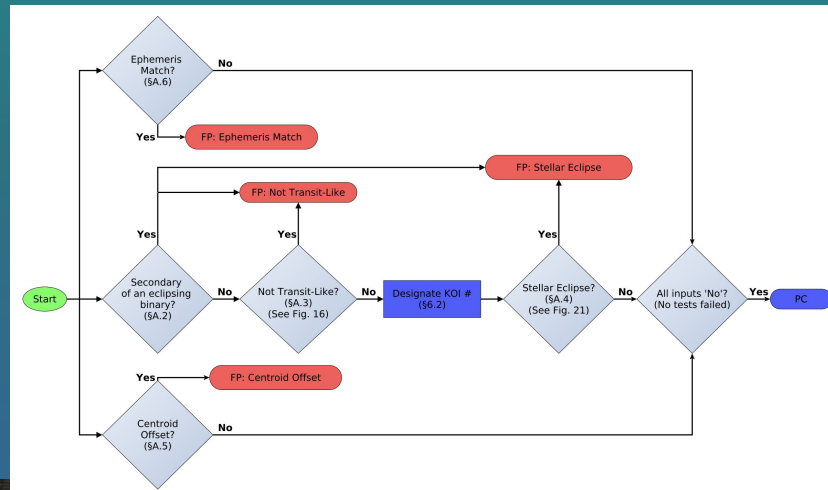
ROBOVETTER - DECISION TREE

“Robovetter” – Thompson et al 2017.

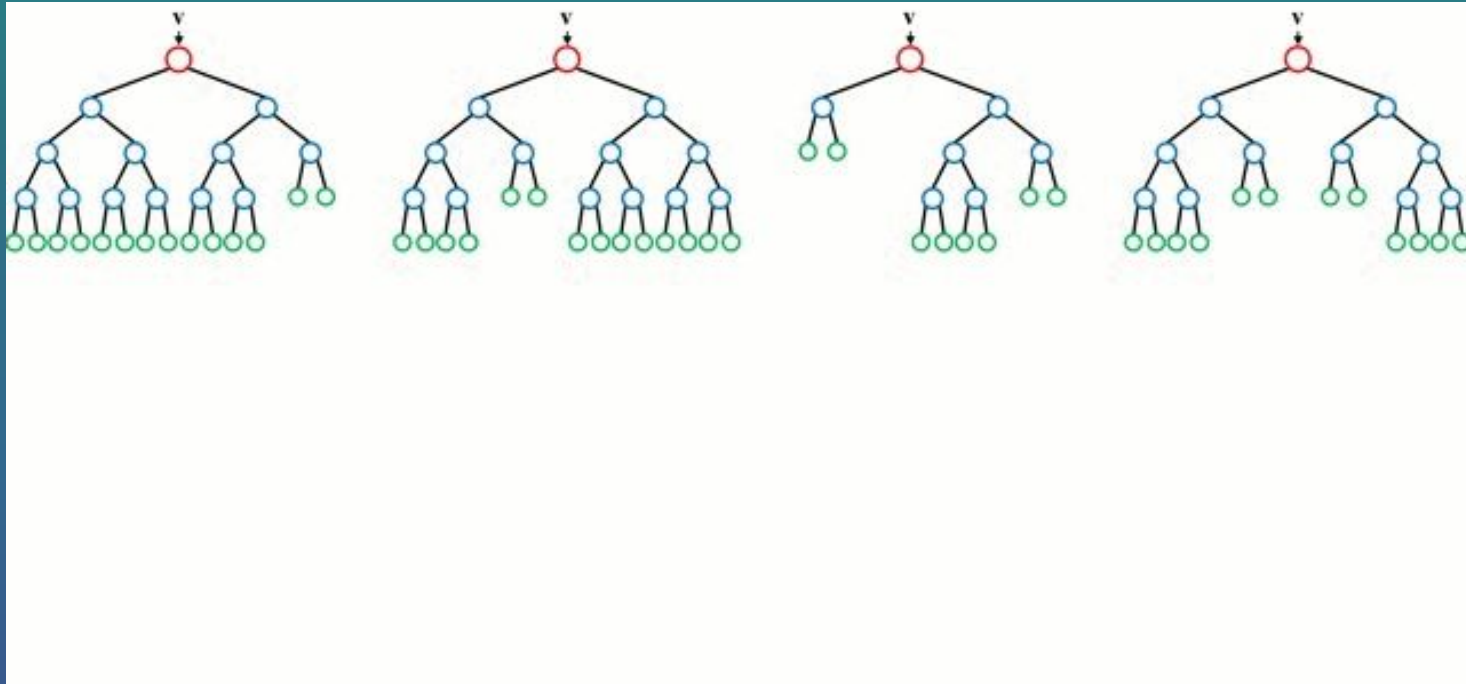
Decision tree classifier used to produce Kepler’s homogenous catalogue in DR25.

Used features processed from lightcurve.

Achieved a recall of around 80% on injected data.



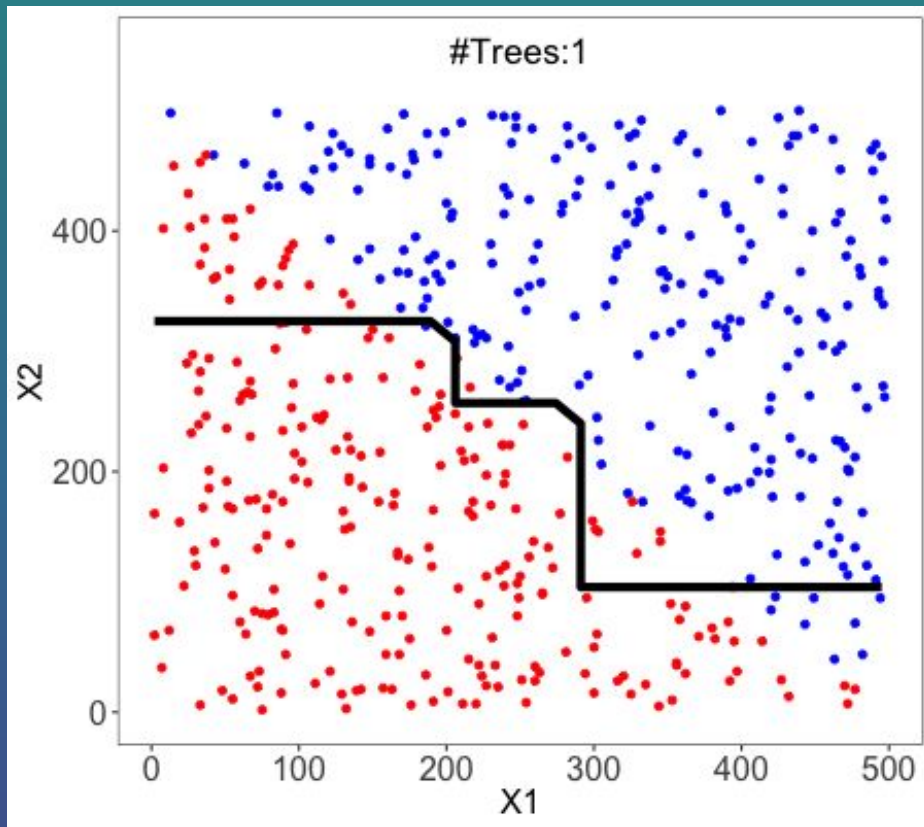
RANDOM FORESTS



- Each tree sees random subset of whole dataset
- Each decision step uses random selection of available

RANDOM FORESTS

- While each tree splits the data “linearly”, averaging of many trees approximates non-linear splits in data.



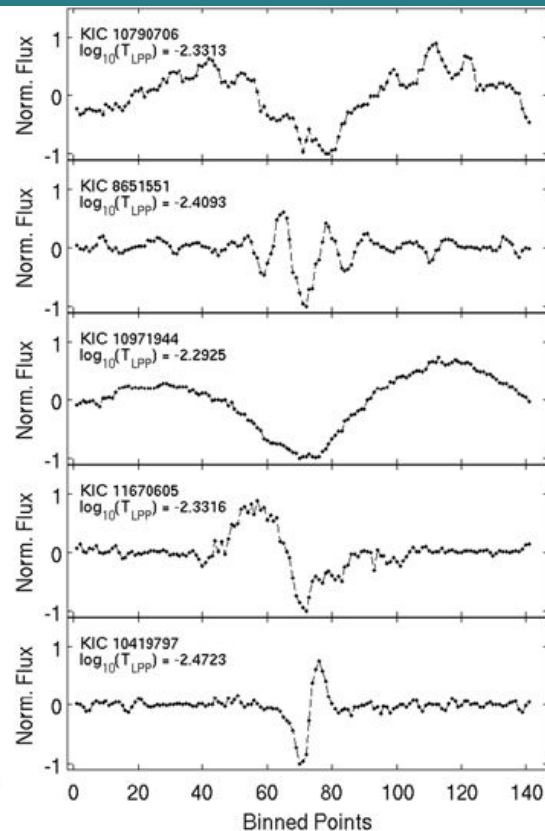
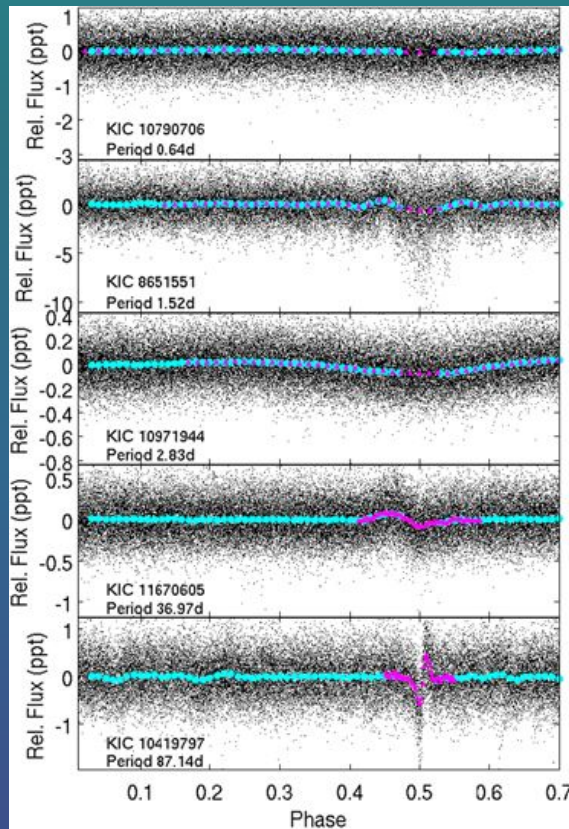
EXOPLANET CLASSIFICATION WITH KNNs

Thompson et al (2015).

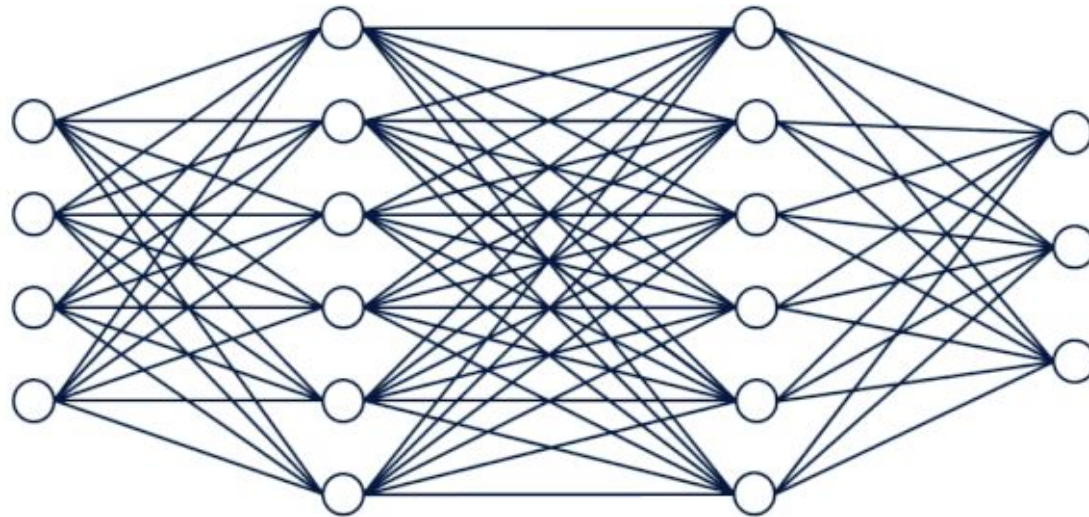
Used a “K-Nearest Neighbours” (KNN) unsupervised approach.

Takes average of nearest labelled features.

Used as inputs binned & normalised phase-folded transits.



NEURAL NETWORKS



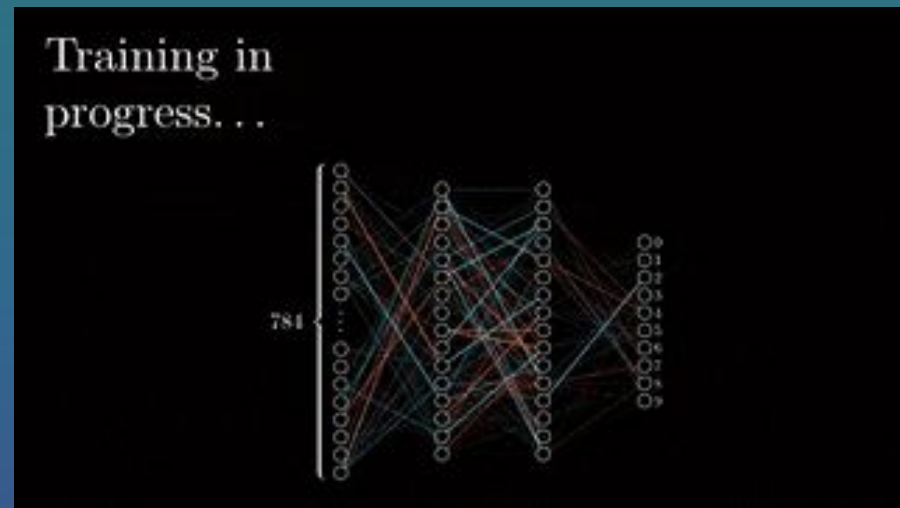
Hidden Layers



NEURAL NETWORKS

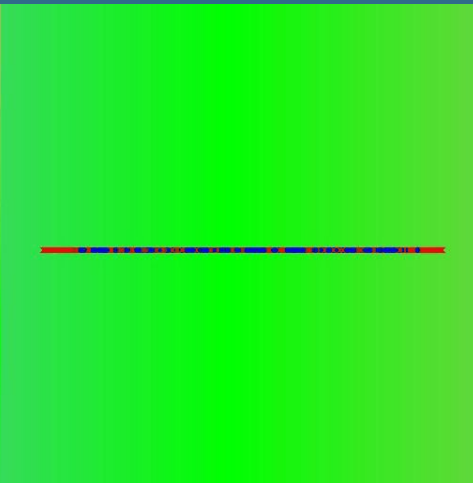
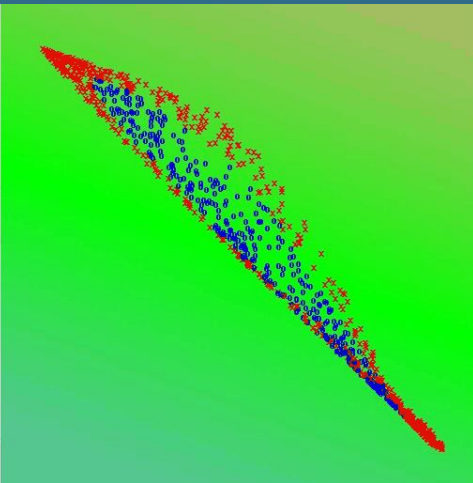
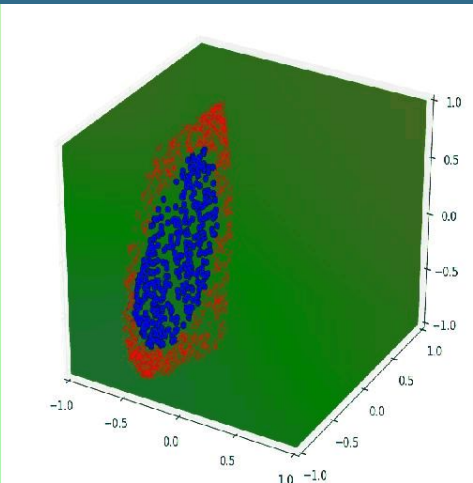
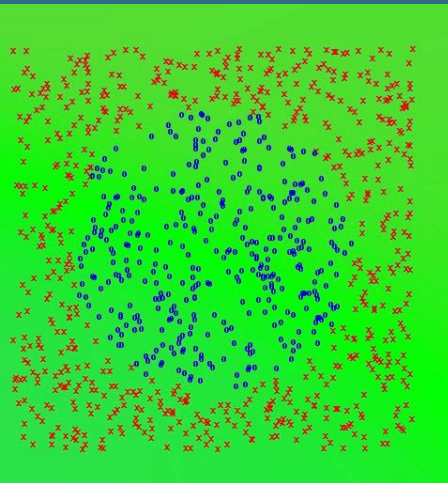
Training neural networks

- Quantify how poorly prediction was compared to ground truth
- Performance is then “back-propagated” through network to weights between neurons.
- These are adjusted such that the updated weight should decrease overall loss



NEURAL NETWORKS

- Neural Networks are not inherently “linear” – can better map irregular parameter spaces
- Hidden layers allow “abstraction” – acts like a new dimension in which to “fold” the (lower dimensionality)



UNSUPERVISED LEARNING (SOMs)

Armstrong et al, 2016

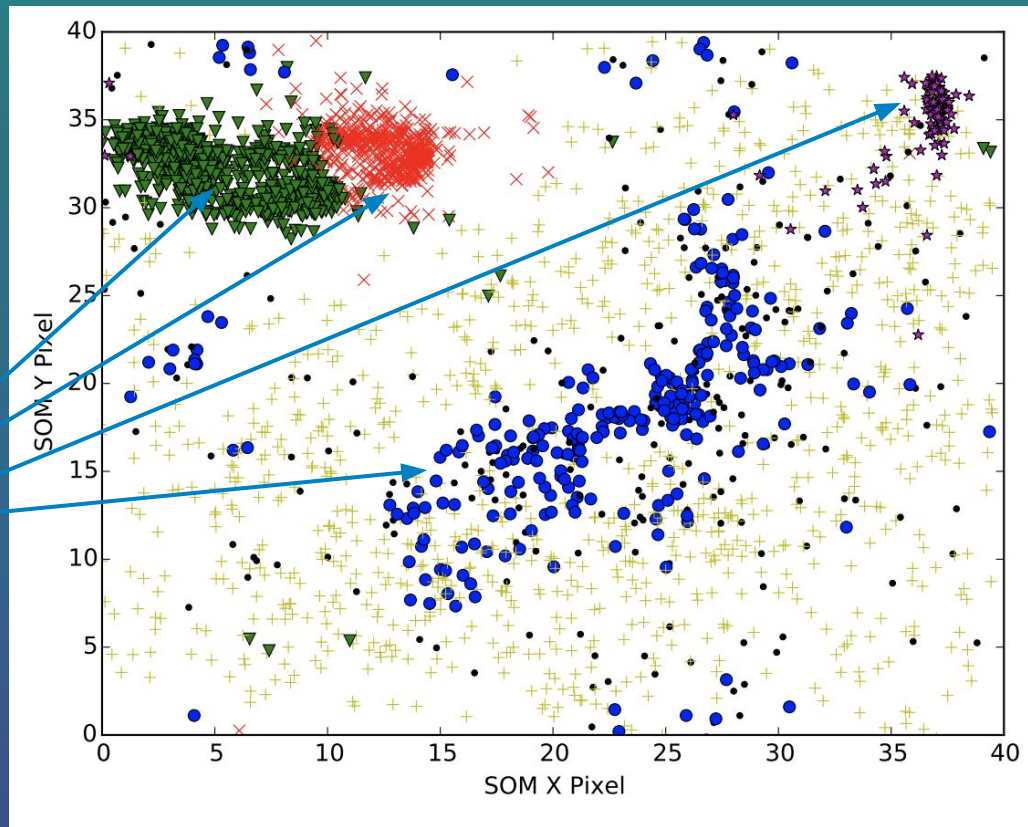
Self-Organising Map (SOM) –
a type of neural network
which reduces dimensionality
without any supervision

Creates isolated regions of
self-similar input data

Performed on 4 K2 campaigns.

Pixel position used as input
into Random Forest.

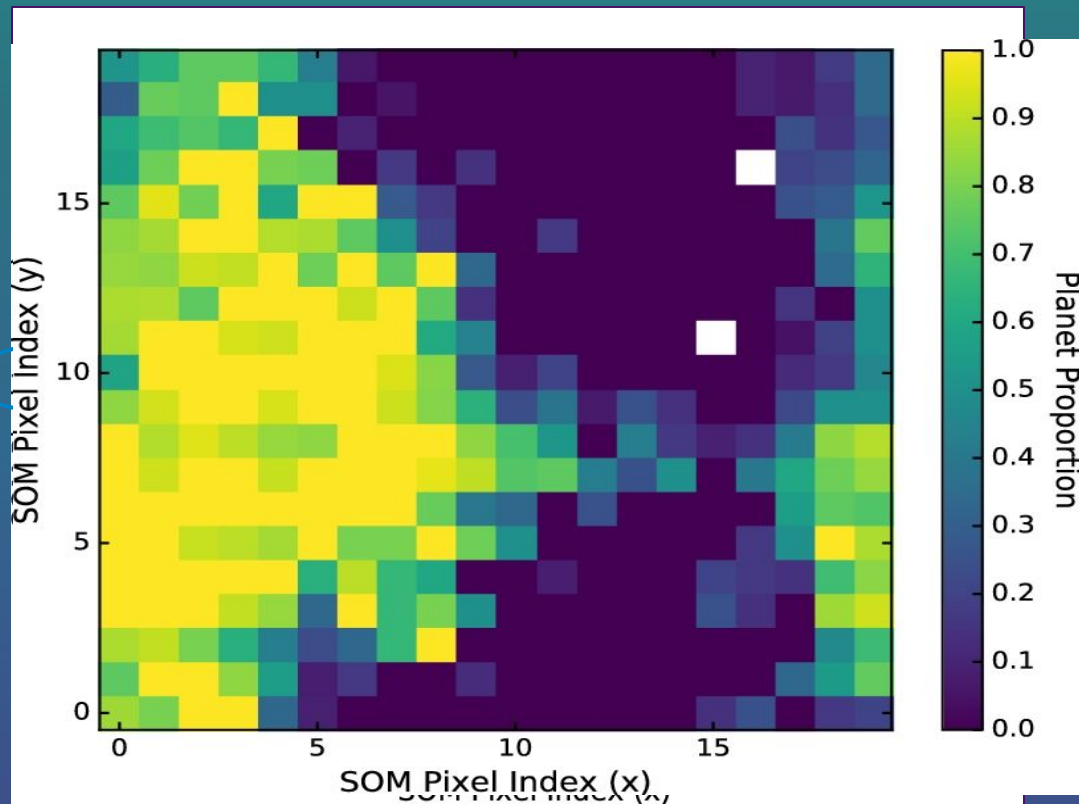
Detached EBs
Contact EBs
RR Lyraes
Delta Scutis



SELF ORGANISING MAPS FOR EXOPLANETS

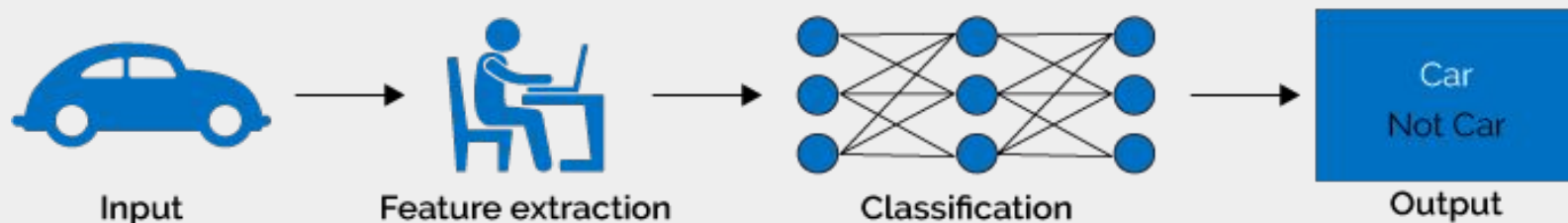
Armstrong et al (2017)

SOM and random forest
applied to Planet
candidates in K2 & Kepler
~79% accuracy on Kepler
planets

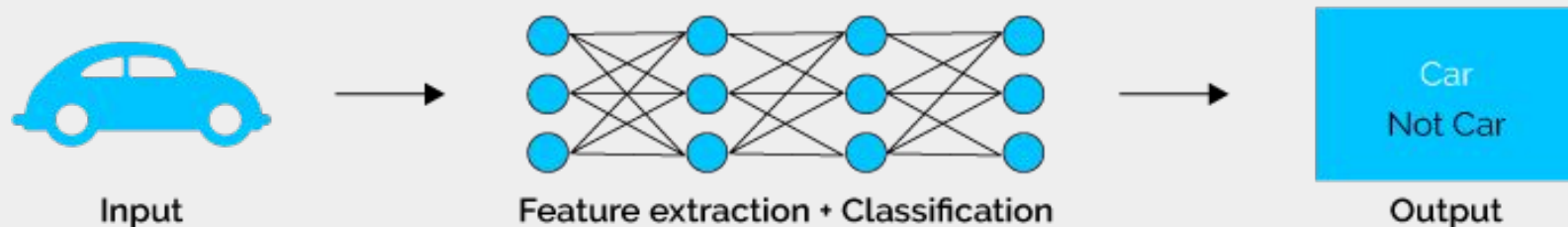


CONVOLUTIONAL NEURAL NETWORKS

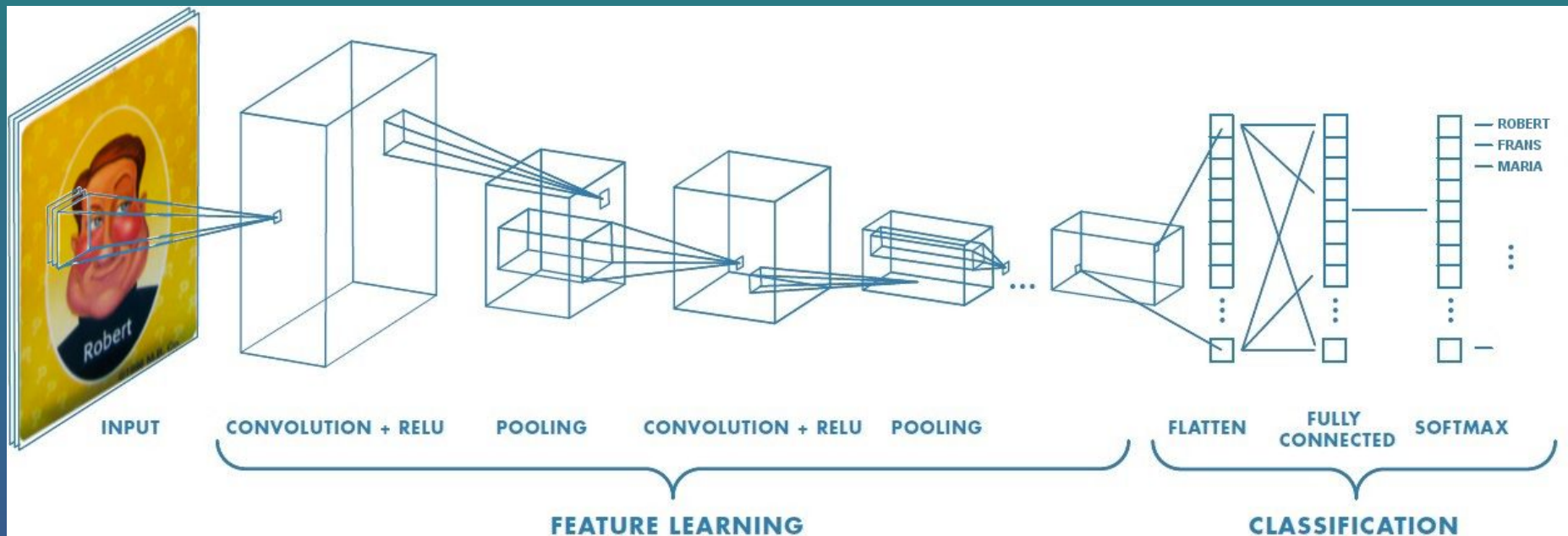
Machine Learning



Deep Learning



CONVOLUTIONAL NEURAL NETWORKS

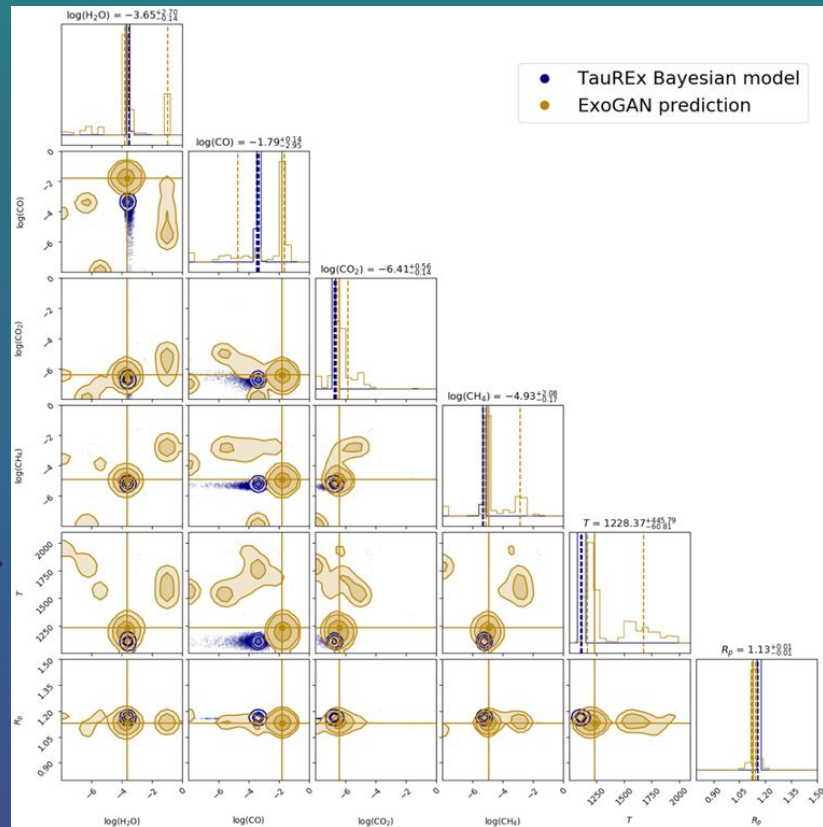
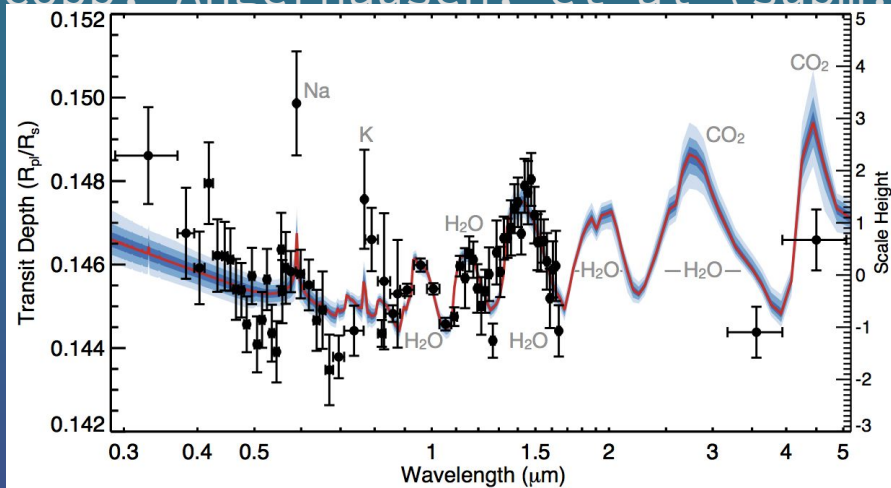


- Raw image “convolved” with range of filters (which themselves are trained with back propagation)
- Enables Feature extraction from the raw data (although raw

CNNs FOR ATMOSPHERIC RETRIEVAL

Waldmann (2015) & Zingales (2018) – RoBERT using Neural networks

Cobb, Angerhausen, et al (subm.)



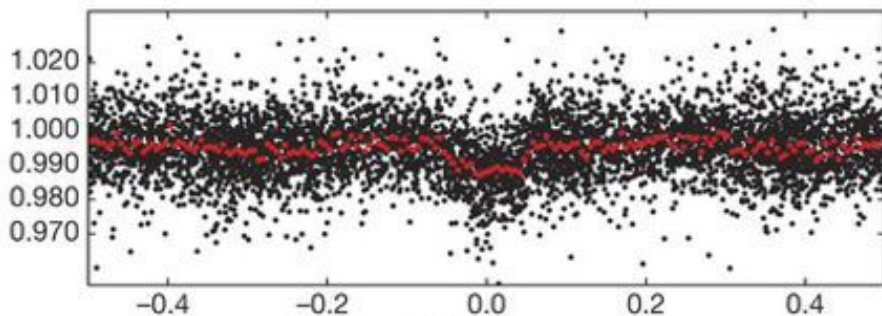
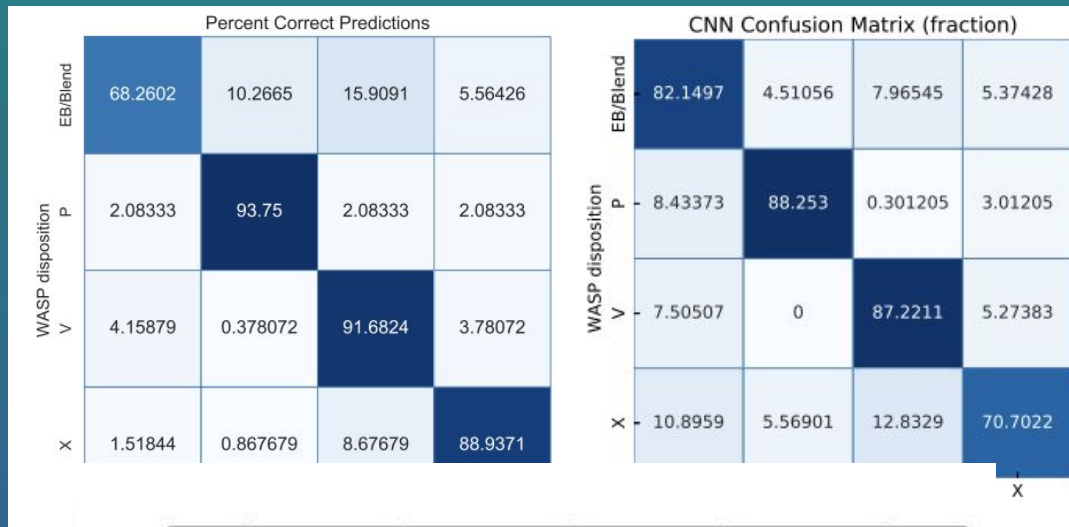
GROUND-BASED TRANSITS WITH RF & CNNs

Schanche et al. (2018)

Classified WASP planet candidates with both Random Forest and Convolutional Neural Network.

CNN gives better average precision, but random forest performs best on planets:

MEarth used Neural



CNNS FOR EXOPLANET DETECTION

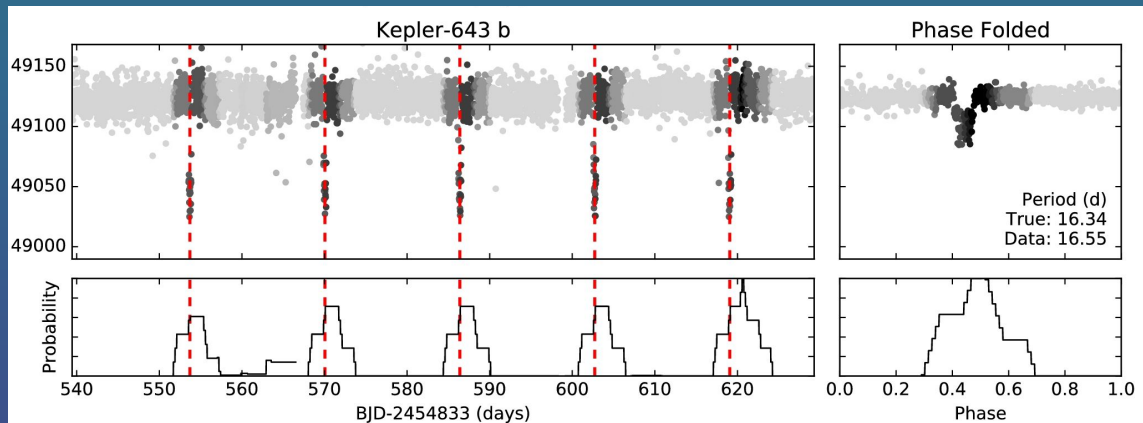
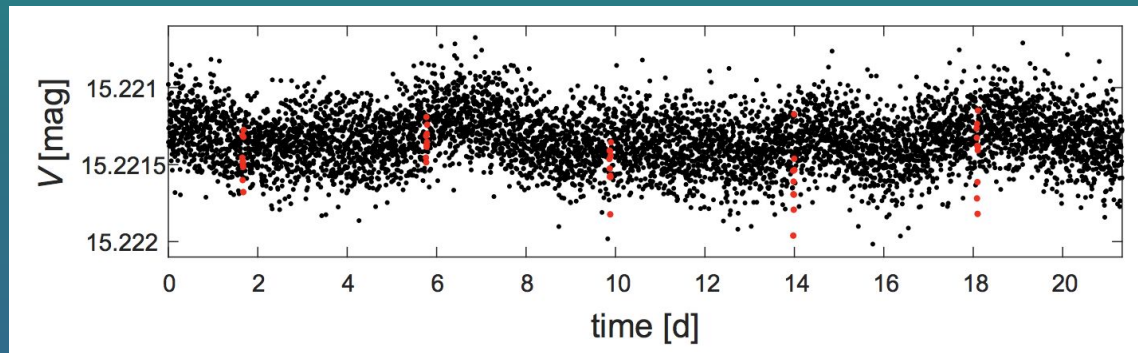
Two parallel papers using neural networks to detect exoplanets:

Zucker et al, (2017)

Pearson et al, (2017)

Difficult as neural networks cannot natively learn “periodicity”.

Neither deal with classifying real planets vs false positives



CLASSIFYING TESS SIMULATIONS

