MLSE, Virtual, 2020

Classifying Transiting Exoplanet Candidates with Deep Learning

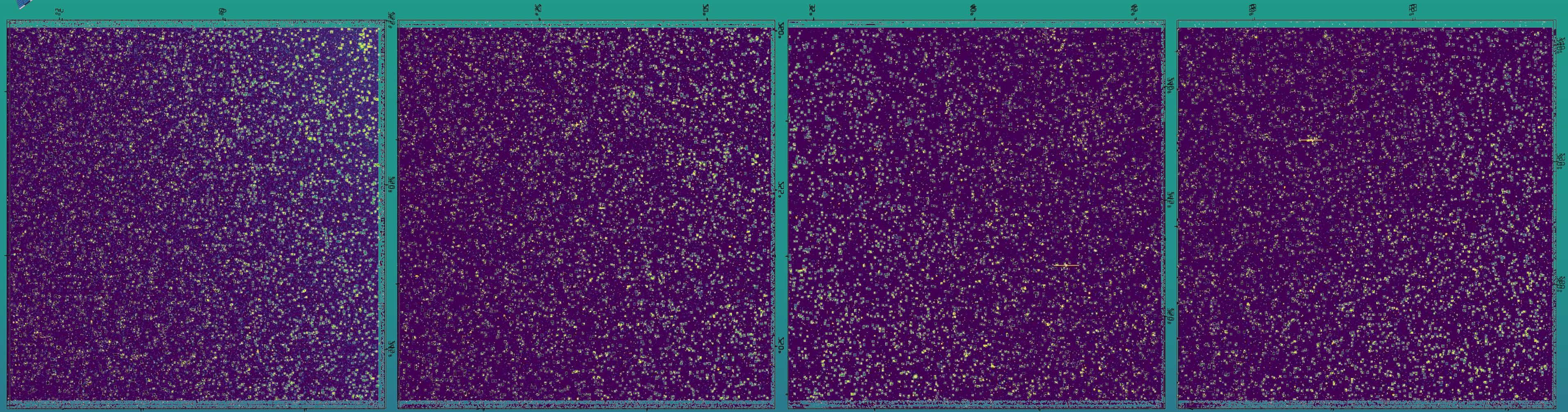
Hugh Osborn

CHESS Fellow at University of Bern & MIT

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

@exohugh

The problem:
From raw data to planets

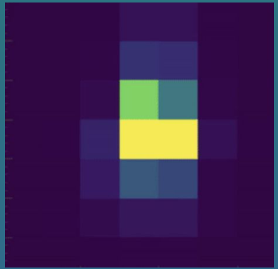


“Postage stamps” for target stars

Typical TESS Raw data

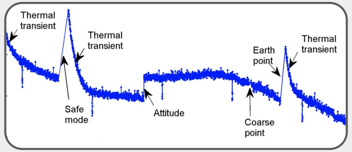
Kepler & TESS pipelines

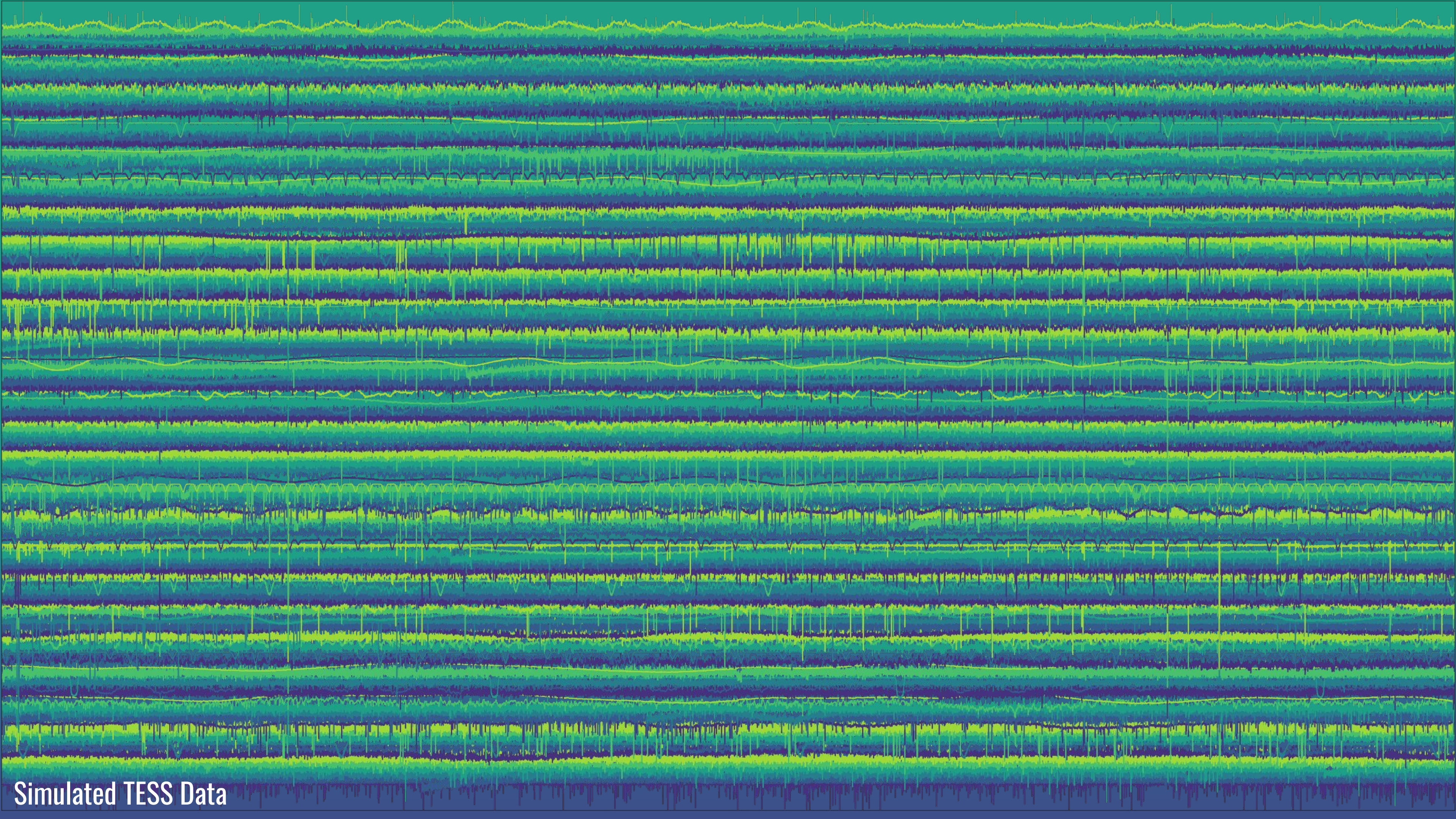
Target Pixel
File (TPF)



Smith+2012, Stumpe+2012

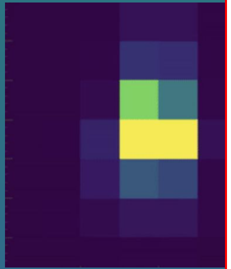
**Aperture Photometry &
Systematic Correction**

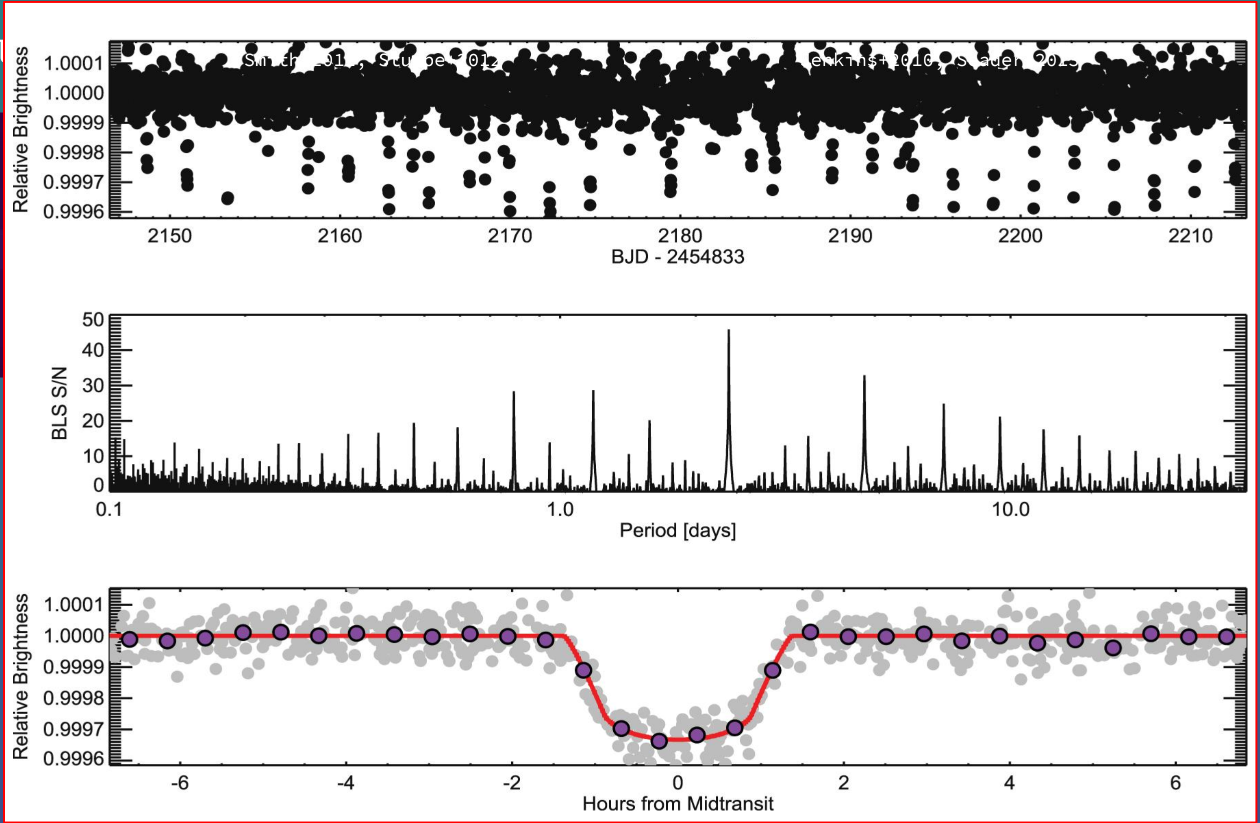


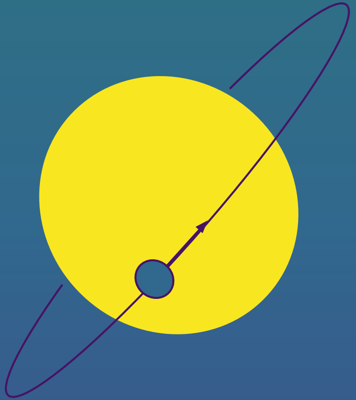

Simulated TESS Data

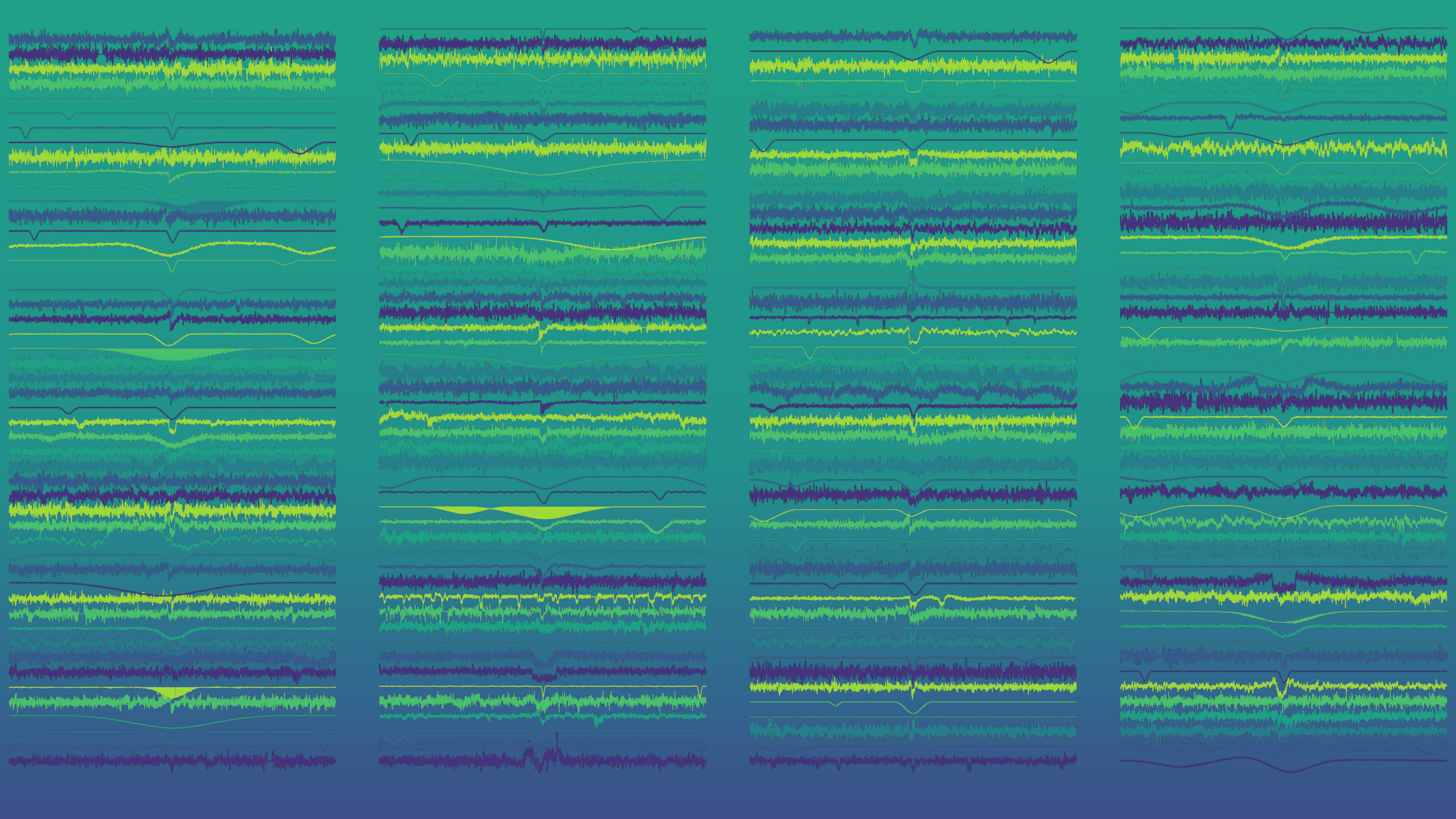
Kepler & TESS pipelines

Target Pixel File (TPF)



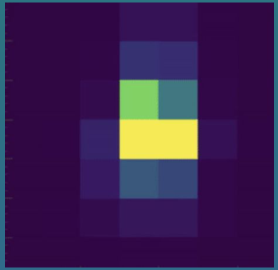






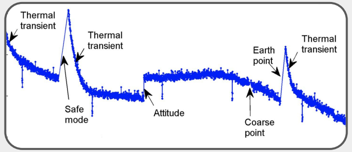
Kepler & TESS pipelines

Target Pixel File (TPF)



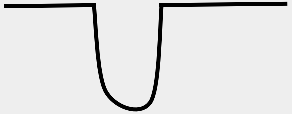
Smith+2012, Stumpe+2012

Aperture Photometry & Systematic Correction



Jenkins+2010, Seader+2013

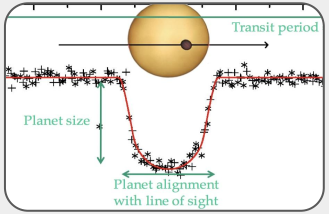
Transiting Planet Search (TPS)



Threshold Crossing Event (TCE)

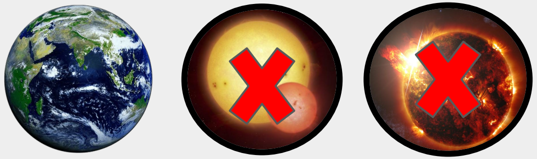
Wu+2010

Data Validation (DV)



Batalha+2013, Burke+2014, Rowe+2015, Mullally+2015

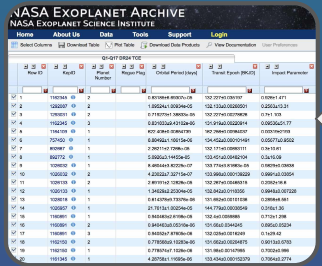
Candidate Classification I.e. Human vetting



Follow-up observations

Confirm/statistically validate planets

Exoplanet Catalogues



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?

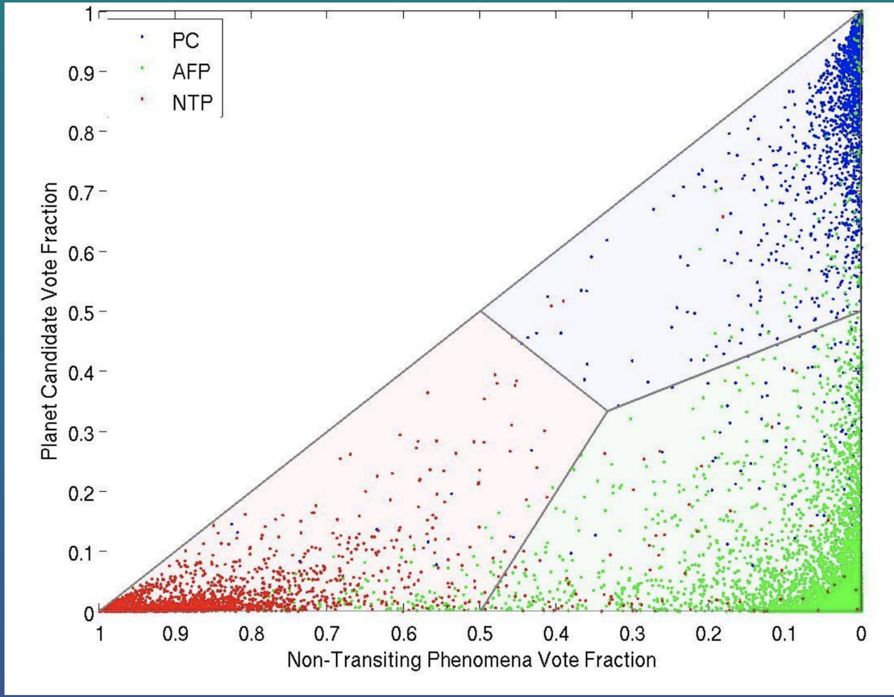
Autovetter - Random Forests

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

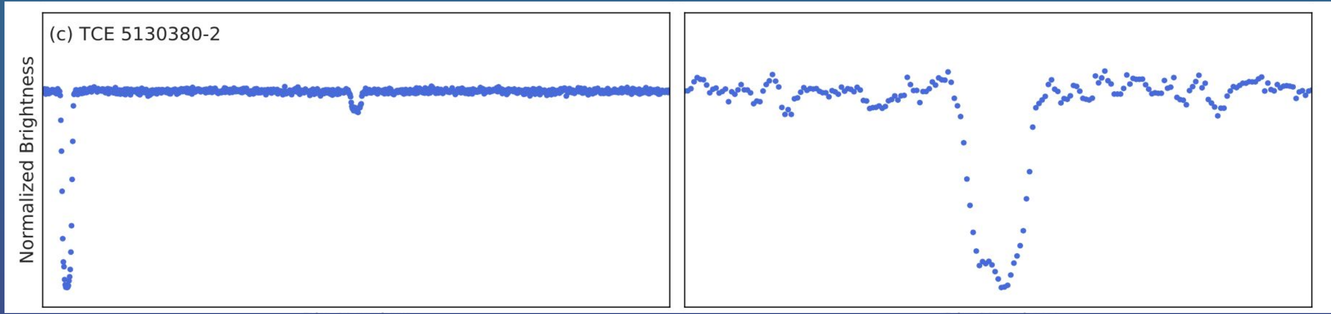
94.15% precision & 97.2% average precision (on human-labelled data)

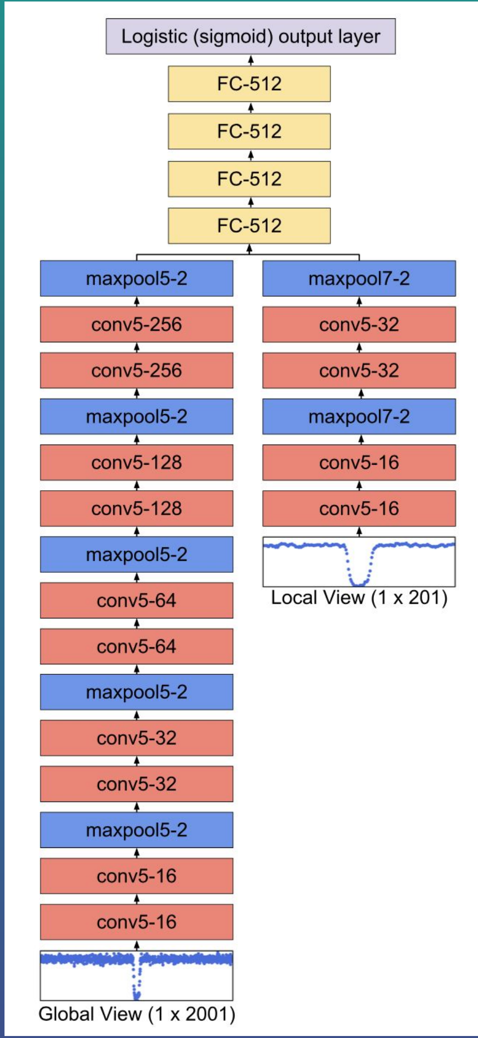


Shallue et al 2018

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]





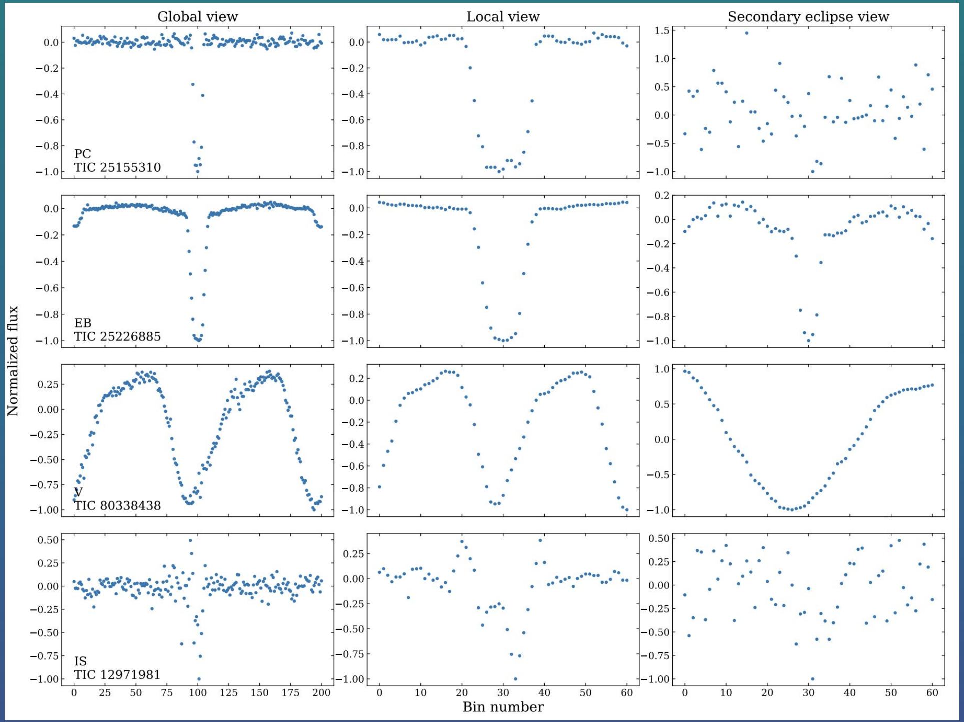
Developments on AstroNet

Application to K2 data (Dattilo et al 2019).

Application to TESS vetting: Yu et al (2019)

Included secondary eclipse region as an input.

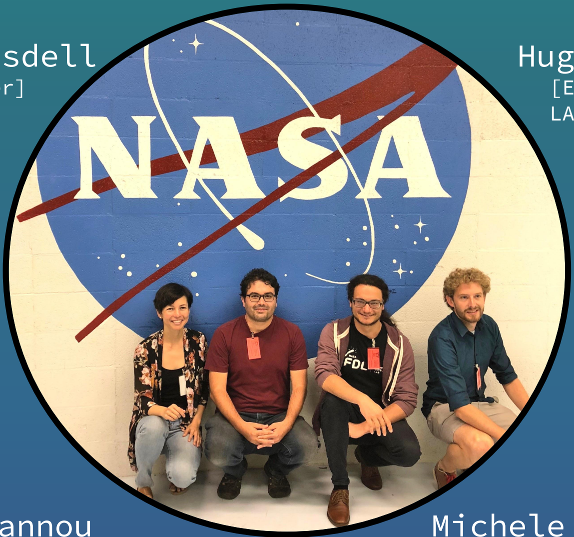
Currently used in TESS vetting at MIT.



2018 NASA FDL-Exoplanet Team

Megan Ansdell
[Exoplaneteer]
UC Berkeley

Hugh Osborn
[Exoplaneteer]
LAM, Marseille



Yani Ioannou
[Deep Learning Expert]
University of Cambridge

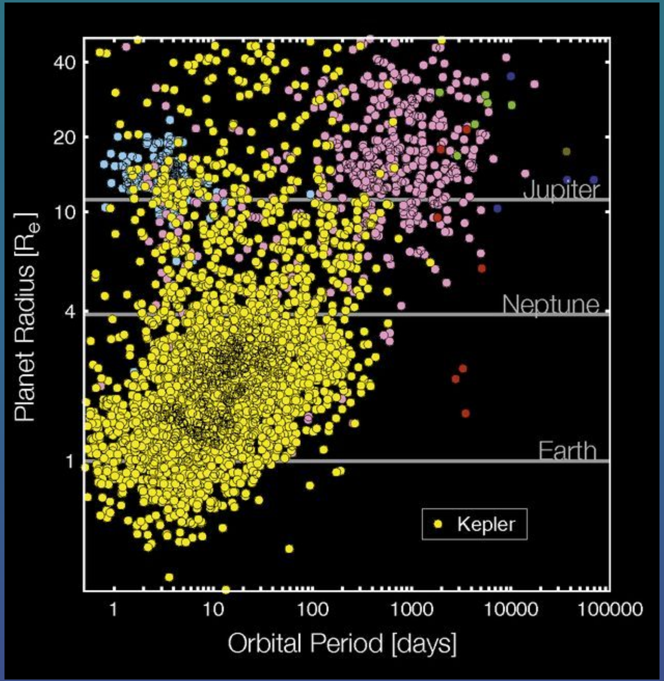
Michele Sasdelli
[Deep Learning Expert]
University of Adelaide

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

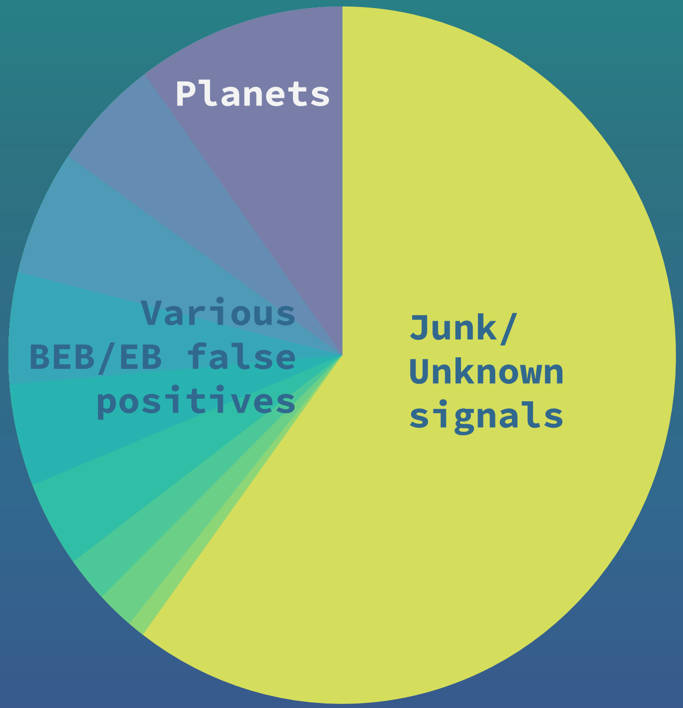
- 16,000 Threshold Crossing Events (TCEs) from Kepler DR24
- Labelled by human vetters
- ~25% planets & ~75% false positives
- Preprocessed the data following Shallue & Vanderburg:
  - Detrended lightcurve
  - Phase-folded onto TCE period
  - Binned to global & local view



Ansdell, Ioannou, Osborn, Sasdelli, et al. (2018)

TESS Input Data & Labels

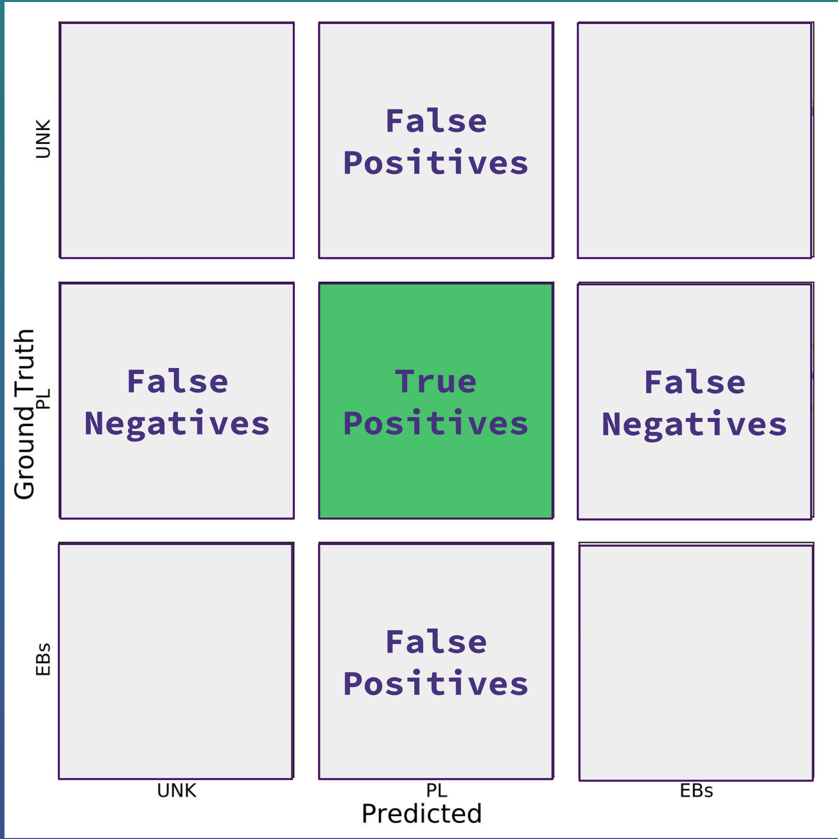
- 4 Simulated sectors - i.e. we know the exact ground-truth
- Pixel-level signal injection, processed with the TESS pipeline
- ~16,000 candidates, ~14% planets
- Preprocessed the data following Shallue & Vanderburg



Osborn, Ansdell, Ioannou, Sasdelli, et al. (2019)

Assessing Model Performance

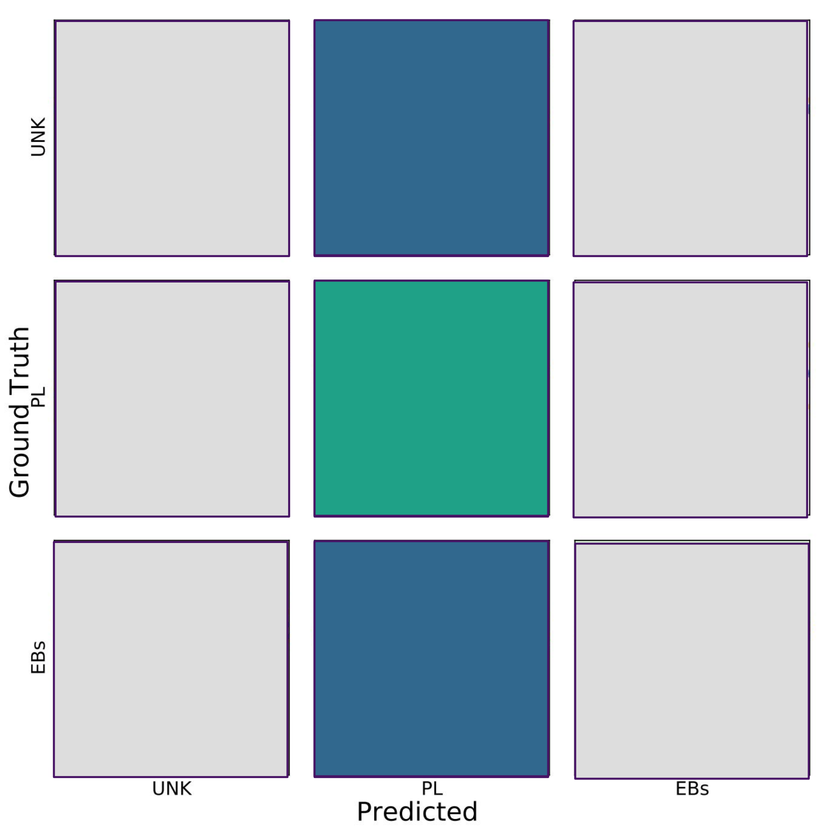
Definitions



Assessing Model Performance

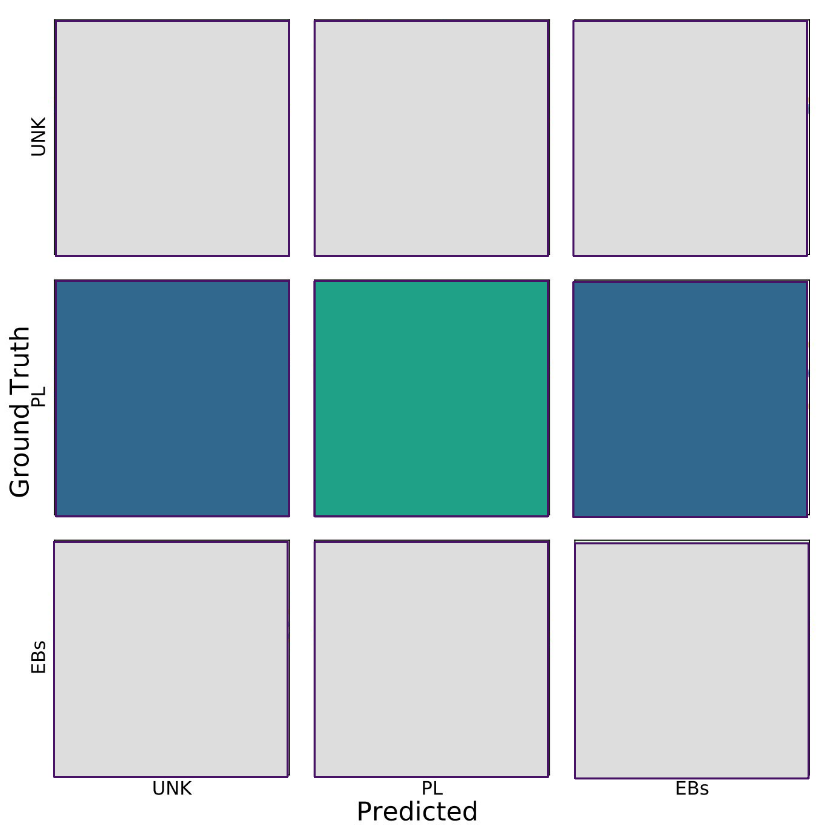
Precision

* also known as accuracy



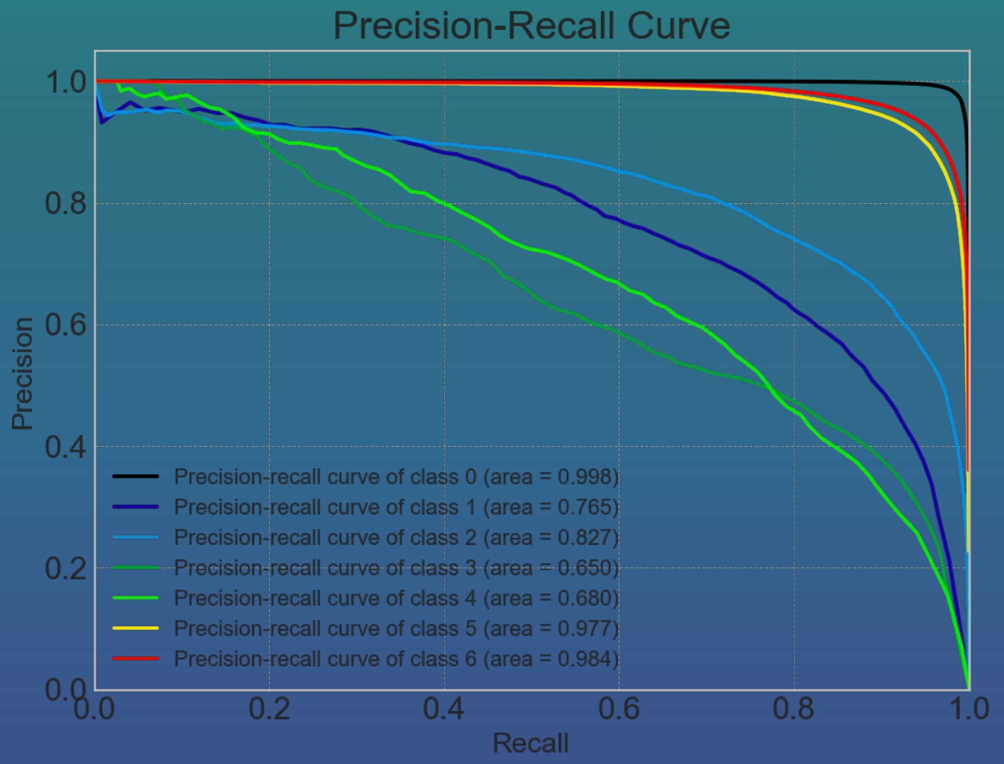
Assessing Model Performance

Recall



Assessing Model Performance

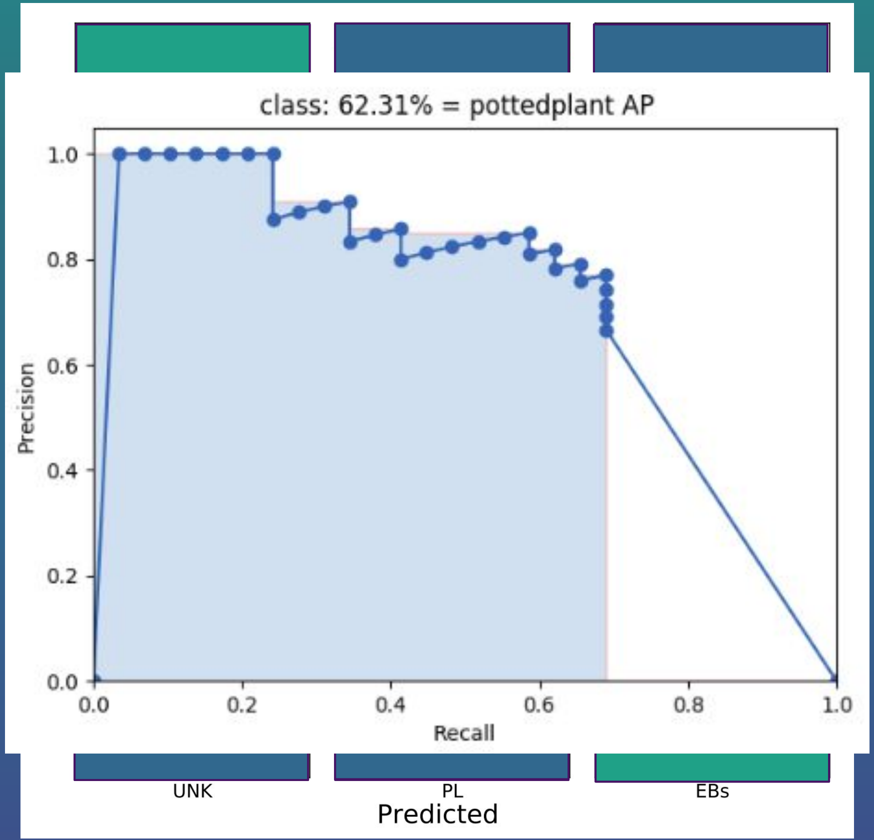
Precision-Recall Curve



Assessing Model Performance

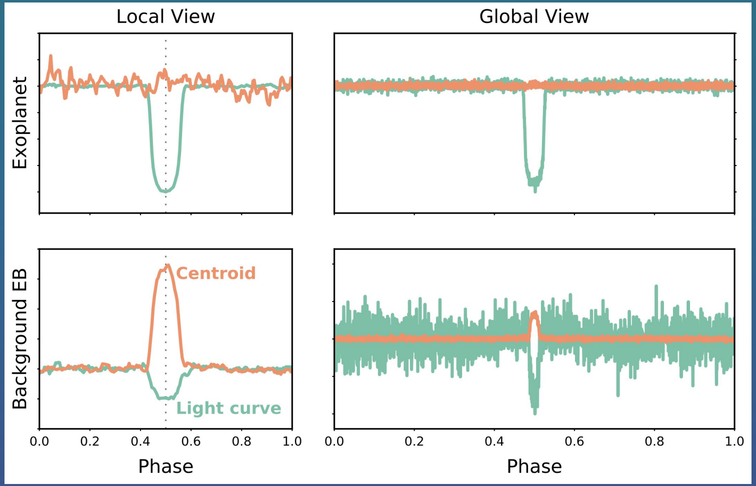
Average Precision

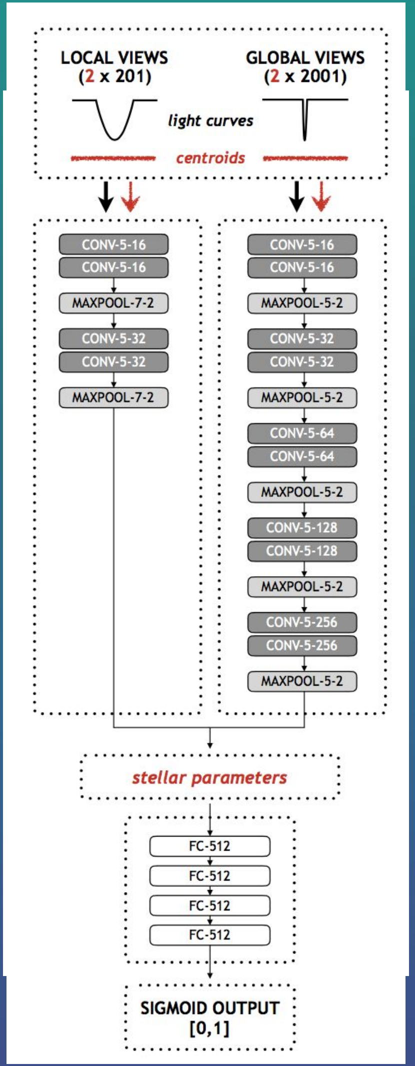
- Weighted average of precision for all classes.
- Functionally similar to Area Under Curve (AUC) for a multi-class classifier - i.e. probability a random positive sample is correctly predicted at any P-R threshold



Domain Knowledge - Centroids

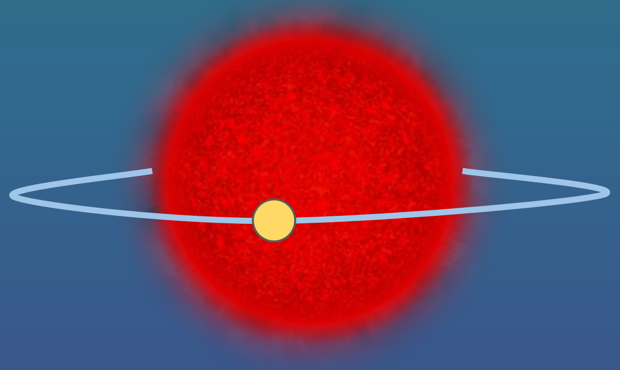
- Position of centre of light over time
- Important for identifying background EBs

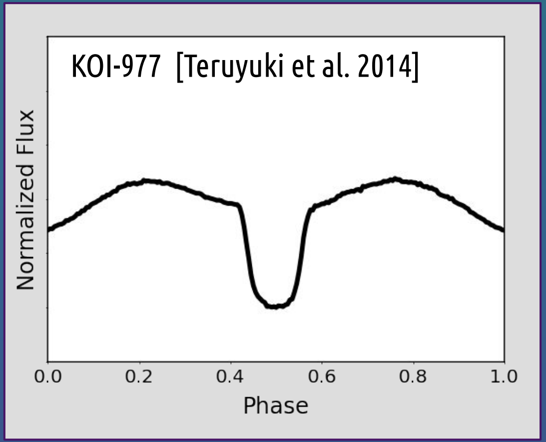


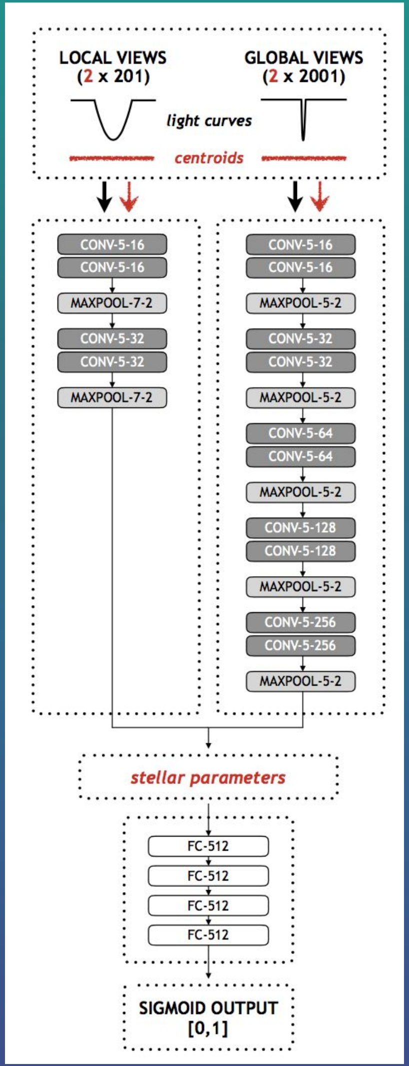


Domain Knowledge - Stellar properties

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

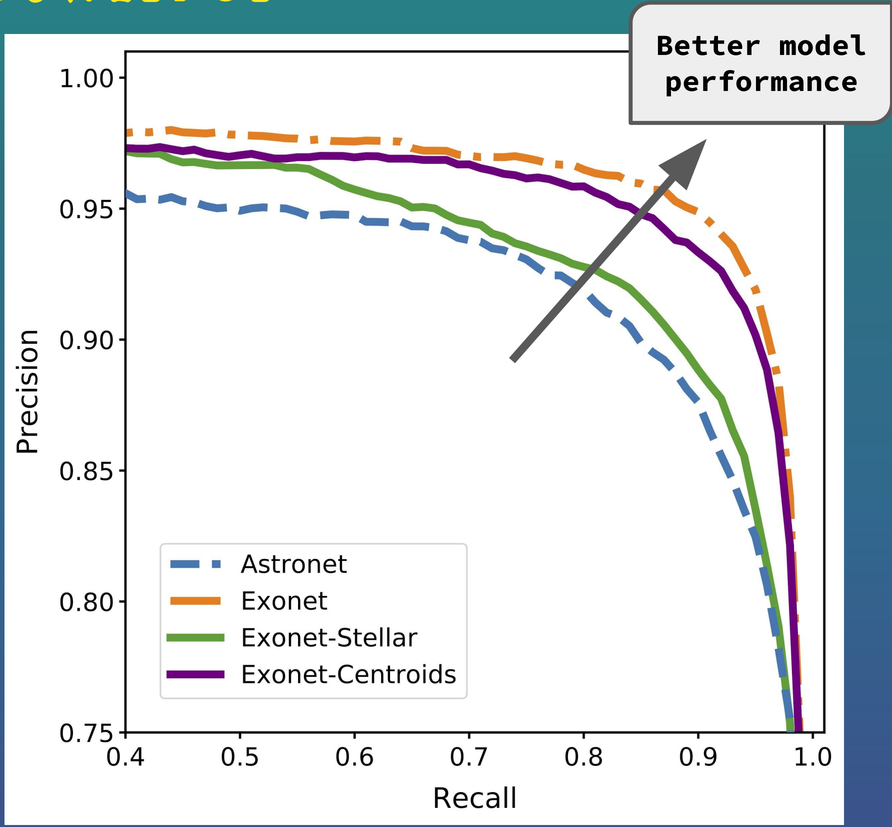






Performance With Domain Knowledge

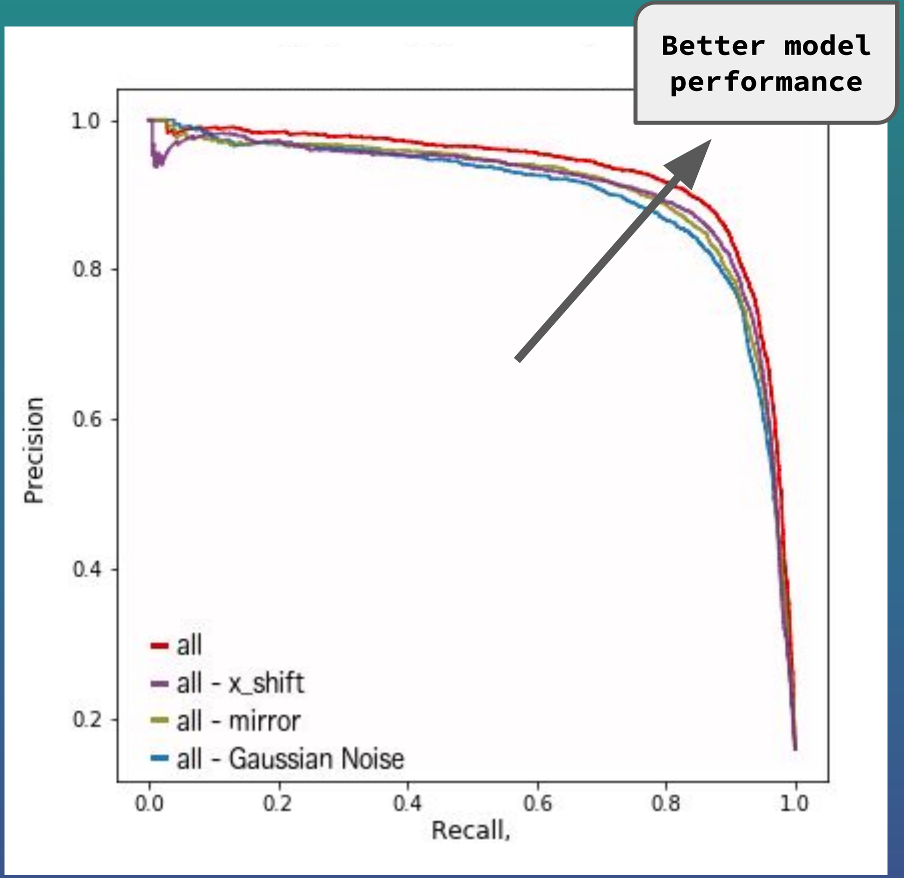
- Centroids & Stellar info both improve performance
- Also helped by cross validation & model ensembling



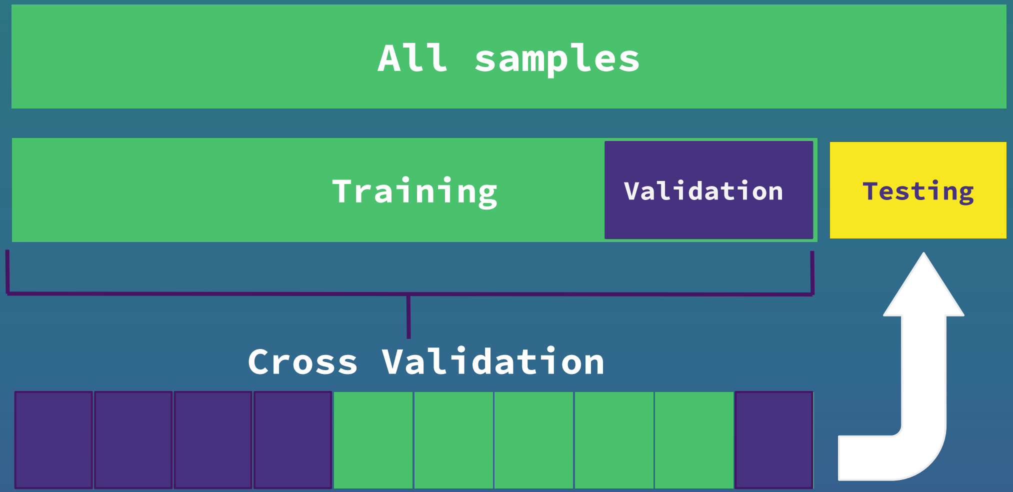
Data Augmentation

- 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”:

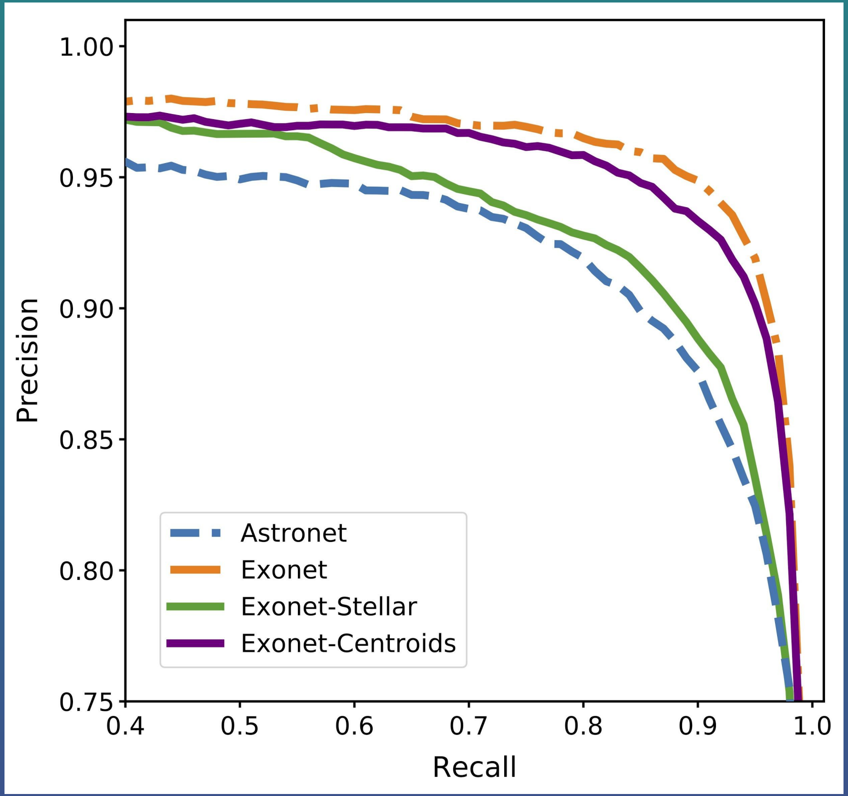
Taking average of models applied to test data

*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.



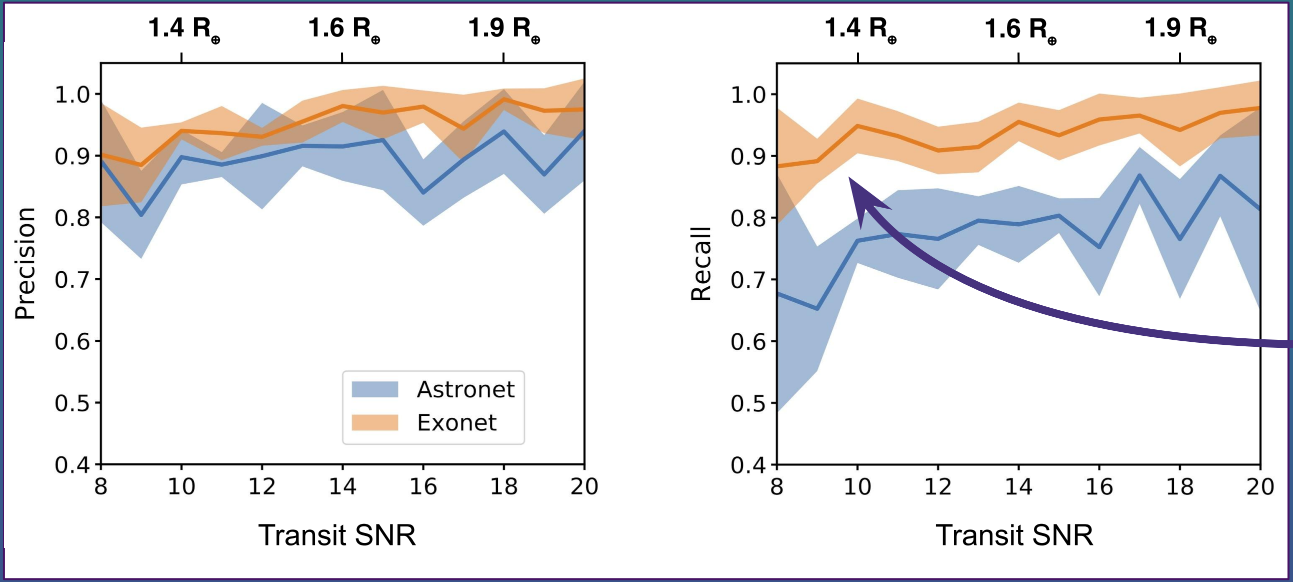
Kepler Performance

- 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



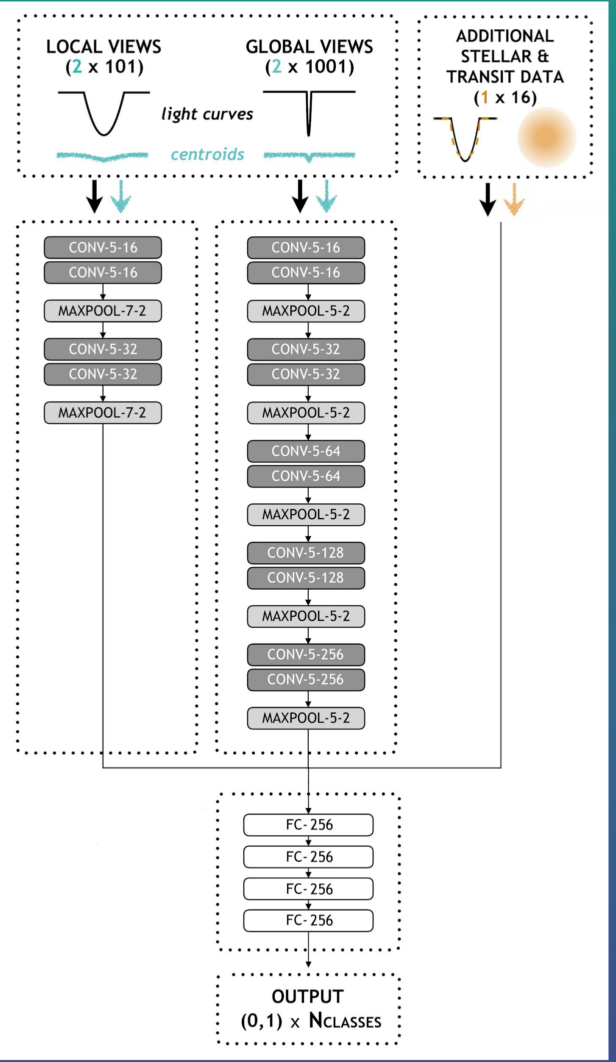
15-20% gains in recall for Earth-sized planets

Classifying TESS Data

Slightly modified from Kepler -> TESS

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

Osborn, Ansdell, Ioannou, Sasdelli, et al. (2019)

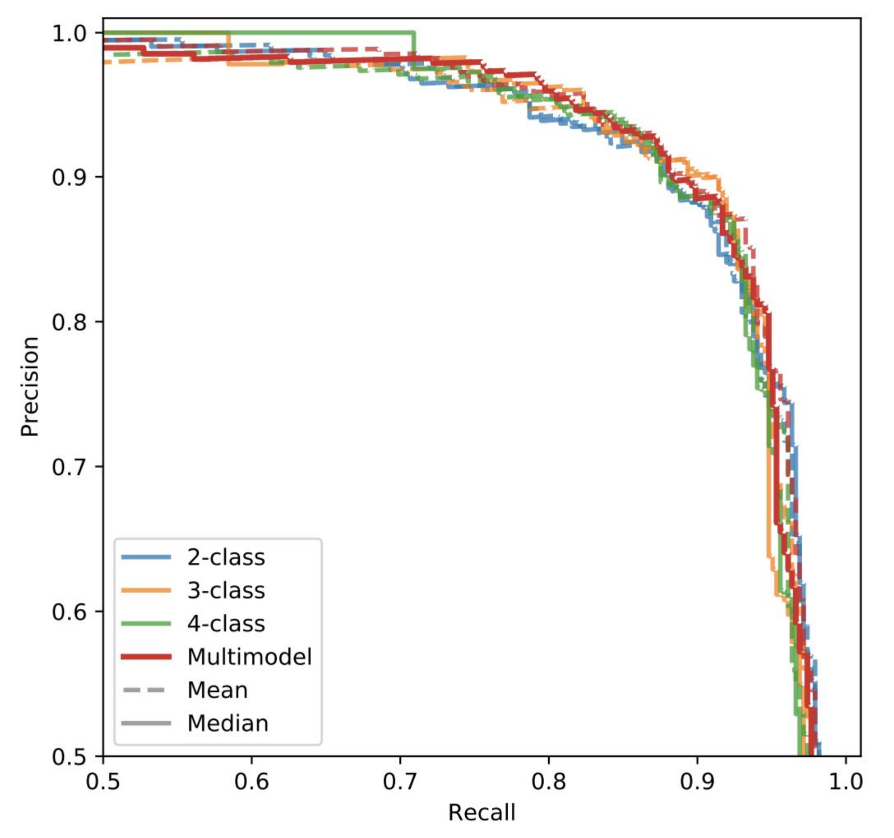


Balanced Batch Sampling

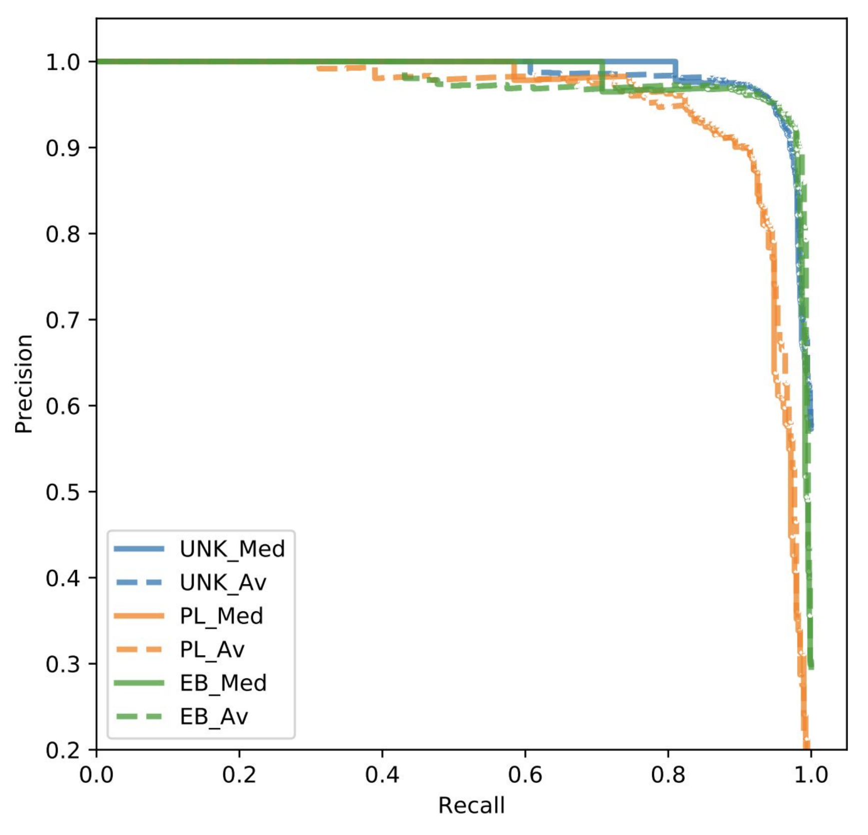
- Models tend to predict the majority class in unbalanced data
- Re-balancing means that each epoch sees same number of samples from each - helps training

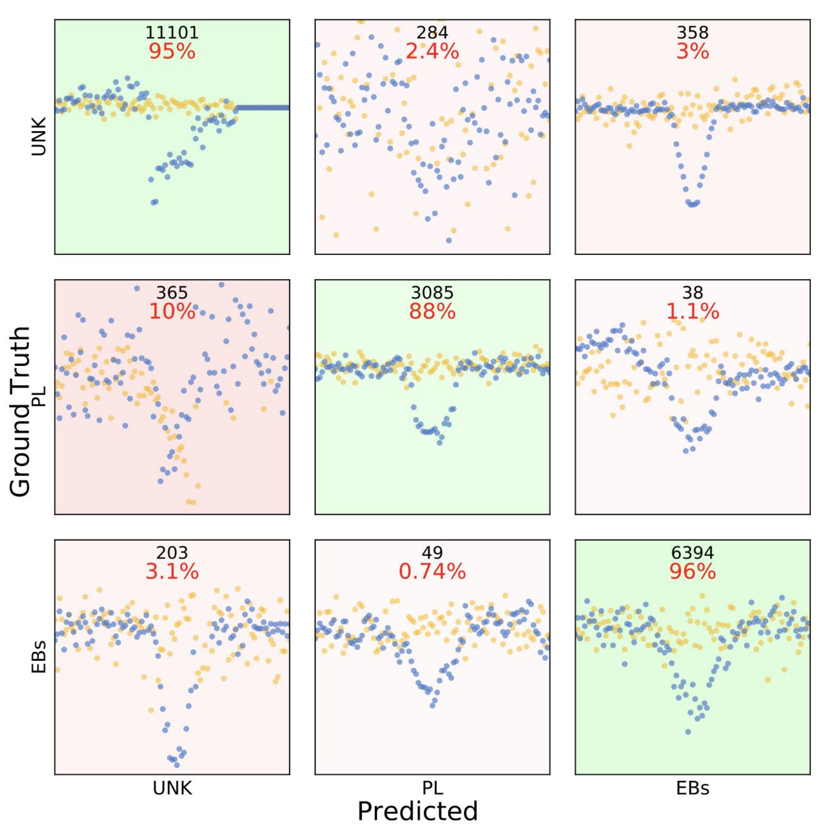
Performance on TESS Simulations

		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**	**90.4**	**90.1**	**95.6**
	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

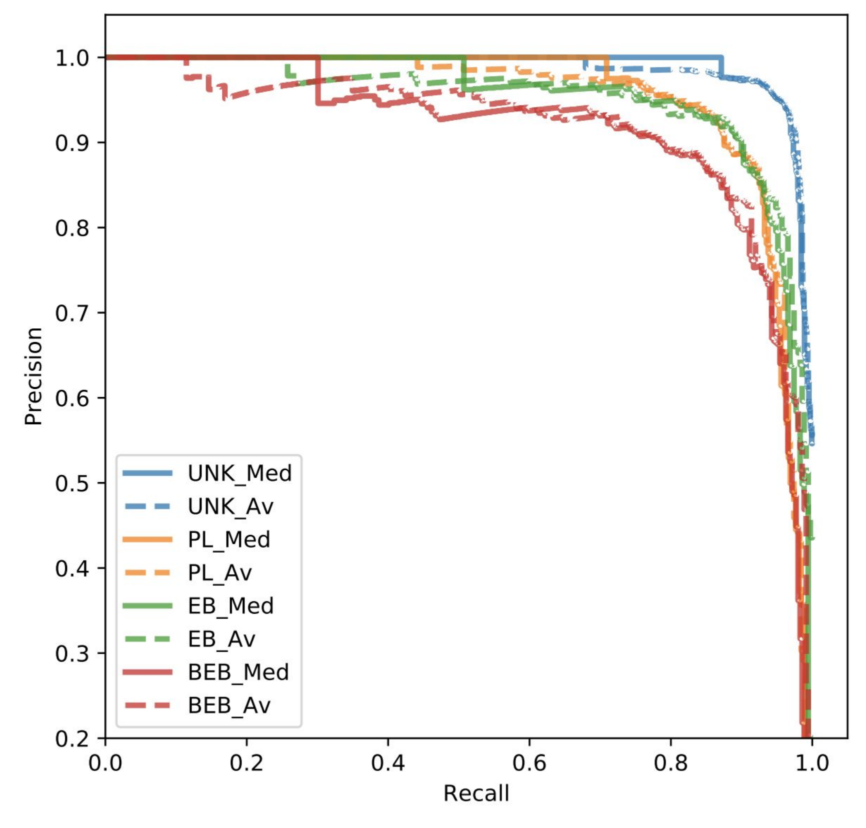


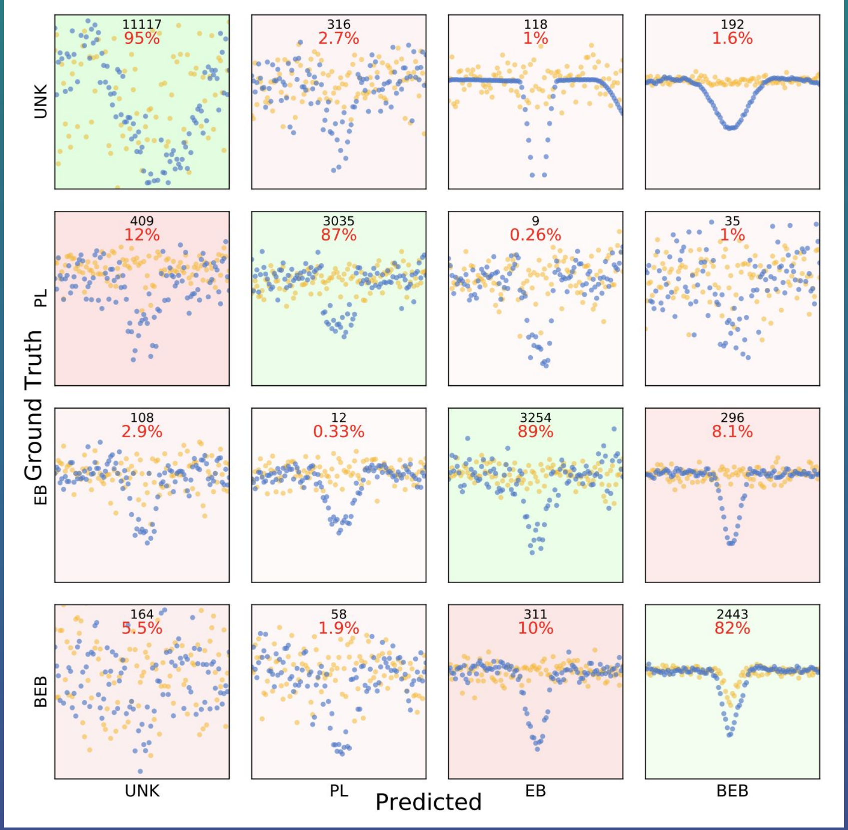
3-class model





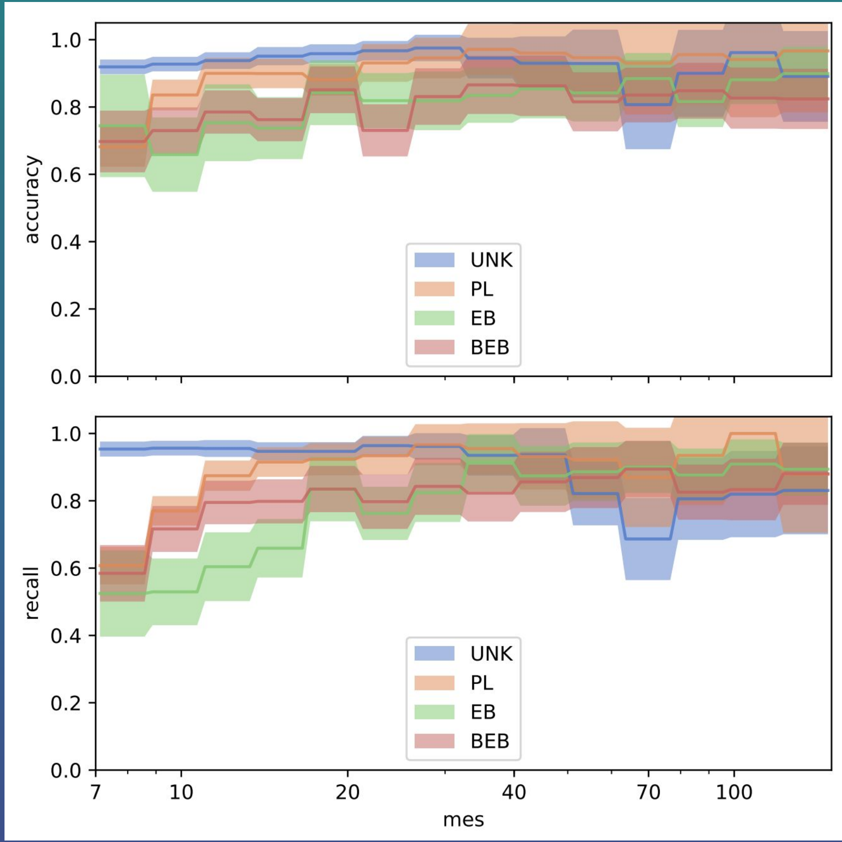
4-CLASS MODEL





Performance as a function of SNR

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

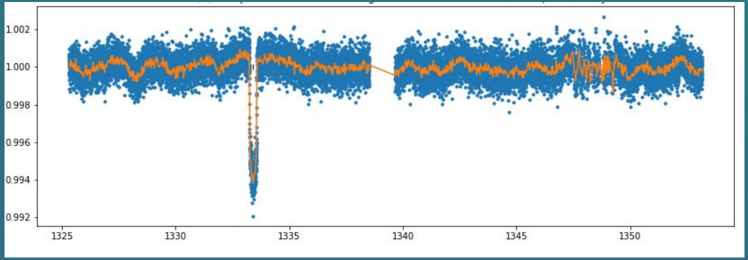


Comparison with Ansdell et al, 2019

	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

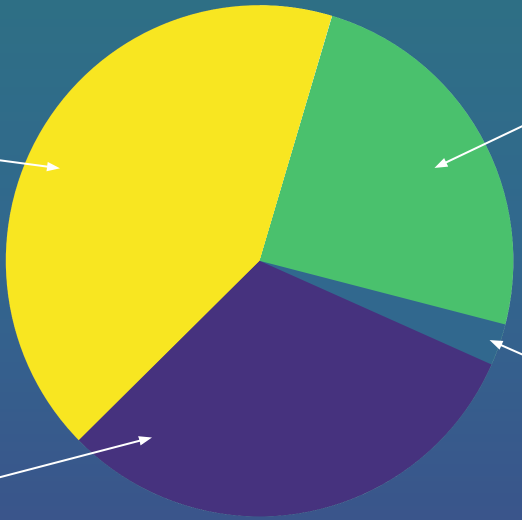
But real data ≠ simulated data

- 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



Recall of only 61% on TOIs

Our model thinks 30% are noise...

14 we suggest are EBs

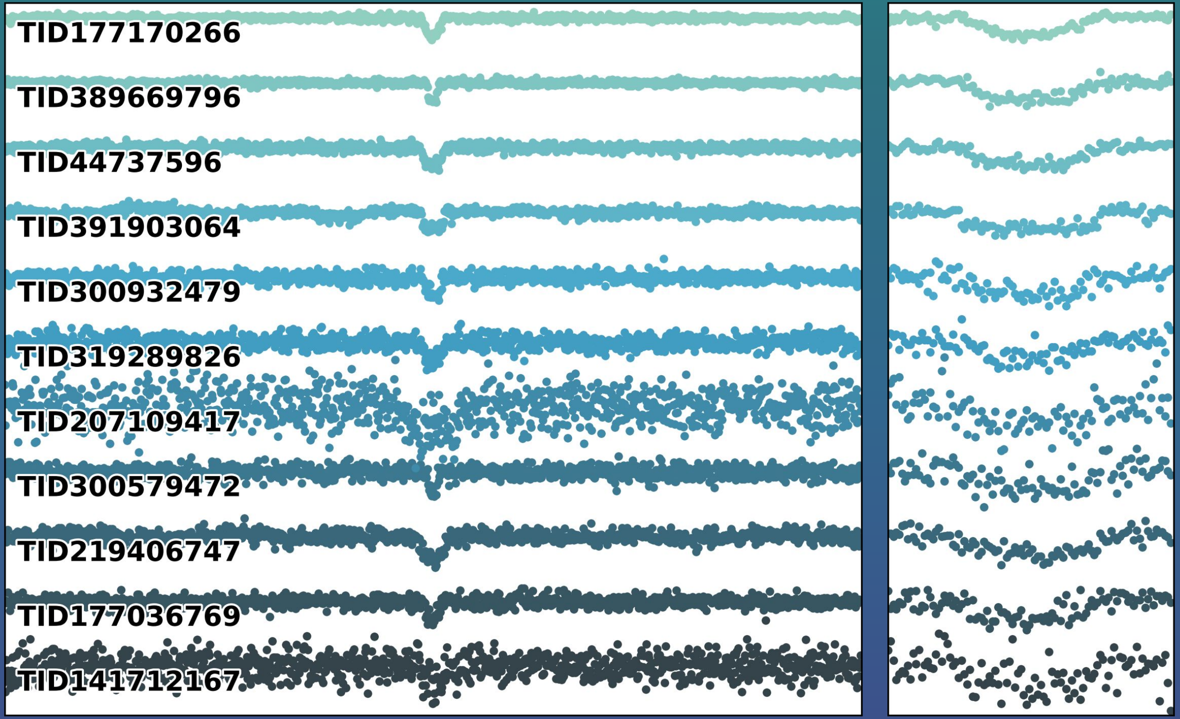
TOIs from QLP FFI pipeline not in our candidate list

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

@exohugh

Classification with Machine Learning


Classes >
Samples ^
< Ground truth
Bill
Claire
Frans
Alfred
Herman
Maria
Alex
Max
Anne
Philip
Tom
Richard
Robert

Decision Tree Classifier

Sample to be classified

Are they an astronomer?

Yes / No

Do they have a beard?

Yes / No

Hair longer than 5cm?

Yes / No

German?

Yes / No

Dan

(Hugh, etc)

^ Predicted class

Decision Tree Classifier

Sample to be classified

Does it have a secondary eclipse

Yes

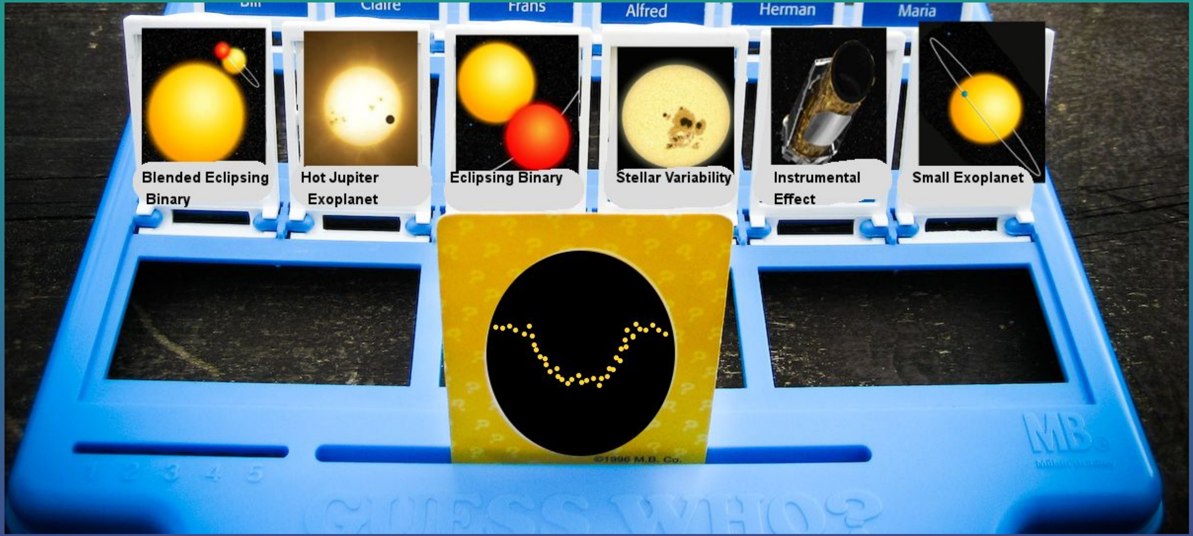
No

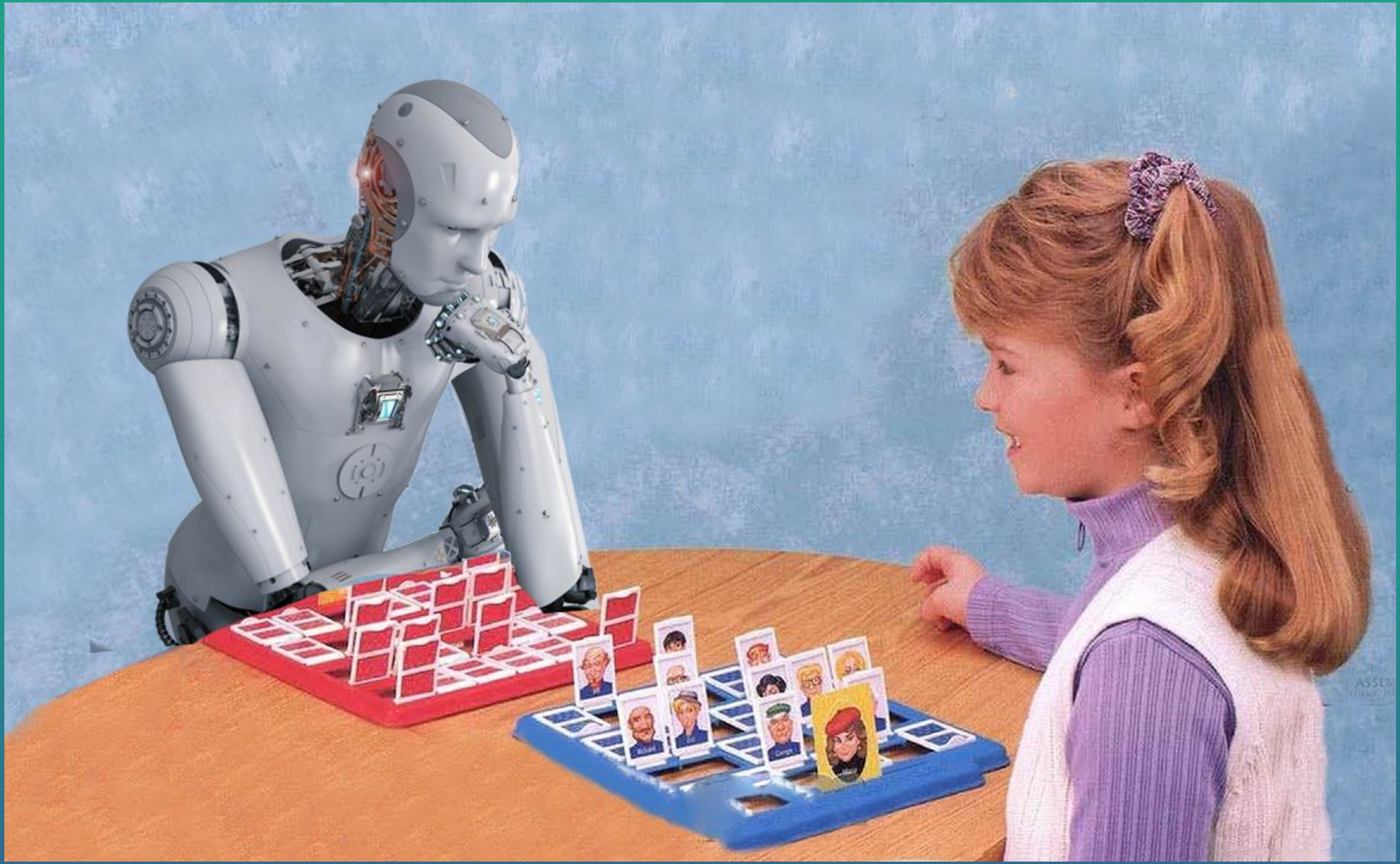
Is the modelled albedo >1

Yes

No

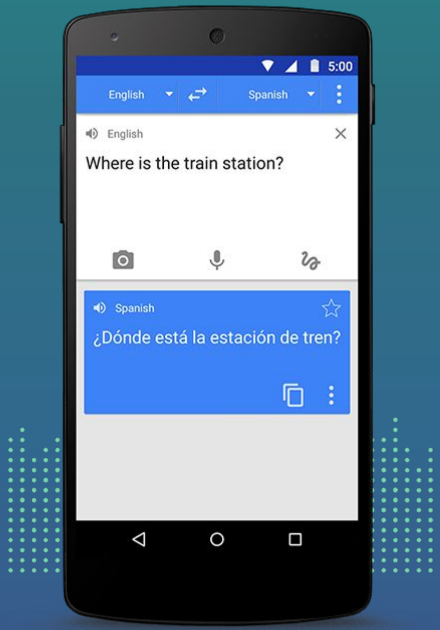
Eclipsing Binary

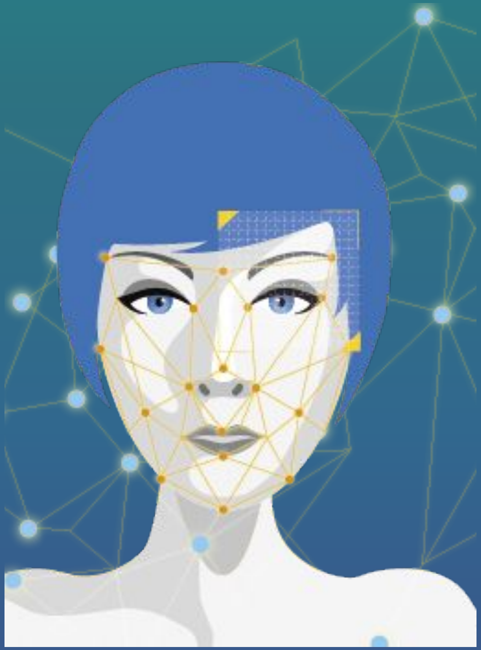


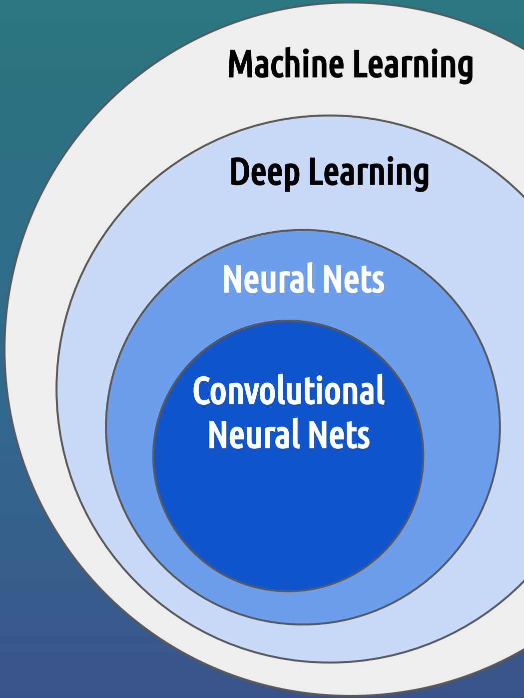


How can we classify with minimal human processing?
With Machine Learning

Machine Learning







Translation

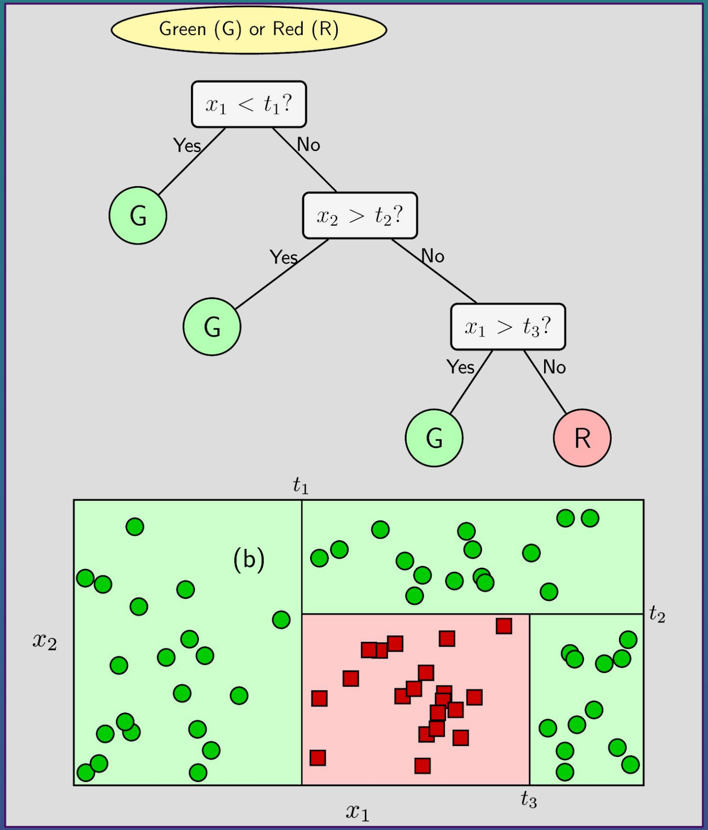
Self-driving cars

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



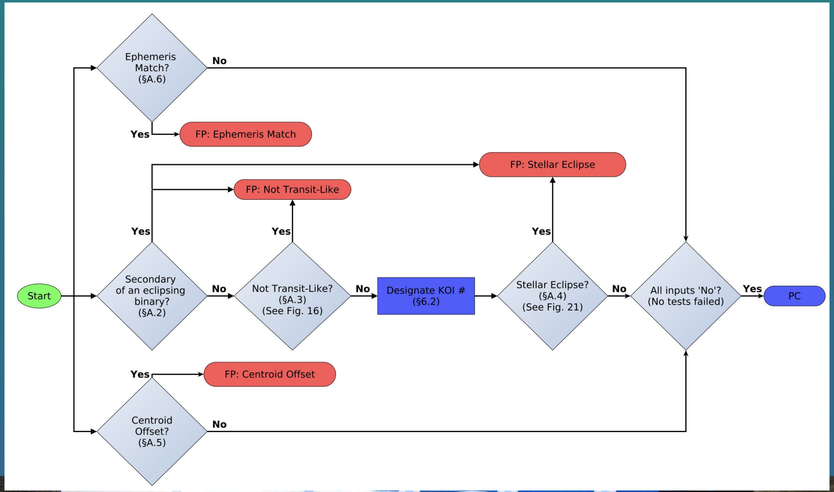
Robovetter - Decision Tree

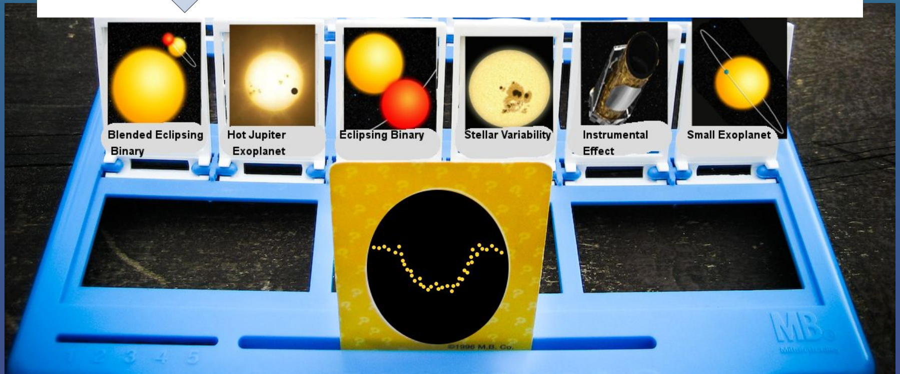
"Robovetter" – Thompson et al 2017.

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

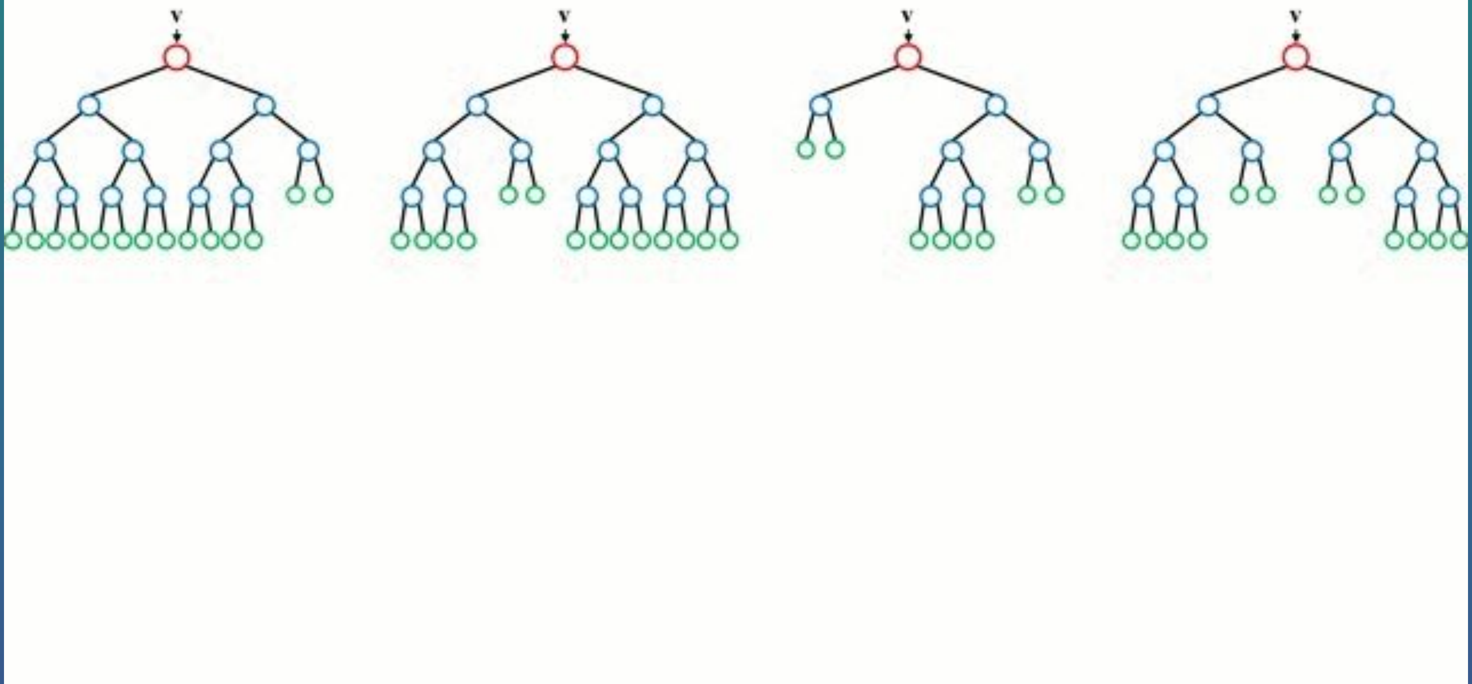
Used features processed from lightcurve.

Achieved a recall of around 80% on injected data.





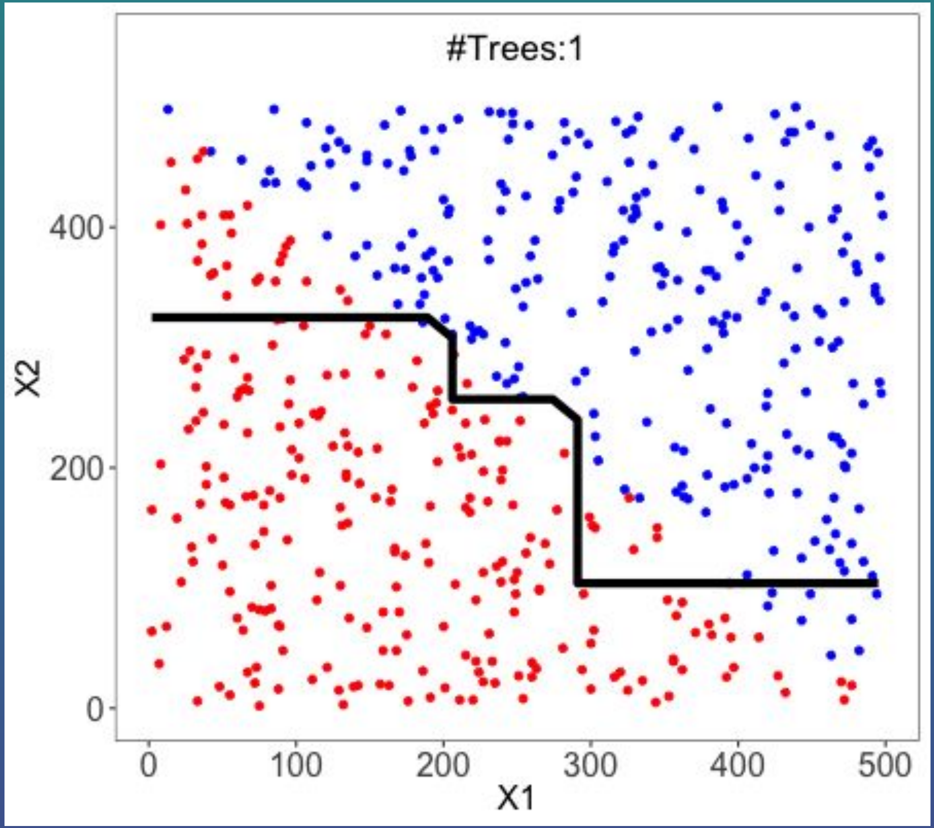
Random Forests



- 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.



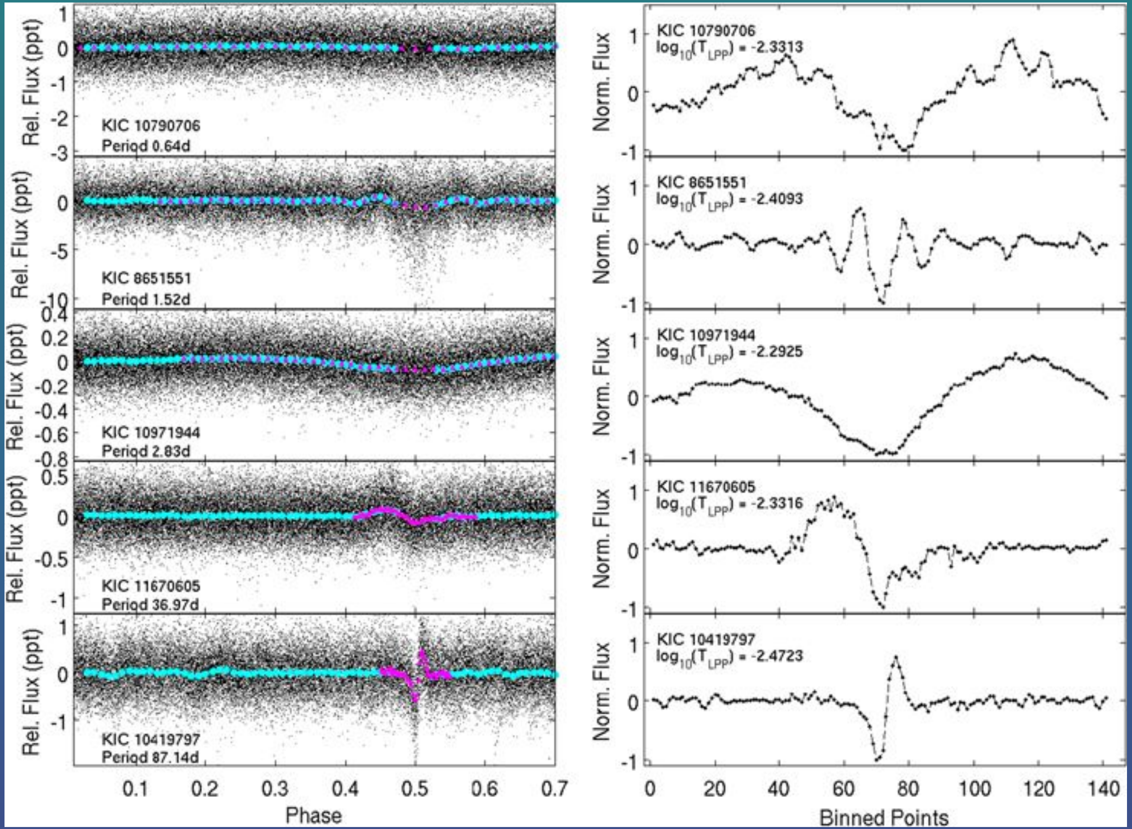
Exoplanet Classification with KNNs

Thompson et al (2015).

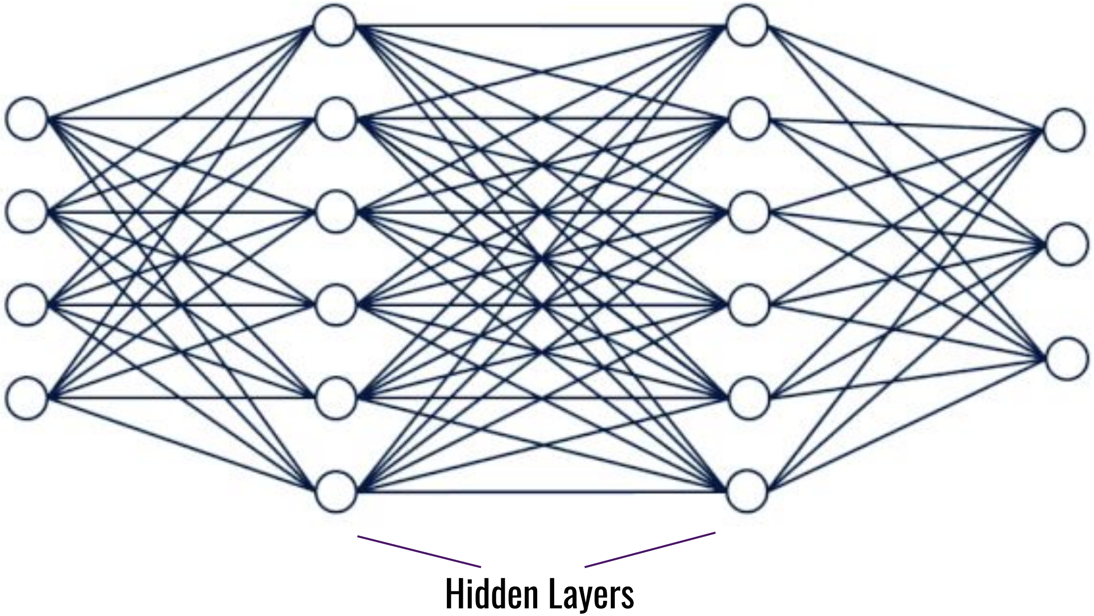
Used a “K-Nearest Neighbours” (KNN) unsupervised approach.

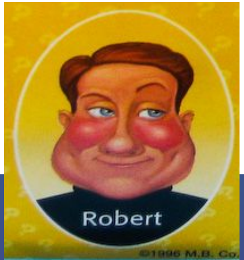
Takes average of nearest labelled features.

Used as inputs binned & normalised phase-folded transits.



Neural Networks

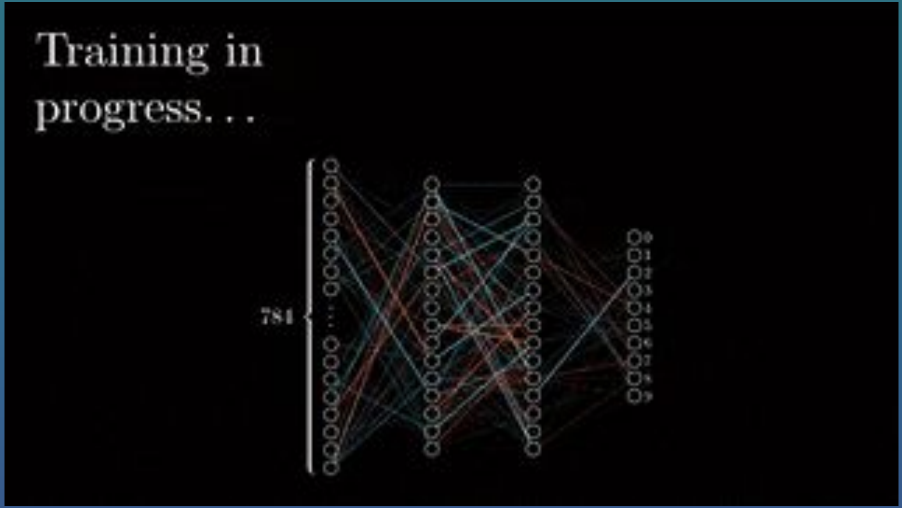




Neural Networks

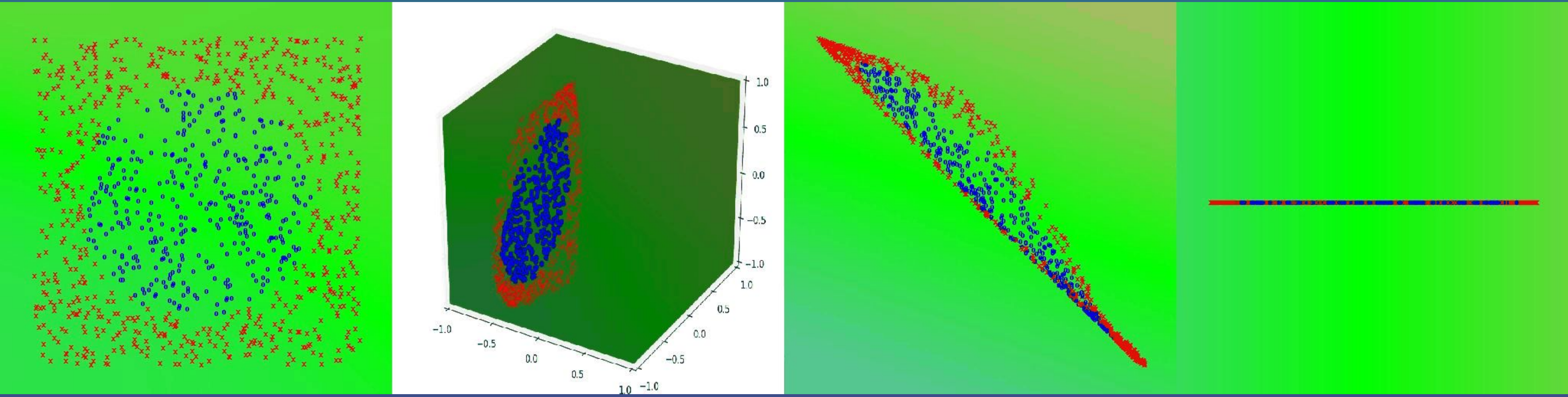
Training neural networks

- 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)

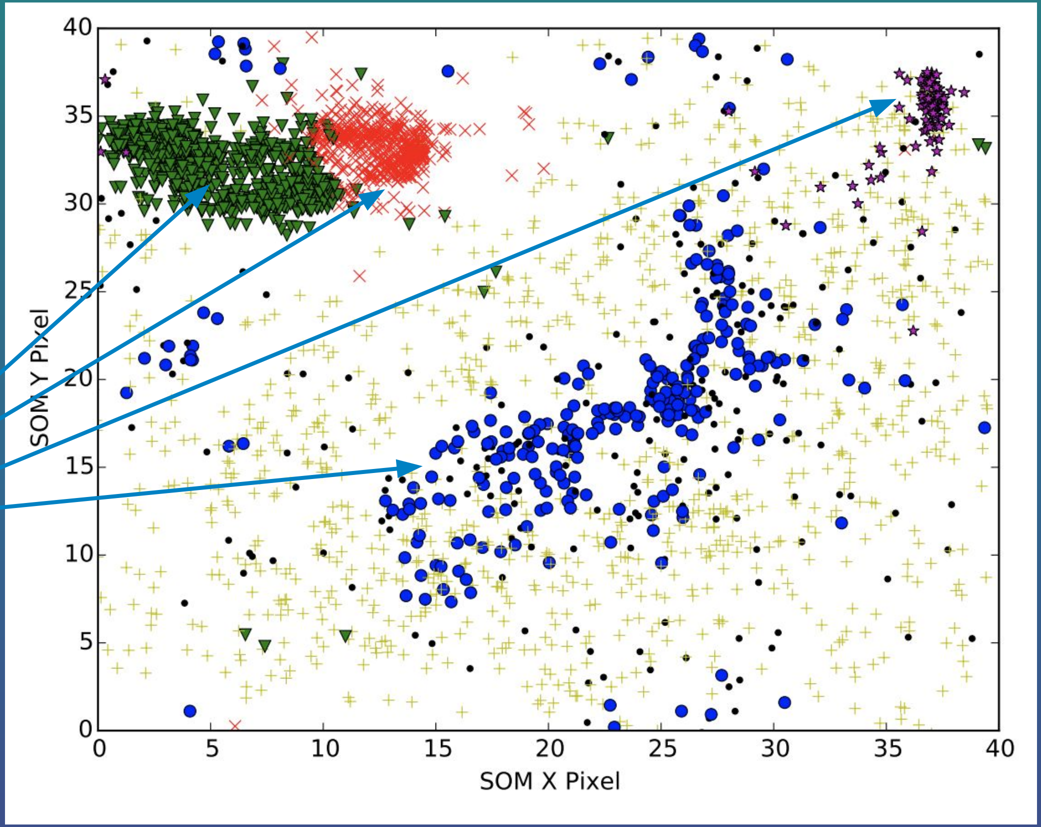
Armstrong et al, 2016

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 an input into Random Forest.



Self Organising Maps for Exoplanets

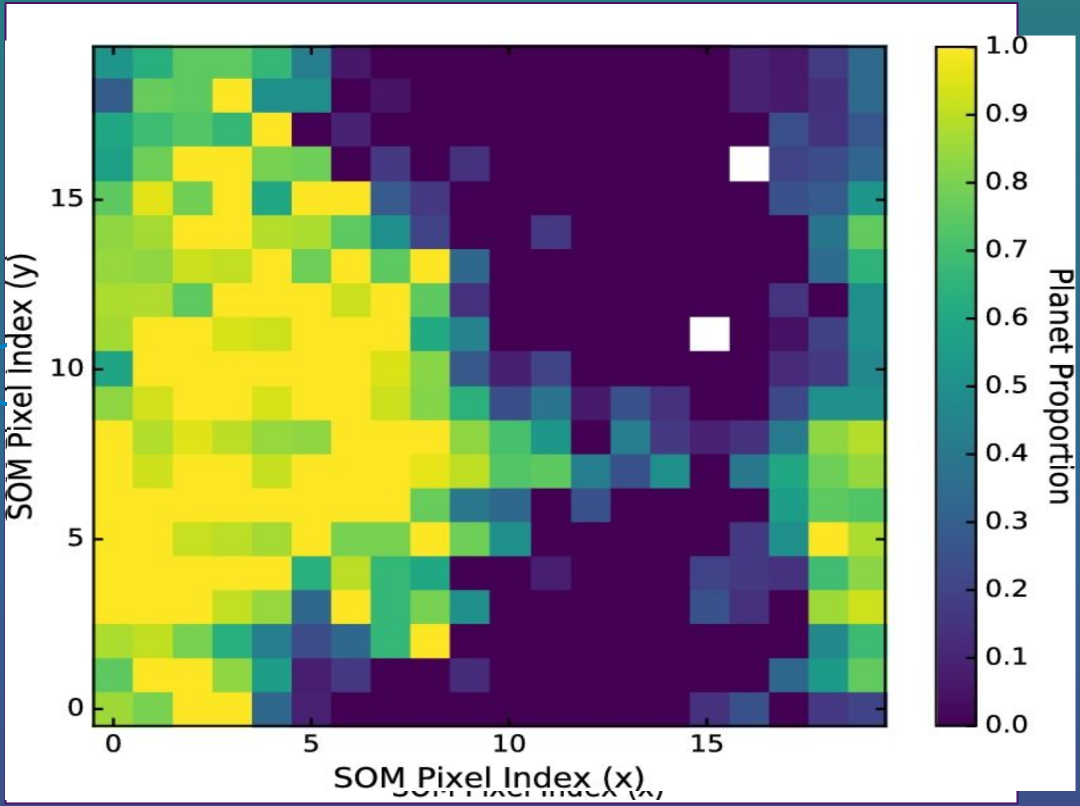
Armstrong et al (2017)

SOM and random forest applied to Planet candidates in K2 & Kepler

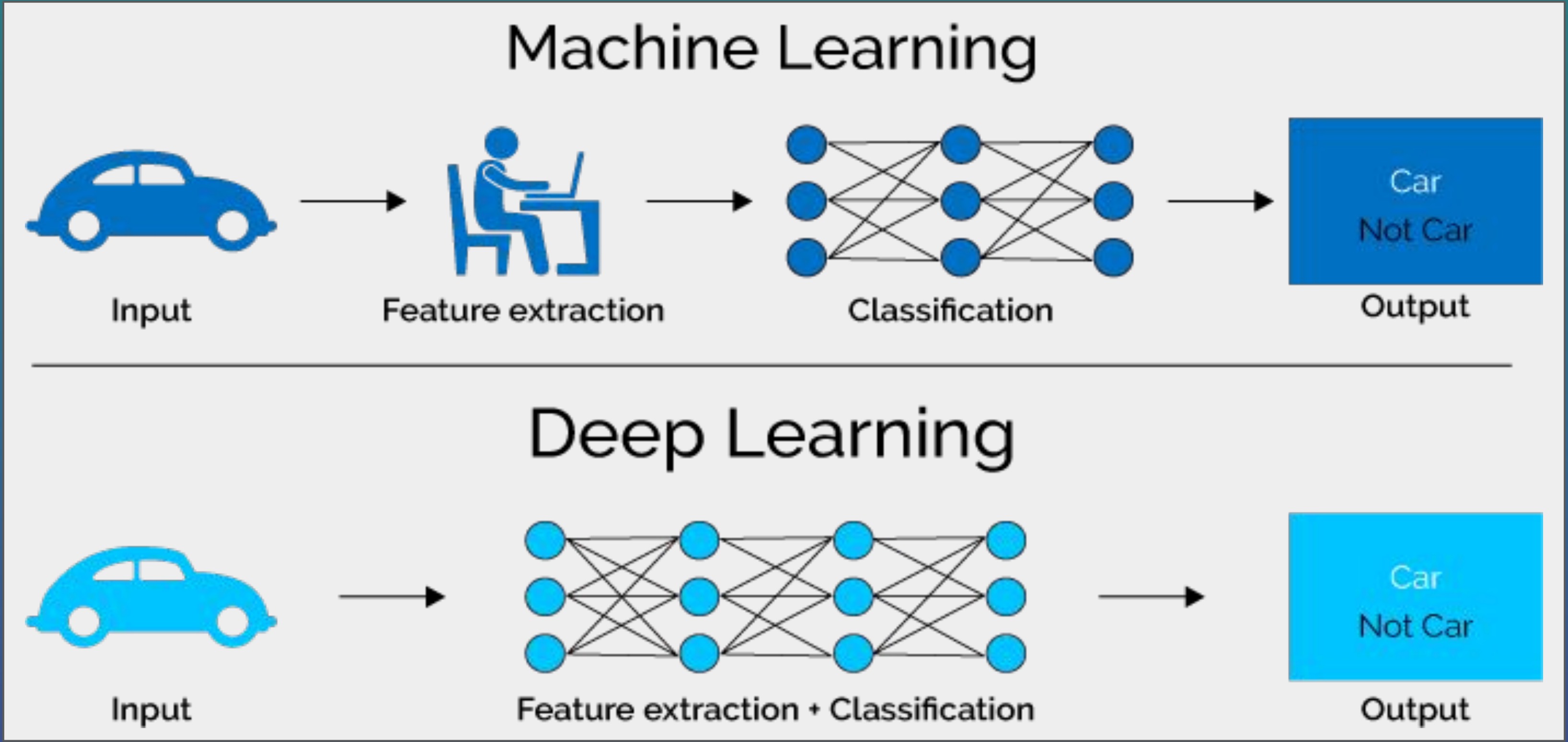
~79% accuracy on Kepler planets

EBs

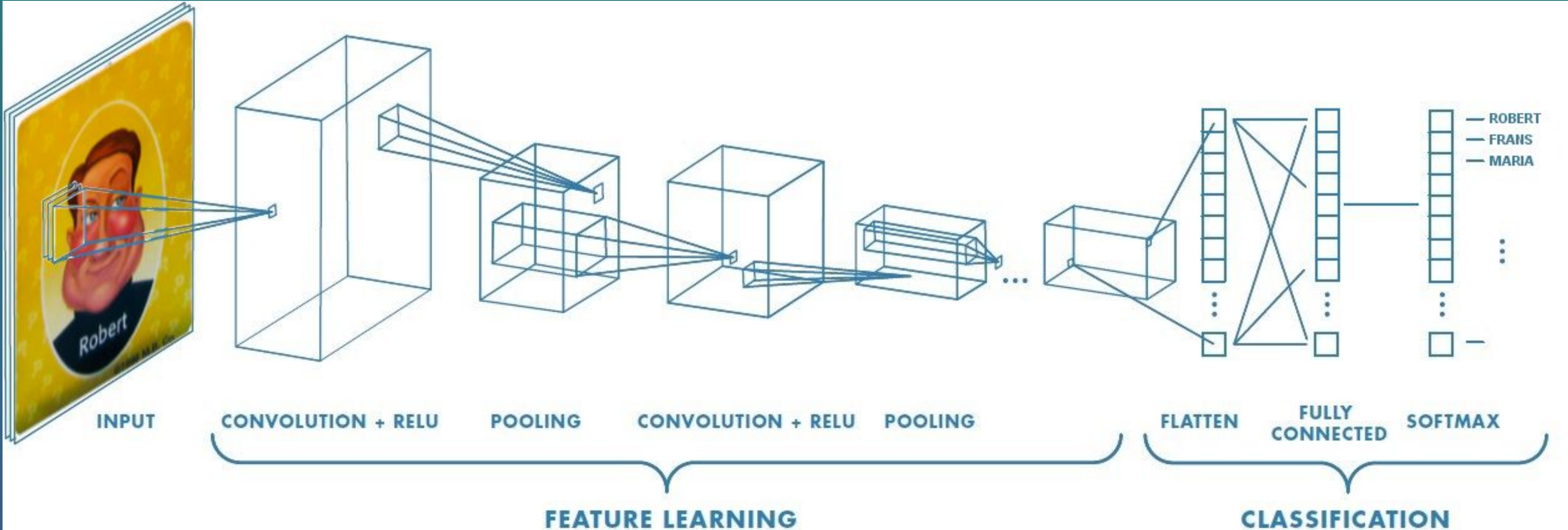
Planets



Convolutional Neural Networks



Convolutional Neural Networks

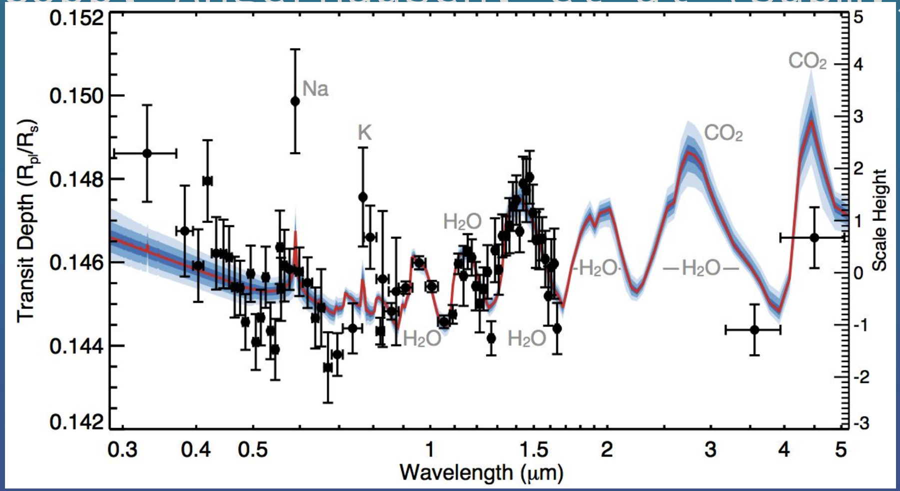


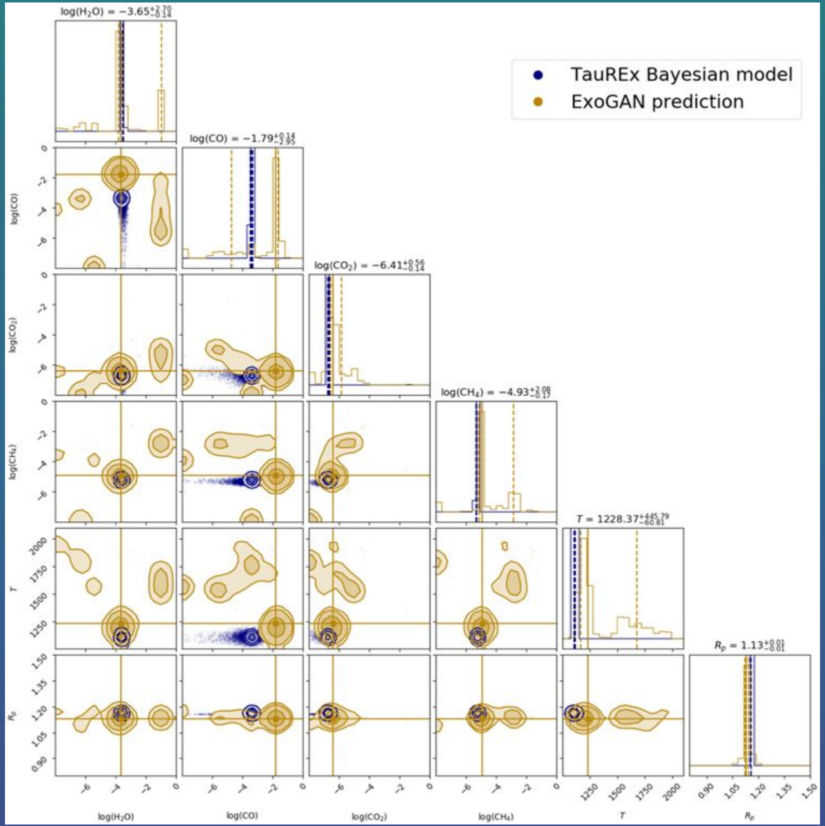
- Raw image “convolved” with range of filters (which themselves are trained with back propagation)
- Enables Feature extraction from the raw data (although raw

CNNs for Atmospheric Retrieval

Waldmann (2015) & Zingales (2018) - RoBErt using Neural networks

Cobb, Angerhausen, et al (subm.)



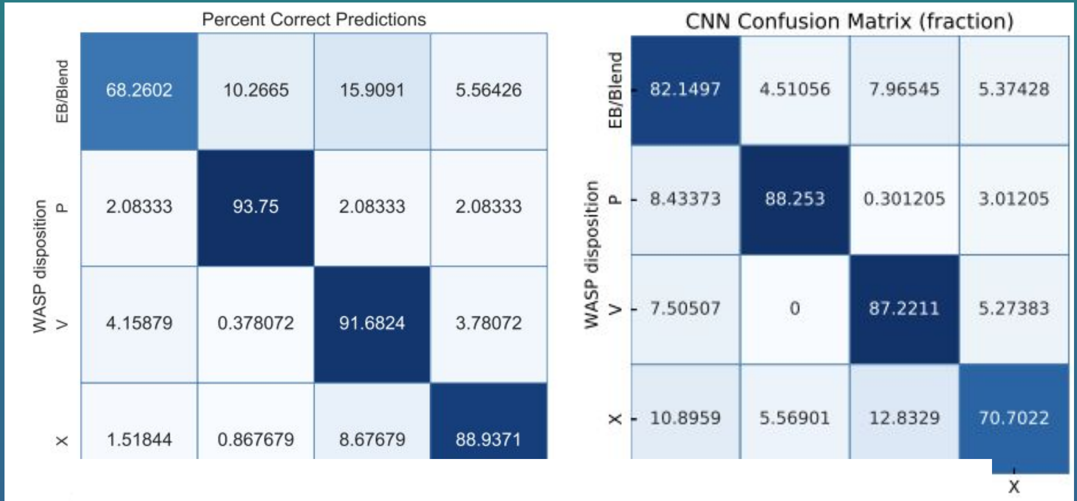


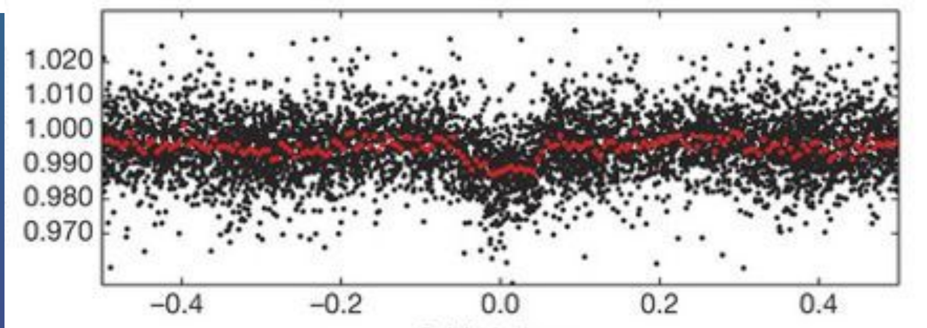
Ground-based transits with RF & CNNs

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

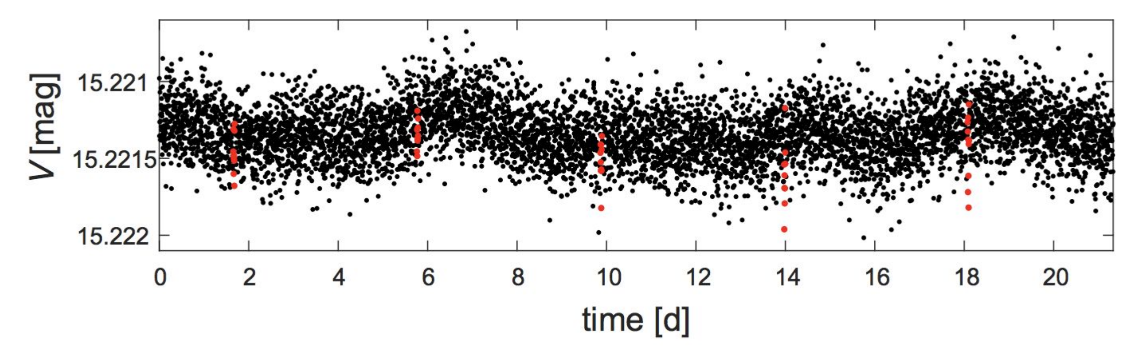
Two parallel papers using neural networks to detect exoplanets:

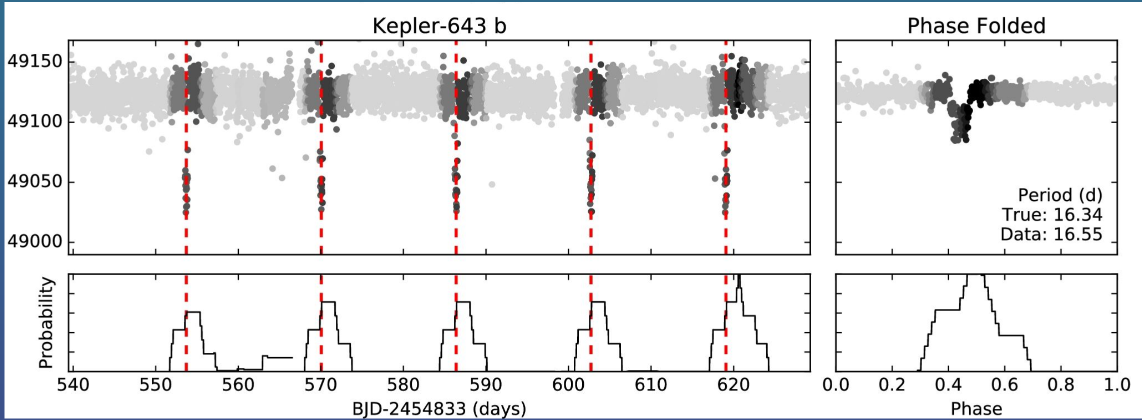
Zucker et al, (2017)

Pearson et al, (2017)

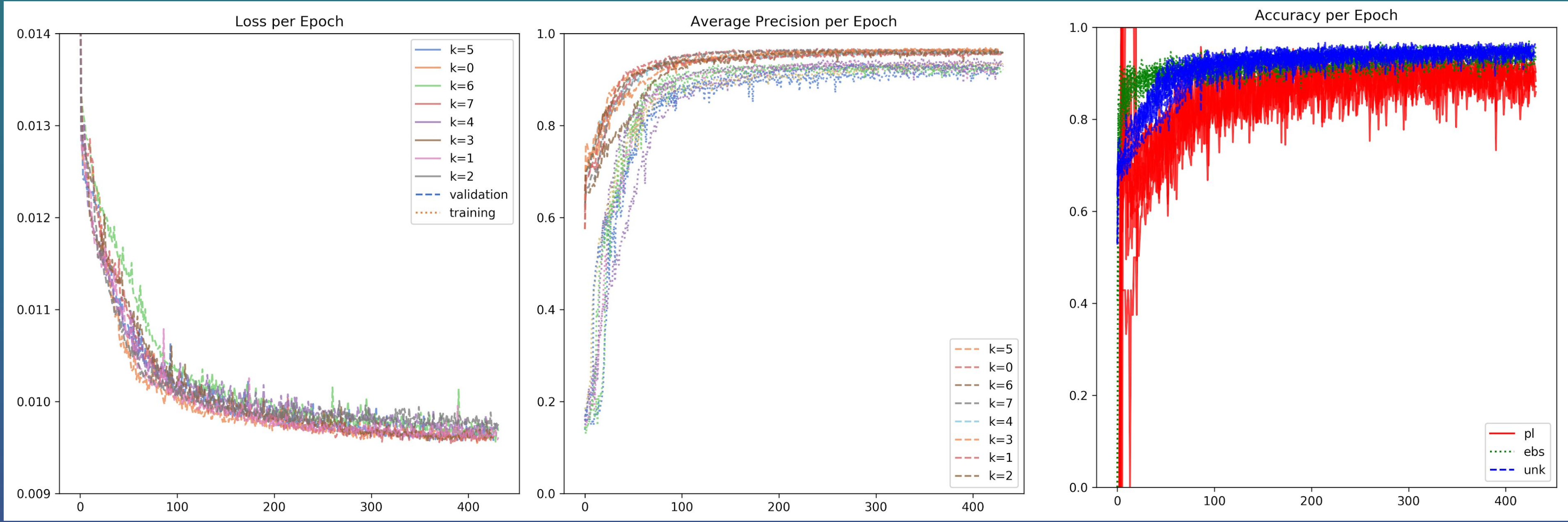
Difficult as neural networks cannot natively learn “periodicity”.

Neither deal with classifying real planets vs false positives





Classifying TESS Simulations



Osborn, Ansdell, Ioannou, Sasdelli, et al. (subm)