A&A 633, A53 (2020)
https://doi.org/10.1051/0004-6361/201935345

Astronomy & Astrophysics

Rapid classification of TESS planet candidates with convolutional neural networks⋆

H. P. Osborn[1,⋆⋆], M. Ansdell[2,⋆⋆], Y. Ioannou[3,⋆⋆], M. Sasdelli[4,⋆⋆], D. Angerhausen[5,6,⋆⋆⋆], D. Caldwell[7,8,⋆⋆⋆], J. M. Jenkins[7,⋆⋆⋆], C. Räissi[9,⋆⋆⋆], and J. C. Smith[7,8,⋆⋆⋆]

[1] Aix-Marseille Université, CNRS, CNES, Laboratoire d'Astrophysique de Marseille, France
e-mail: hugh.osborn@lam.fr
[2] Center for Integrative Planetary Science, University of California at Berkeley, Berkeley, USA
[3] Machine Intelligence Lab, Cambridge University, Cambridge, UK
[4] Australian Institute for Machine Learning, University of Adelaide, Adelaide, Australia
[5] Center for Space and Habitability, University of Bern, Bern, Switzerland
[6] Blue Marble Space Institute of Science, Seattle, USA
[7] NASA Ames Research Center, California, USA
[8] SETI Institute, California, USA
[9] Institut National de Recherche en Informatique et en Automatique, Rocquencourt, France

Received 22 February 2019 / Accepted 8 April 2019

ABSTRACT

Aims. Accurately and rapidly classifying exoplanet candidates from transit surveys is a goal of growing importance as the data rates from space-based survey missions increase. This is especially true for the NASA TESS mission which generates thousands of new candidates each month. Here we created the first deep-learning model capable of classifying TESS planet candidates.
Methods. We adapted an existing neural network model and then trained and tested this updated model on four sectors of high-fidelity, pixel-level TESS simulations data created using the *Lilith* simulator and processed using the full TESS pipeline. With the caveat that direct transfer of the model to real data will not perform as accurately, we also applied this model to four sectors of TESS candidates.
Results. We find our model performs very well on our simulated data, with 97% average precision and 92% accuracy on planets in the two-class model. This accuracy is also boosted by another ∼4% if planets found at the wrong periods are included. We also performed three-class and four-class classification of planets, blended and target eclipsing binaries, and non-astrophysical false positives, which have slightly lower average precision and planet accuracies but are useful for follow-up decisions. When applied to real TESS data, 61% of threshold crossing events (TCEs) coincident with currently published TESS objects of interest are recovered as planets, 4% more are suggested to be eclipsing binaries, and we propose a further 200 TCEs as planet candidates.

Key words. planets and satellites: detection – methods: analytical

1. Introduction

In the next two years, the NASA Transiting Exoplanet Survey Satellite (TESS) mission (Ricker et al. 2014) is likely to more than double the number of currently known exoplanets (Sullivan et al. 2015; Huang et al. 2018a; Barclay et al. 2018). It will do this by observing 90% of the sky for up to one year and monitoring millions of stars with precise-enough photometry to detect the transits of extrasolar planets across their stars (e.g. Huang et al. 2018b; Vanderspek et al. 2019; Wang et al. 2019). Every ∼27.1 day "sector" monitors the light of tens of thousands of stars which are then compiled into 1D "light curves", detrended for instrumental systematics, and searched for signals similar to transiting planets. However, those signals with exoplanetary origin are dwarfed by signals from false positives – those from artificial noise sources (e.g. systematic errors not removed by detrending), or from astrophysical false positives such as binary stars and variables. The best way to classify exoplanetary signals is therefore a key open question.

Answers until now include human vetting, both by teams of experts (Crossfield et al. 2018) or members of the public (Fischer et al. 2012), vetting using classical tree diagrams of specific diagnostics (Mullally et al. 2016), ensemble learning methods such as random forests (McCauliff et al. 2015; Armstrong et al. 2018), and deep learning techniques such as neural networks (Shallue & Vanderburg 2018; Schanche et al. 2018; Ansdell et al. 2018). The current process of vetting TESS candidates involves a high degree of human input. In Crossfield et al. (2018), 19 vetters completed the initial vetting stage of 1000 candidates or threshold crossing events (TCEs), with each candidate viewed by at least two vetters. However each TESS campaign has so far produced more than 1000 TCEs, and a simple extrapolation suggests as many as 500 human work hours may be required per month to select the best TESS candidates.

The first attempts at classification using neural networks have tended to use exclusively the light curve (e.g. Shallue & Vanderburg 2018; Zucker & Giryes 2018). In Ansdell et al. (2018), we modified the 1D light-curve-only neural network

⋆ Full Tables A.1 and A.2 are only available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/633/A53
⋆⋆ NASA FDL 2018 participant.
⋆⋆⋆ NASA FDL 2018 mentor.

approach to candidate classification of Shallue & Vanderburg (2018) to include both centroids and stellar parameters, subsequently improving the precision of classification. In this paper we show results on adapting those models to both simulated and real TESS data, the first time deep learning has been performed for TESS.

2. Datasets

2.1. TSOP-301

As no flight data existed at the start of the project, we relied on multiple end-to-end simulations performed by the TESS team. Three such runs were considered for use: an initial one-sector run named ETE-6 (Jenkins et al. 2018) used for the final TESS mission ground segment integration test, a 2.5-sector run (two whole sectors and a further sector including only the overlap region) named TSOP-280 used for the final validation and verification of the TESS science processing pipeline (Jenkins et al. 2016), and a four-sector run called TSOP-301[1] which was specifically designed to create a test set for machine learning and to characterize detection characteristics of the TESS pipeline. We focused on the TSOP-301 run, which had the most data and the most complete set of simulated features.

TSOP-301 was a full four-sector end-to-end run of the TESS science processing pipeline. To help facilitate the development of the science processing operations center (SPOC) pipeline, it was necessary to produce simulated flight data with sufficient fidelity and volume to exercise all the capabilities of the pipeline in an integrated way. Using a physics-based TESS instrument and sky model, the simulation tool named *Lilith* (Tenenbaum et al., in prep.) creates a set of raw TESS data which includes models for the charge-coupled devices (CCDs), readout electronics, camera optics, behaviour of the attitude control system (ACS), spacecraft orbit, spacecraft jitter and the sky, including zodiacal light, and the TESS input catalog (TIC). The end product is an array of expected instrumental artifacts and systematic errors (e.g. cosmic rays, CCD readout errors, thermal-induced focus errors, and spacecraft jitter-induced pointing errors). The model also incorporates realistic instances of stellar astrophysics, including stellar variability, eclipsing binaries, background eclipsing binaries, transiting planets, and diffuse light.

This simulated raw image dataset is then passed through the SPOC pipeline providing full integration tests of the science processing from raw pixel calibration, to transiting planet search (Jenkins et al. 2010; Seader et al. 2013), to the generation of archivable data products such as tables of TCEs and data validation products (Twicken et al. 2018; Li et al. 2019). Full instrumental and astrophysical ground truth is generated for each *Lilith* run and can be used as a training set.

In TSOP-301 we simulated four sectors using the then-current TESS target management and selection with the use of version 6 of the TIC. There were 16 000 targets per sector and many targets were near the ecliptic pole resulting in many targets being observed for more than one sector. Realistic planet distributions based on current understood planet populations were not used, and instead a distribution was generated with good overlap with the desired TESS planet detectability in order to provide a machine-learning classifier with a good distribution of signals to train on. A fraction of 20% of all targets had planetary transits, the distributions for which are seen in Fig. 1. An additional 20% had eclipsing binaries (EBs) or background eclipsing binaries (BEBs) in order to give the classifier a good set of potential astrophysical false positives. Using appropriate dataset balancing (i.e. Sect. 3.2), this difference should have minimal effects on the performance of a deep-learning model.

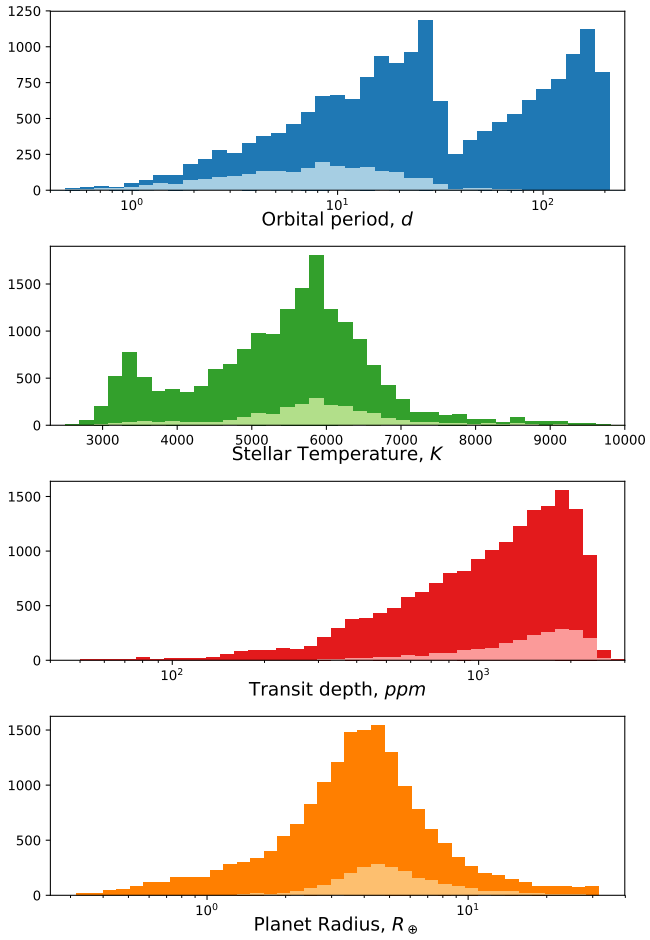

Fig. 1. Distribution of injected planet signals as a function of key inputs. The split distribution in the *upper plot* is due to an injected distribution into the multi-sector regions that was flat in linear period space. The lighter colour shows those injections present in TCEs.

2.2. Pre-processing

TESS light curves were pre-processed in a method similar to that of Shallue & Vanderburg (2018) and Ansdell et al. (2018), iterating over each TCE to produce binned phase-folded "global" (full light curve showing the entire phase between −0.5 and 0.5) and "local" (zoomed in on the transit between $-2.5t_{\rm dur}$ to $2.5t_{\rm dur}$) views of both the light and centroid curves (see Fig. 3 in Shallue & Vanderburg 2018 or Fig. 1 in Ansdell et al. 2018). A light curve from which instrumental systematic errors have been removed is compiled for each target before it is searched for transiting planets (the so-called pre-search data conditioning, or PDC, light curve), and then a light curve with all non-planetary signals detrended is produced after a candidate detection has been found (the data validation, or DV, light curve). Unlike Shallue & Vanderburg (2018) and Ansdell et al. (2018), which used exclusively the PDC light curves, we used both of these time series. These were both accessed from the TESS MAST pages[2].

[1] Called "TSOP-301" from the TESS operations issue tracking ticket which initiated the run.

[2] https://archive.stsci.edu/tess/bulk_downloads.html

The DV time series contain unique time series for each candidate planet, with flux during the transits of previously detected TCEs removed. We took the detrended (LC_DETREND) DV light curve where these were available, using the initial (LC_INIT) light curve if not. Centroid information is found exclusively in the PDC files, and comes in two types – PSF centroids (which are calculated using a model of the TESS point spread function) and the MOM centroids (which are simply the weighted centre of light within the TESS aperture). We extracted row and column PSF centroids where available, as these are typically more accurate, but reverted to MOM when these were missing. In both cases, the median was subtracted giving relative x- and y-shifts in the centre of light.

Anomalies greater than 3.5-sigma from surrounding points are also removed from each time-series. Both time series were then phase folded and median binned into global and local views. We primarily use the DV light curve for the final "views" of the TCEs, however in some cases the gaps around previously detected transits cause large gaps in the final views. We pick a threshold of 50%, above which the PDC light curve views are instead used. These light curves are then normalised using the detected depth such that the median out-of-transit is 0.0 while the transit depth is at −1.0. The row and column centroids are first added in quadrature, and then also phase folded and binned into global and local views. To normalise the centroids, the out-of-transit median is subtracted, the time series is then inverted (to match the "flux drop" of a transit), and finally it is multiplied by the ratio between the out-of-transit RMS in the (normalised) light curve and this new centroid curve. This is done to turn centroid curves without significant deviations into flat lines (rather than amplify low-significance structure).

Some TCEs remain with large numbers of gaps in the phase-folded views (due to detection near or on gaps in the light curve), which is problematic as not-a-numbers ("NaNs") are undifferentiable, and therefore cause immediate errors when present in data seen by a machine-learning model. Models with NaNs and without anomalies removed initially struggled to train, likely as a result of too many objects with missing data. To avoid this, we remove 4577 TCEs for which more than 25% of the local view is missing. Only 4% of these constituted planetary candidates, therefore the overall fraction of planets actually increased due to this cut. We also filled any gaps missing data in the remaining views with zeros (which matches the median out-of-transit value), although our white-noise augmentation step (see Sect. 3.4) means the model sees Gaussian noise for these missing values.

2.3. Stellar and transit parameters

The neural networks will classify data using the shape and distribution of the input transit data. However, extra information can be found by using other parameters which may also help classification. This includes stellar parameters, which testing in Ansdell et al. (2018) showed provided a boost of around 0.5% in accuracy for planet classification (potentially as a result of identifying large stars unlikely to be planet hosts). However, the planetary injections performed by *Lilith* effectively choose random stars rather than following any physical correlations (such as trends in planet occurrence with metallicity or stellar mass), and therefore stellar parameters are unlikely to provide as big a boost. Some transit phenomena may also not be represented in the light curve data but may aid classification, the most obvious being depth and duration – both an overly deep and an overly long eclipse may suggest an eclipse of two similar-sized objects. However, one or both of these are lost during global and local view normalisation. We therefore added the following additional data: from the transit search: the orbital period, transit duration, the semi major axis scaled to stellar radius (a/R_s), the number of transits $N_{\rm TRANS}$, the transit S/N, and the transit depth and ingress duration. Derived from the transit model fit parameters we added the radius ratio R_p/R_s, the impact parameter b, the ratio of the maximum multiple event statistic (MES, a proxy for S/N; Jenkins et al. 2002) to the expected MES from the single event statistic (SES; i.e. the S/N of the strongest signal) SES $\sqrt{N_{\rm trans}}$, the logged planet radius divided by an arbitrary planetary boundary (set at 13 $R_\oplus$), and the logged ratio of the transit duration over the expected duration for a planetary transit given the stellar density and orbital period. Furthermore, from stellar parameters we added the TESS band magnitude, stellar radius, total proper motion, stellar log g, stellar metallicity, and stellar effective temperature. We took these values from the DV light-curve fit headers provided for each TCE.

All these additional data were then normalised by subtracting the median and dividing by the standard deviation.

2.4. Labels

Unlike for real flight data, the ground truth of our simulated TESS dataset is known precisely. However, the injected signals are never recovered perfectly during transit search – some may be found at the incorrect period, or with incorrect durations, and so on. Therefore, the degree of correlation between the injected signal and the recovered TCE must be computed – we adapted the code of the TESS team which sums in quadrature the number of cadences that overlap between the in-transit (or in-eclipse) points from an injection and those from the detection, setting a threshold of 0.75.

We split eclipsing binaries into their primary and secondary dips, therefore recovering both signals. We also searched for injections recovered at an integer multiple of the real period, finding a handful of equal-depth eclipsing binaries detected at half the real period. Although complex labels were generated for each target (e.g., EB_secondary or BEB_at_2P), we collated all labels from the same source to give between four (planet, eclipsing binary, background eclipsing binary, non-astrophysical signal) and two (planet and not planet) labels, depending on the model used.

2.5. TESS sectors 1 to 4

At the point of submission, data from four TESS sectors have been released. Catalogues of TCEs have been compiled from the ∼16 000 two-minute cadence targets observed for each sector[3]. In total, this gives 7562 TCEs from 3266 unique TESS IDs, which includes duplications between sectors. Of these, 370 have been published as TESS objects of interest (TOIs). Their identification comes from candidates identified by the "quick-look pipeline" (QLP, Fausnaugh et al. 2018) which are then manually vetted in the manner of Crossfield et al. (2018).

The TESS light curves were processed in the same way as for the simulated data (see Sect. 2.2). We also performed the same removal of light curves that had more than 20% of points in either of the phase-folded views missing. This led to the removal of 2197 candidates.

[3] http://archive.stsci.edu/tess/bulk_downloads/bulk_downloads_tce.html for sectors 1–3, and we took the information in the released DV light curves to build a TCE catalogue for campaign 4.

Despite being generated from the pixels with realistic noise sources, the simulated data are unlikely to be identical to the real data in some key ways, especially in terms of unexpected systematic noise sources. This likely includes the second orbit of sector 1 which has higher than average systematic noise due to unexpected noise in fine pointing. However, some injected noise sources have been identified as not present in the real data, such as the sudden pixel sensitivity dropouts (SPSDs) which were present in *Kepler*.

3. Machine learning models

3.1. Architecture

In `Astronet` (Shallue & Vanderburg 2018) and `exonet` (Ansdell et al. 2018), a series of convolutional layers are applied to the local and global views, with the larger global view having a deeper convolutional structure. These are then combined together as inputs for a series of fully connected (FC) layers (equivalent to a linear Artifical Neural Network, or ANN, layer) before outputting a single estimate of class membership. This estimate, between 0 and 1 for each input sample, may be naively thought of as an estimate of the probability of class membership, however without priors and Bayesian methodology, it should not be directly taken as a measure of planetary probability. Figure 2 gives an overview of the model architecture.

We maintained the convolutional filter sizes and architecture from `Astronet`, with four 1D convolutional layers for the local view, and eight for the global view. Max pooling is performed every two layers to reduce the overall size of the tensor. With the number of input data points shrunk by a factor of two (see Sect. 3.5), the final fully connected layers were similarly shrunk from 512 to 256 neurons. The dimensionality of the output depends on the model loss function, with either a single class probability estimate per object (binary) or an estimate per class per object (multi-class).

For binary models, the binary cross entropy loss function (`BCEloss` in `pytorch`) was used, whereas for multi-class models, a cross entropy loss (`CrossEntropyLoss` in `pytorch`) function was used. For gradient descent, we used Adam (Kingma & Ba 2014) as an optimizer with a starting learning rate of around 2×10^{-5}.

In all cases, we trained until the output of the loss function, when applied to validation data, had stopped decreasing; a sign that the model is well-fitted but has not yet begun to be overfitted. This was between 200 and 500 epochs, depending on the learning rate and number of classes used.

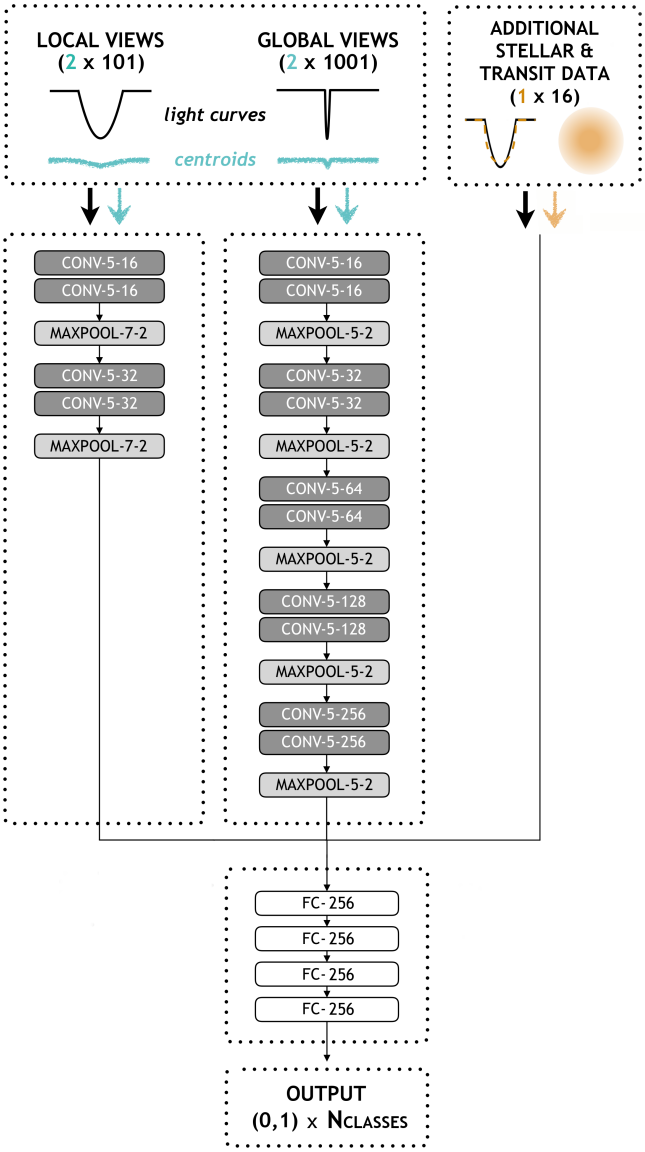

Fig. 2. Convolutional neural network architectures used in this work. "CONV": a 1D convolutional layer, with the two following numbers referring to the kernel size and the number of filters; "MAXPOOL" refers to the process of 1D max pooling the tensor, and the numbers refer to kernel size and stride; "FC" is a fully connected or linear ANN layer where the number shows the number of neurons.

3.2. Balanced batch sampling

Training a neural network using a dataset with an unbalanced class distribution is difficult (see, e.g. Chawla et al. 2004), since the learning algorithm inevitably biases the model towards learning the majority class. In the case of the *Kepler* dataset used in Ansdell et al. (2018), the two classes (planet and non-planet) were more closely balanced than the data here. This was partly because candidates labelled as "unknown" by human vetters (Batalha et al. 2013; Burke et al. 2014; Mullally et al. 2015; Rowe et al. 2015) were classified and removed from the DR24 sample. However, such a step is not available with our TESS dataset, hence only 14% of the TCE dataset are planets. It is therefore necessary to perform dataset balancing in order to train the network. We took an approach that involves resampling the input data (rather than, e.g. weighting the loss function). We did this by balancing the mini-batches used in training, meaning each training epoch sees an equal number of samples from each class (see, e.g., He & Garcia 2008).

3.3. Cross validation

To test the model while retaining as much of the data as possible for training, we used cross validation. This splits the data into k parts, and independently trains such that a different subsection of data is kept as validation data each time, while $(k-1)$ parts are used for training. We used $k = 8$ for all models here in order to use all available graphics processing units (GPUs).

3.4. Augmentation

Augmentation is the process of modifying training data in order to generate similar but not identical samples, thereby increasing

Table 1. Results on testing different model augmentation and input view sizes.

Model	Avg. precision
201/2001 bins	92.0 ± 0.7%
101/1001 bins	92.7 ± 0.7%
Without white noise	89.6 ± 0.7%
Without phase-inversion	90.4 ± 0.7%
Without phase-shifts	90.5 ± 0.7%
Without any augmentation	85.2 ± 0.7%

Notes. The 101-bin and 201-bin models are with all three methods of data augmentation. Testing of individual augmentation techniques was performed by removing each individual method in turn from the 101-bin model. Four-fold cross validation was used for this testing, and the numbers given are on the validation dataset.

the effective number of samples. This therefore helps preserve against over-fitting. We used three methods of augmentation: white Gaussian noise was added to each light and centroid curve, with the amount chosen randomly between 0 and the out-of-transit RMS of each light- and centroid-curve; a shift of between −5 and 5 bins was applied to the phase; and 50% of all time series were mirrored in phase. These were tested using cross validation on a baseline binary model to assess whether augmentation improved model training, with the results seen in Table 1. It was found that adding each improved the overall precision of the model, with the removal of Gaussian white noise augmentation having the greatest effect (3.1% decrease in average precision, or AP).

3.5. Input array dimensions

For both the `astronet` and `exonet` models, input arrays of 2001 and 201 in size were used for global and local views, respectively. Reasons for this included that long-period planet candidates seen with *Kepler* needed at least a single bin on the global view, and that high-resolution local view allows the in/egress of small planets to be resolved. TESS, which will find shorter-period planets which are on average larger than those of *Kepler*, therefore may not need such wide bins. We tested whether or not reducing the number of bins by a factor of two improved performance with TESS (see Table 1). This shows that a smaller light curve view does indeed improve model performance, likely because increasing S/N in each bin outweighs the effects from low phase resolution in TESS. Halving the number of bins also increases run speed.

4. Results with simulated data

To best assess the accuracy and precision of each model, we performed an "ensemble" or bagging method, taking all eight models trained during cross validation and applying these to the test data taking the mean across all of these eight class membership probabilites. Ensembles typically outperform single models and guard against models which may find local minima (see, e.g. Dietterich 2000). Although different initialisation weights are usually used for each model in an ensemble, we used the same random initialisation weights. However, a test with the binary model confirms that, due to each model seeing different training and validation datasets, there is no difference in performance. We used the 10% of data which were randomly left out of the training/validation set. These results are shown for each model in Table 2.

Table 2. Accuracy, recall, and average precision for the trained model on the test set, using a mean of the estimated class membership probability from all eight ensemble models.

	Accuracy	Recall	Average precision
Binary			97.3%
Planet	91.8%	87.8%	95.2%
Not planet	97.6%	98.5%	99.4%
Three-class			97.1%
Planets	90.4%	90.1%	95.6%
EBs	95.1%	95.1%	96.9%
Unknown	94.8%	94.9%	97.7%
Four-class			96.3%
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%

Notes. For overall average precision, a "micro" average is performed which better accounts for imbalanced datasets.

Here we define accuracy as the fraction of all estimated class members that are correct (TP/(TP + FP), often also called precision); recall as the fraction of all objects of a class which are estimated correctly (TP/(TP + FN)); and average precision as the average accuracy (or precision) for all classes, weighted by the class frequency (a so-called micro average, implemented with `scipy`'s `average_precision_score`)[4].

We find the binary model gives the best planet accuracy, while the three-class model gives both the highest average precision on planets and on all classes.

In Fig. 3 we show a comparison of the ROC (receiver operating characteristic) curves for all three class categories on exclusively planets. These show near-perfect agreement suggesting the addition of other classes does not inhibit a model from differentiating planets. In Figs. 4 and 6 we show the ROC curves for each class in the multi-class, three-class, and four-class models respectively, when applied to the test dataset. We calculate both the median and mean values across all eight ensemble models. Due to the tendancy of class probabilities to cluster near 0 or 1, the median gives a higher precision at more restrictive thresholds (e.g. low recall) while the mean gives higher recall for less restrictive thresholds (e.g. low precision)[5]. In Figs. 5 and 7 we show confusion matrices for the three- and four-class models using cross validation data, including randomly selected local view light curves for each class. In Fig. 8 we compare the performance of recall and accuracy with respect to MES using cross validation data for the four-class model.

5. Application to real TESS data

We directly applied the trained models to those TCEs from the first four sectors of real TESS data (see Table A.1). As well as

[4] The micro-averaged average precision simplifies to (TP + TN)/(TP + TN + FP + FN) in the binary case.

[5] This is most likely due to the sigmoid layer, which distributes class probability estimations close to either 0 or 1. For example, the mean of classifications [0,0,0,0,0,1,1,1] is more inclusive for misidentified candidates (0.375) than a median (0.0).

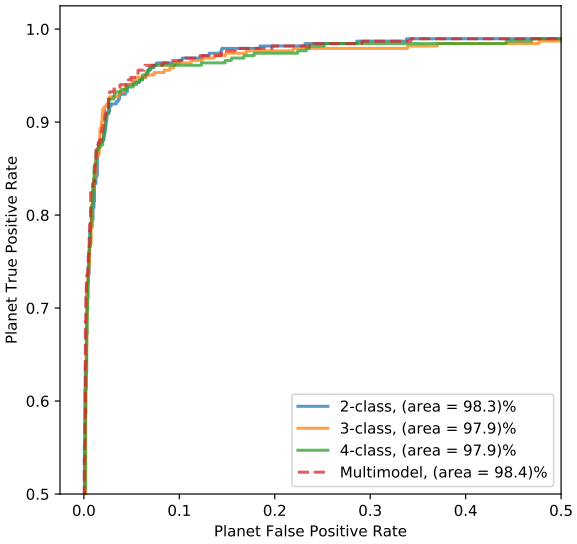

Fig. 3. Receiver operating characteristic curve for planets in all three models. The ROC metric varies the class probability threshold and compares, at each new value, the recall of a particular model (or true positive rate) with its false-positive rate. The curve drawn then allows comparison of classification rates even across models trained with different balances of planets and false positives. Random guessing would produce a diagonal line through the origin to [1,1], which would not be visible on this zoomed figure. Perfect models lie as close as possible to the top-left corner. We show mean-averaged ensemble models for the three model types (binary, three-class and four-class models). We also plot "multimodel" averages which produce an ensemble of the estimated planet probabilities from each sample in all three of those models. The area under each curve (ROC-AUC), a statistic which corresponds to the probability that the correct class outranks the incorrect one, is also shown for each model in the legend.

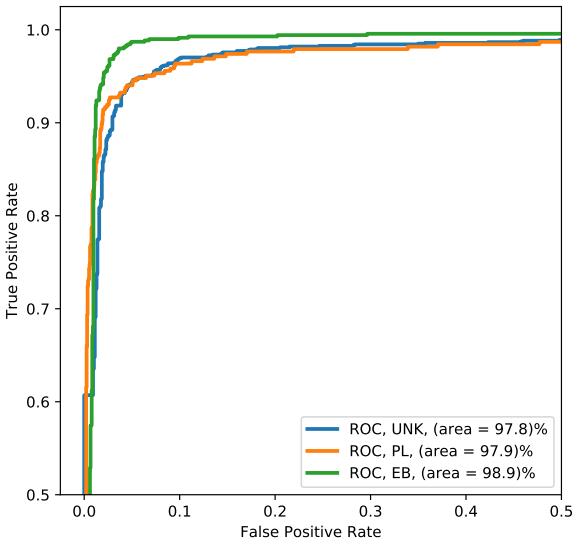

Fig. 4. Receiver operating characteristic curve for our three-class model. "UNK" refers to unknown, or non-astrophysical sources; PL refers to planets; and EB refers to eclipsing binaries.

using the ensemble models from Sect. 4, we also compiled an average planet class

To check how well the model was performing, we loaded the TOIs published so far on the TESS alerts database[6], which is

[6] https://tev.mit.edu/data/, accessed 2019/02/08, a log-in is required.

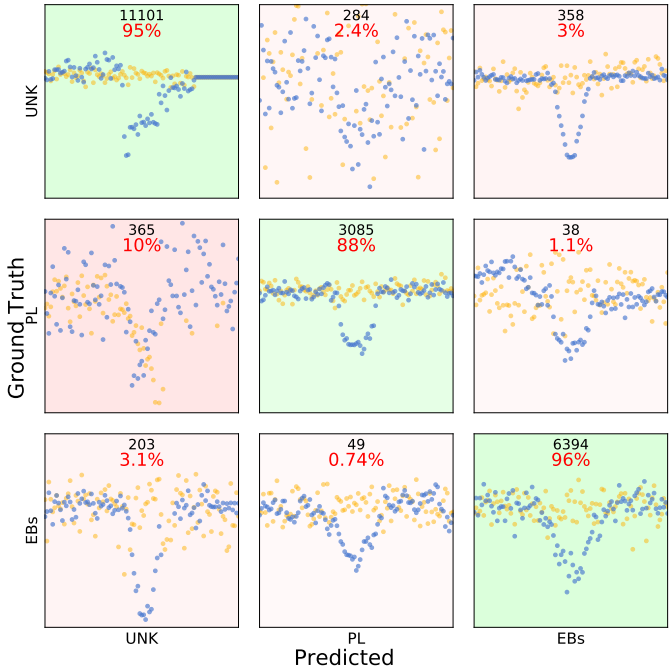

Fig. 5. Confusion matrix, using real data from the cross validation dataset, for the three-class model. Labels as in Fig. 4. The binned and phase-folded input data from a single randomly selected object are shown in each subplot, with the light curve in blue and the centroid curve in orange. The black number is the total number of objects classified in this subset, while the red number shows recall to two significant figures, i.e. the proportion of each class that is estimated to be a member of this class (hence horizontal rows always sum to 100% within rounding errors). Objects on the diagonal are correctly classified and coloured green, while those outside are mis-classified and are coloured red. The strength of the background colours is proportional to the percentage (i.e. recall) in each box.

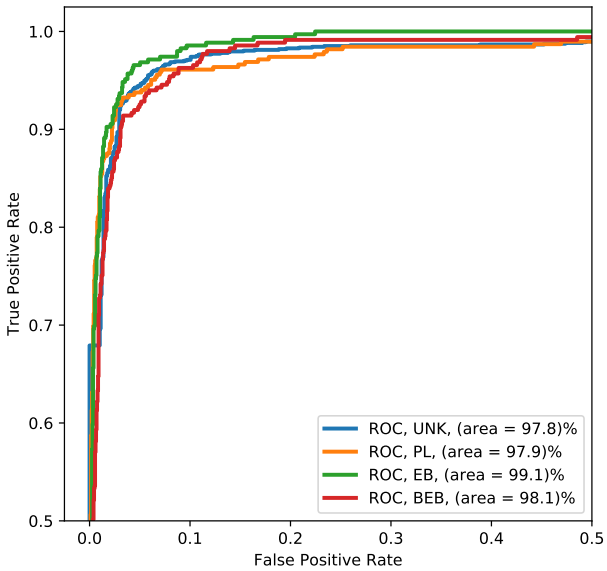

Fig. 6. Receiver operating characteristic curve for the four-class `exonet` model. The EB and BEB models perform poorly in comparison to Fig. 4, mostly due to confusion with each other (see Fig. 7).

compiled using candidates from both the SPOC pipeline (which come from the TCE tables used here) and the so-called quick-look pipeline (QLP). Planet candidates are then identified by

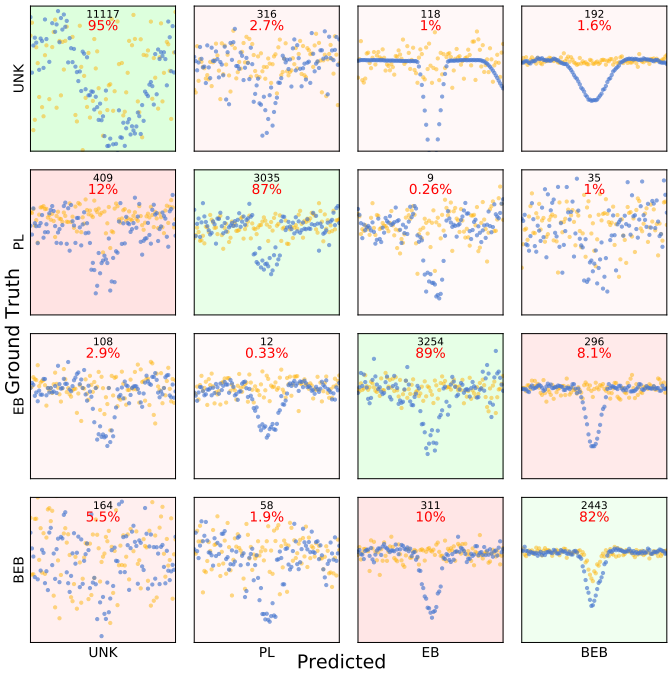

Fig. 7. Confusion matrix, as in Fig. 5 but for the four-class `exonet` model.

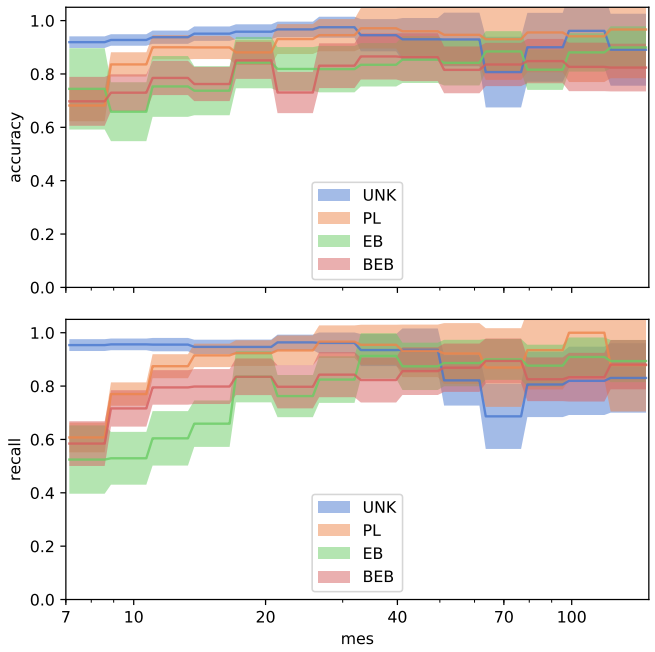

Fig. 8. Comparison of recall and accuracy as a function of the MES for each of the classes in our four-class model.

manual vetting as in Crossfield et al. (2018). As expected, our TCE list contains all but one of the 146 SPOC-derived TOIs, but only 46 of the 207 QLP-derived TOIs. Including duplications due to candidates identified in multiple sectors, we found 353 TCEs which corroborated with 201 TOIs (based on a combination of period, epoch, duration, and depth matches). Of those TCEs, 61% were classified as planets with a threshold of >50% in our average classifier. Ignoring duplications and taking the classification from the most sectors (or otherwise averaging the classifications) gives 112 out of 212 TOIs in agreement.

Table 3. TESS objects of interest with a high likelihood of being astrophysical false positives.

TID	TOI	UNK_3	PL_3	EB_3	EB_4	BEB_4
2760710	227.01	0.090	0.000	0.910	0.836	0.024
279740441	273.01	0.574	0.000	0.426	0.000	0.835
425934411	142.01	0.506	0.001	0.493	0.000	0.678
272086159	176.01	0.001	0.000	0.999	0.879	0.003
272086159	176.01	0.002	0.000	0.998	0.800	0.000
237924601	252.01	0.006	0.000	0.994	0.003	0.873
92226327	256.01	0.150	0.000	0.850	0.558	0.276
425934411	142.01	0.157	0.005	0.838	0.000	0.979
425934411	142.01	0.033	0.000	0.966	0.000	0.853
237924601	252.01	0.021	0.107	0.872	0.125	0.592
231702397	122.01	0.199	0.390	0.411	0.057	0.640
307210830	175.02	0.186	0.412	0.401	0.000	0.610
176778112	408.01	0.352	0.014	0.634	0.000	0.520
355703913	111.01 [(a)]	0.055	0.010	0.935	0.005	0.408
237924601	252.01	0.110	0.399	0.491	0.011	0.515

Notes. We only show all classes for the three-class model, and the split EB and BEB classes from the four-class model. [(a)]Marks HATS-34b.

Of the 43 planets known before launch (the majority being hot jupiters), 95.2% were classified as planets by the model, with the exception of WASP-18b (planetary probability estimate ($p_{\rm pl}$) = 49.9%) which has a detectable secondary eclipse (Shporer et al. 2019), and HATS-34b ($p_{\rm pl}$ = 39%) which has a V-shaped transit (b = 0.94, de Val-Borro et al. 2016).

Of the 14 planets already confirmed by TESS from the TOI list, we strongly identify TOIs 123b, 135b, 125b, 120b, 125c, and 174b, ($p_{\rm pl} > 0.9$), weakly recognise TOIs 216b, 197b, 144b and TOI-136b, ($0.3 < p_{\rm pl} < 0.9$), and misidentify TOI-125c, TOI-216c, and TOI-256b & c.

A further 15 are ranked as EBs or BEBs (with $p > 0.5$ in either model), which we list in Table 3. However, a quick manual vetting of these signals does not come to the same conclusion, with TIDs (TESS ID, as defined in the TIC, Stassun et al. 2018) 2760710, 92226327, 231702397, and 307210830 still possible planet candidates. A total of 95 and 82 objects are classed as "unknown" (e.g. from a non-astrophysical source) with the three- and four-class models, respectively.

Interestingly, a further 200 TCEs not classified as TOIs are estimated to be members of the planet class (see Table A.2). These are spread across 144 unique TESS objects, with 57 of those having a class probability greater than 90%. After viewing these 200 TCEs, we plotted a handful of the most promising planetary candidates in Fig. 10.

6. Discussion

6.1. Comparison with Ansdell et al. (2018)

In Ansdell et al. (2018), we achieved an average precision of 98%, with an accuracy on planets of 97.5%. In this study, we are unable to achieve a similarly high average precision or accuracy, with 97.3% average precision for the binary model and 92% accuracy on planets in the three-class model. A number of limitations could explain this discrepancy, which we cover in turn here.

The most obvious is in the presence of more false positives in the TESS input data, whereas some non-astrophysical false

positives (objects labelled as unknown by Batalha et al. 2013; Burke et al. 2014; Mullally et al. 2015; Rowe et al. 2015) were removed from the samples in both Shallue & Vanderburg (2018) and Ansdell et al. (2018). The abundance of non-astrophysical false positives in this TESS dataset may also be caused in part by the reduction in the minimum number of transits from three to two, allowing two non-periodic noise sources to combine to give a candidate signal (far more difficult in the $N_t \geq 3$ case). Another discrepancy is in the source of labels: for this study the ground truth was, for the most part, known absolutely thanks to simulations. In the *Kepler* dataset, only those signals identified as planet candidates by humans were positive classes, introducing a possible human bias. For example, planets that are more difficult to identify (e.g. those affected by systematic noise) may have been missed in human vetting, improving the overall quality of the planet class.

It may seem like another difference might be the proportion of low-S/N planet candidates in TESS compared to *Kepler*, which could be intrinsically higher due to the larger average flux uncertainties. However, this is not the case, and the distribution of injected TESS planets and *Kepler* planet candidates is similar in terms of S/N. Instead, low-S/N TESS candidates are found, on average, at a larger planetary radius. This itself may be problematic, as large planets are more easily confused with eclipsing binaries, although the difference is likely minimal.

Another significant difference between the TESS and *Kepler* datasets is in the centroids. The uncertainty in the centre of light (i.e. centroid) is determined by two things – the total number of photons and the number of pixels that light is spread over. TESS suffers in both of these cases when compared to *Kepler*, with fewer photons (a direct correlation with the higher average noise in TESS), and larger pixels compared to the point spread function (PSF). This means centroids are noisier, and TESS may not see a centroid shift on an object for which *Kepler* was able to. This discrepancy may also explain why we initially found that adding centroids caused problems with model training. Another reason is the increased presence of NaNs in the input data arrays, due partly to the shorter baseline and decrease in the minimum number of transits to two compared to three in *Kepler*.

Another source of the discrepancy is in noise in the labels. Although it may appear to simple to identify true signals from no signal at all in simulated data, this is not necessarily true. For example, single transits and eclipses were frequently detected by TESS, with an incorrect period and/or with a second transit detected corresponding to some systematic noise or gap. Our correlation metric would in these cases discard this as a "near miss". However, the neural network is indeed seeing the signal of an astrophysical source.

A manual inspection of those candidates estimated to be planets by the three-class model reveals that 69% of those 284 objects with "unknown" ground truth were in fact co-incident with planetary injections. Of those, 44% came from monotransits, 25% came from period confusion in multi-planet systems (e.g. a period near resonant with two or more planets, producing a planet-like phase-folded transit in combination), and the other 31% had other origins such as single or half-transits left by the first iteration of transit detection which were then identified in the incorrect period in the second search; and transits close to, but not at, the correct period which became "smeared" in phase-space. Immediately including these in the correctly identified boxes improves planet accuracy in the cross-validation results for the three-class model from 90.3 to 95.7% and the average precision across all classes to 95.0%, nearly matching that of Ansdell et al. (2018). A similar increase would be expected in the ensemble test data. However, this poses a question as to exactly what constitutes a bona fide planet detection, and whether planets on missed periods constitute a true detection or not. One improvement might be to apply a continuous label from the degree of correlation between injection and recovered signal rather than a pure binary label. This however is beyond the scope of this work.

6.2. Comments on multiclass and binary models

We attempted to train both binary and multi-class models partly because we assumed that the simplicity of a binary model may improve performance. We thought that a multi-class model with specific knowledge of the source of the possible false positive contaminant may aid planet follow-up. For example, class confusion between a planet and a background eclipsing binary may lead to the need for high-resolution imagery, whereas confusion between a planet and a non-astrophysical signal may lead to follow-up photometric observations of a future transit.

However, our results suggest that there is minimal difference between binary and multi-class models, as Fig. 3 suggests. In fact, the highest average precision on planets in the binary and three-class models were equal at AP = 95.6%; this is likely to be even higher if monotransits and other near-miss planetary signals are included as true positives.

Figure 9 also shows how all three models perform worse when classifying planets at lower S/N, with only between 60 and 70% of planets with $7 < S/N < 8.5$ detected. This decrease is expected, as the threshold for detections which become TCEs (7.2σ) is set such that the fraction of signals at this threshold which are from real astrophysical sources (both planets, EBs and BEBs) is 50%. Far fewer than 50% of TCEs with MES $\sim 7.1\sigma$ are therefore likely to be planets, and our value of up to ~70% shows a marked improvement.

Where multiclass models did have a noticeable negative effect was on the accuracy of identifying EB and BEB objects in the four-class model. When normalised by depth and duration, blended binaries have identical shapes to eclipsing binaries, although often at lower S/N. These two classes were therefore frequently confused, as Fig. 7 shows, leading to lower average precision for the model as a whole.

Another noticeable result from the multiclass models is that our model performs best at identifying non-astrophysical false positives. This is especially true at low S/N (see Fig. 8) where the recall and accuracy on these unknown signals actually increase. This may suggest however that the model is in some ways "playing the system" and applying the class of "unknown" for all transit events with high noise, where such signals dominate. However, the accuracy of the classification also suggests that the model is learning the systematic noise inherent to the data, and is therefore able to separate these signals from the astrophysical classes. This in itself is extremely important as often these systematic noise sources are varied and are unable to be modelled, and are therefore difficult to distinguish using classical techniques.

6.3. Real TESS data

Directly applying a model trained on simulated data to real data is a risky strategy. Although the planet and EB signals are likely to appear physically similar, the characteristics of the systematic noise are likely extremely different. However, a recall of 61% on the TOI list is a relatively good sign that the model is transferable. Especially given the different techniques and even pipeline

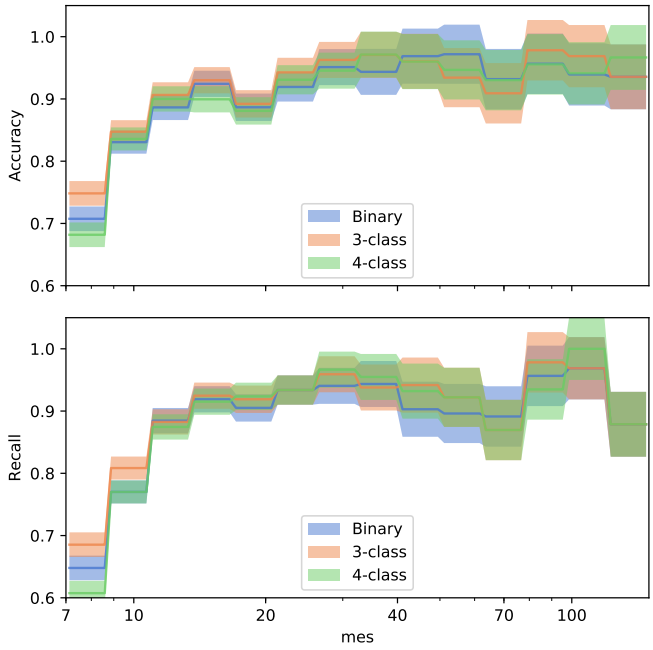

Fig. 9. Comparison of recall and accuracy for planets as a function of the MES across all three models. We note the y-scale has been re-scaled between 60 and 100%.

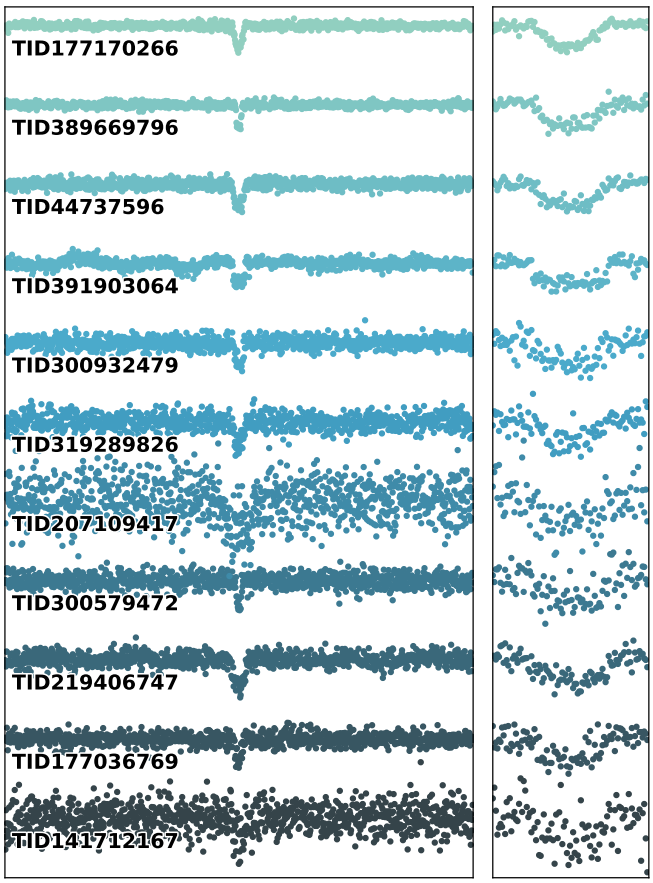

Fig. 10. Small selection of candidates (TCEs) that have not been classified as TOIs but nonetheless appear to be good planet candidates. These come from a manual search for planet-like signals amongst those 200 TCEs with high $p_{\rm pl}$ estimations from our CNN models. *Left panel*: "global" view for the whole phase; *right panel*: "local" view. These are sorted by transit S/N. See Table A.2 for full information on all estimated planets.

used to create the TOI list, and given the as-yet unknown ground truth of those targets in the TOI list. Indeed, our models suggest a handful of TOIs are indeed most likely to be astrophysical false positives.

Our model also suggested nearly 100 further TCEs have high ($p_{\rm pl} > 0.95$) planetary class probabilities. These include a 760 ppm signal from HD 55820 (TIC391903064), a 4 ppt transit on HD 270135 (TIC389669796), and a 500 ppm signal from TIC207109417. Such targets are ripe for follow-up, and we hope future vetting by improved models will confirm these signals and identify even more. Unfortunately, a quick look at those predicted planets also reveals many clear binaries, although the majority have transit-like eclipse shapes due to a large radius ratio, such as binary companions of giant stars. This may be due to our input training set lacking objects of this nature.

Clearly our recall and accuracy on planet candidates in the simulated data is not matched when applying that to real data. However, without full knowledge of the ground truth in the real TESS data, assessing model performance will be intrinsically more challenging. In order to best represent the realistic noise sources, one could inject and recover realistic transit signals in real TESS data. However, this work was started before real data was available, and performing injections and recovery in real TESS light curves is beyond the scope of this paper. We intend to perform such a task in a future publication.

7. Conclusion

The classification of candidates in exoplanet transit surveys can be a long and labour-intensive process when manual techniques are used. Neural network-based classifiers `Astronet` (Shallue & Vanderburg 2018) and `exonet` (Ansdell et al. 2018) have proven themselves to be extremely accurate (98% average precision in Ansdell et al. 2018) and once trained can classify potential planets extremely rapidly. We set out to apply such models to TESS-like data.

To do this, we followed the `Astronet` technique of using local and global "views" of the phase-folded, binned photometric data for each candidate, as well as the improvements of `exonet` – namely including the centroids and stellar parameters. In order to improve results, we also added data augmentation by mimicking additional noise sources, and use balanced batch sampling to normalise the unequal number of samples of each class in training data.

Using four sectors of pixel-level simulations with injected planets and false-positive populations (known as TSOP-301), we trained three models with varying numbers of source classes using cross validation. We achieve average precision as high as 97.3% with accuracy on planet populations as high as 91.8%. This is despite limitations when compared to those results using *Kepler* data, such as lower-significance centroids, a large population of non-astrophysical false positives (which were partly removed in the *Kepler* ML dataset), and a higher degree of confusion between real planet signals and noise due to the lowered threshold on the number of transits from three to two. Indeed, when positives from confusion between planet injections or monotransits identified at the wrong period are included in the "planet" class, accuracy rises to as high as 95.7%.

We also show that our models perform well even at low-S/N with accuracy on planets as high as 75% for signals with $S/N < 8.5$. Our use of multi-class models may also aid targeted follow-up observations by providing class probabilities for different false-positives which may impact follow-up strategy. The

high accuracy in non-astrophysical false positives also suggest that our neural network is able to learn patterns in sources with low-significance systematic noise. This could therefore push planet detection closer to the theoretical S/N limit than is possible with classical vetting techniques.

Once these models were trained on simulated data, we applied them to real TESS candidates from sectors 1 to 4. Although no ground truth exists to test the performance of this model, we recover more than 60% of the currently identified TOI list, including more than 95% of all planets identified before the mission. We also identify 14 TOIs as likely false positives. However, the use of confirmed TESS planets as a training set, plus injections of simulated transits into real flight data, would improve our confidence in such classifications. This will form the next step in this ongoing project.

Acknowledgements. Software used includes: Astropy (Astropy Collaboration 2013, 2018), PyTorch (Paszke et al. 2017), Astronet (Shallue & Vanderburg 2018), Jupyter (Kluyver et al. 2016), SciPy (Jones et al. 2001), Scikit Learn (Pedregosa et al. 2011), Matplotlib (Hunter 2007). This work, along with Ansdell et al. (2018), are the result of the 2018 NASA Frontier Development Lab[7] (FDL) exoplanet challenge, and we are extremely grateful to the organisers, mentors, and sponsors for providing this opportunity. This material is based on work supported by Google Cloud. M.A. also gratefully acknowledges the NVIDIA Corporation for donating the Quadro P6000 GPU used for this research. We thank Adam Lesnikowski, Noa Kel, Hamed Valizadegan and Yarin Gal for their helpful discussions. This paper includes data from the TESS mission; funding for the TESS mission is provided by the NASA Explorer Program. The data used in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26 555. M.A. also acknowledges support from NSF grant AST-1 518 332 and NASA grants NNX15AC89G and NNX15AD95G/NEXSS. We acknowledge the use of public TESS Alert data from pipelines at the TESS Science Office and at the TESS Science Processing Operations Center. H.P.O. acknowledges support from Centre National d'Etudes Spatiales (CNES) grant 131 425-PLATO.

References

Ansdell, M., Ioannou, Y., Osborn, H. P., et al. 2018, ApJ, 869, L7
Armstrong, D. J., Günther, M. N., McCormac, J., et al. 2018, MNRAS
Astropy Collaboration (Robitaille, T. P., et al.) 2013, A&A, 558, A33
Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, AJ, 156, 123
Barclay, T., Pepper, J., & Quintana, E. V. 2018, ApJS, 239, 2
Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, ApJS, 204, 24
Burke, C. J., Bryson, S. T., Mullally, F., et al. 2014, ApJS, 210, 19
Chawla, N. V., Japkowicz, N., & Kotcz, A. 2004, ACM Sigkdd Explorations Newsletter, 6, 1
Crossfield, I. J., Guerrero, N., David, T., et al. 2018, ApJS, 39, 5
de Val-Borro, M., Bakos, G., Brahm, R., et al. 2016, AJ, 152, 161
Dietterich, T. G. 2000, in International Workshop on Multiple Classifier Systems (Berlin: Springer), 1
Fausnaugh, M., Huang, X., Glidden, A., Guerrero, N., & TESS Science Office. 2018, AAS Meeting Abstracts, Vol. 231, 439.09
Fischer, D. A., Schwamb, M. E., Schawinski, K., et al. 2012, MNRAS, 419, 2900
He, H., & Garcia, E. A. 2008, IEEE Trans. Knowl Data Engineering, 21, 1263
Huang, C. X., Shporer, A., Dragomir, D., et al. 2018a, AJ, submitted [arXiv:1807.11129]
Huang, C. X., Burt, J., Vanderburg, A., et al. 2018b, ApJ, 868, L39
Hunter, J. D. 2007, Comput. Sci. & Eng., 9, 90
Jenkins, J. M., Caldwell, D. A., & Borucki, W. J. 2002, ApJ, 564, 495
Jenkins, J. M., Chandrasekaran, H., McCauliff, S. D., et al. 2010, SPIE Conf. Ser., 7740, 77400D
Jenkins, J. M., Twicken, J. D., McCauliff, S., et al. 2016, Proc. SPIE, 9913, 99133E
Jenkins, J. M., Tenenbaum, P., Caldwell, D. A., et al. 2018, Res. Notes AAS, 2, 47
Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python
Kingma, D. P., & Ba, J. 2014, ArXiv e-prints [arXiv:1412.6980]
Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in Positioning and Power in Academic Publishing: Players, Agents and Agendas, eds. F. Loizides & B. Schmidt (Amsterdam: IOS Press), 87
Li, J., Tenenbaum, P., Twicken, J. D., et al. 2019, PASP, 131, 024506
McCauliff, S. D., Jenkins, J. M., Catanzarite, J., et al. 2015, ApJ, 806, 6
Mullally, F., Coughlin, J. L., Thompson, S. E., et al. 2015, ApJS, 217, 31
Mullally, F., Coughlin, J. L., Thompson, S. E., et al. 2016, PASP, 128, 074502
Paszke, A., Gross, S., Chintala, S., et al. 2017, NIPS 2017 Workshop Autodiff Submission
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, Proc. SPIE, 9143, 914320
Rowe, J. F., Coughlin, J. L., Antoci, V., et al. 2015, ApJS, 217, 16
Schanche, N., Cameron, A. C., Hébrard, G., et al. 2018, MNRAS, 483, 5534
Seader, S., Tenenbaum, P., Jenkins, J. M., & Burke, C. J. 2013, ApJS, 206, 25
Shallue, C. J., & Vanderburg, A. 2018, AJ, 155, 94
Shporer, A., Wong, I., Huang, C. X., et al. 2019, AJ, 157, 178
Stassun, K. G., Oelkers, R. J., Pepper, J., et al. 2018, AJ, 156, 102
Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., et al. 2015, ApJ, 809, 77
Twicken, J. D., Catanzarite, J. H., Clarke, B. D., et al. 2018, PASP, 130, 064502
Vanderspek, R., Huang, C. X., Vanderburg, A., et al. 2019, ApJ, 871, L24
Wang, S., Jones, M., Shporer, A., et al. 2019, AJ, 157, 51
Zucker, S., & Giryes, R. 2018, AJ, 155, 147

[7] https://frontierdevelopmentlab.org/

Appendix A: Additional tables

Table A.1. TCEs that correspond to detected TOIs, with the estimated class probability as given by our CNN models.

TCE info							Binary	3-class estimates			4-class estimates				All
TESS ID	TOI	Sect	Period (d)	Epoch	t_D (hr)	δ (ppm)	PL	UNK	PL	EB	UNK	PL	EB	BEB	PL
77031414	241.01	2	1.387	1355.196	2.290	14 260	0.998	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.999
122612091	264.01	3 4	2.217	1387.831	3.846	4180	0.997	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.999
1129033	398.01	4	1.360	1410.985	2.156	16 380	0.997	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.999
290131778	123.01	1	3.309	1325.375	5.672	3230	0.997	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.999
144065872	105.01	1	2.185	1326.506	2.865	11 840	0.997	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.999
230982885	195.01	2	2.073	1355.490	2.588	13 510	0.997	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.999
184240683	250.01	2	1.628	1355.508	2.376	13 800	0.997	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.999
267263253	135.01	1	4.127	1325.784	4.495	10 070	0.997	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.999
25375553	143.01	1	2.311	1325.582	3.416	6950	0.996	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.999
388104525	112.01	1 2 3	2.500	1327.410	2.877	14 940	0.996	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.999
25155310	114.01	1 2 3	3.289	1327.521	3.416	7180	0.996	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.999
38846515	106.01	1 2 3	2.849	1326.745	3.785	7500	0.996	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.999
422655579	388.01	4	2.903	1413.143	5.048	4660	0.995	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.998
402026209	232.01	2	1.338	1355.185	2.157	27 440	0.995	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.998
231670397	104.01	1	4.087	1327.673	5.597	3610	0.993	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.998
149603524	102.01	1 2 3	4.412	1326.079	3.779	14 030	0.992	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.997
92352620	107.01	1	3.950	1328.299	4.580	12 930	0.990	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.997
336732616	103.01	1	3.548	1327.253	3.488	10 400	0.989	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.996
166836920	267.01	3	5.752	1387.960	5.386	5390	0.986	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.995
281459670	110.01	1 2	3.174	1328.040	2.723	15 600	0.984	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.995
170634116	413.01	4	3.662	1412.892	3.784	12 460	0.980	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.993
257567854	403.01	4	3.533	1411.902	3.443	11 280	0.980	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.993
238176110	116.01	1	2.799	1326.689	2.365	16 850	0.978	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.993
260609205	219.01	1 2 3	4.462	1328.755	5.410	21 140	0.976	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.992
183537452	192.01	2	3.923	1356.415	2.616	11 500	0.968	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.989
403224672	141.01	1	1.008	1325.539	1.499	220	0.967	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.989
219253008	268.01	4 3	5.066	1415.524	6.158	1880	0.964	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.988
97409519	113.01	1	3.373	1327.053	2.640	17 150	0.964	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.988
204376737	231.01	2	3.361	1357.395	2.579	24 220	0.963	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.988
183979262	183.01	1 2 1 2	3.431	1326.104	3.130	8980	0.962	0.001	0.999	0.000	0.003	0.996	0.001	0.0	0.986
35857242	400.01	4	3.635	1413.315	4.005	8890	0.957	0.000	1.000	0.000	0.000	1.000	0.000	0.0	0.986
234994474	134.01	1	1.401	1326.033	1.261	560	0.956	0.000	0.999	0.001	0.000	1.000	0.000	0.0	0.985

Notes. These are ranked by the average planet score. 201 rows, see associated CDS table of all TCEs.

Table A.2. TCEs which do not correspond to TOIs, but to which our model gives a high average score for the planet class.

TCE info						Binary	3-class estimates			4-class estimates				All
TESS ID	Sect	Period (d)	Epoch	t_D (hr)	δ (ppm)	PL	UNK	PL	EB	UNK	PL	EB	BEB	PL
92352621	1	3.950	1328.299	4.560	14 230	0.997	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.999
139256217	1	2.199	1325.350	3.821	18 410	0.997	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.999
259470701	4	2.543	1412.514	3.845	17 670	0.995	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.998
231275247	2	4.008	1354.114	4.995	21 100	0.995	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.998
272357134	1	4.197	1328.549	5.810	17 080	0.994	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.998
272357134	1 2 3	4.197	1328.549	5.810	17 030	0.992	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.997
272357134	1 2	4.197	1328.549	5.811	16 990	0.992	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.997
237342298	3	2.692	1386.312	4.341	31 710	0.991	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.997
231275247	3	4.008	1386.177	5.031	21 460	0.991	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.997
272357134	4	4.197	1412.487	5.797	17 060	0.991	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.997
272357134	3	4.197	1387.306	5.813	17 090	0.990	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.997
272357134	2	4.197	1357.927	5.812	16 860	0.989	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.996
231275247	1 2 3	4.008	1354.114	5.009	21 170	0.988	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.996
349308842	4	3.582	1410.985	4.156	32 620	0.985	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.995
170749770	4	5.306	1412.463	6.497	16 800	0.985	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.995
159984211	3	3.660	1386.931	6.411	10 150	0.982	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.994
237342298	4	2.692	1413.233	4.334	30 650	0.982	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.994
159984211	4	3.660	1412.553	6.462	10 990	0.980	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.993
350480660	4	4.464	1413.136	3.759	27 390	0.980	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.993
350480660	3	4.464	1386.354	3.739	27 110	0.976	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.992
64108432	4	2.779	1413.462	3.579	17 940	0.970	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.990
115115136	1	3.962	1327.888	4.558	32 680	0.968	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.989
268529943	3	4.301	1388.935	3.768	20 290	0.967	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.989
167554898	4	4.453	1413.030	3.887	6540	0.965	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.988
260756218	1	1.955	1325.391	2.300	26 760	0.941	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.980
422844353	4	6.635	1413.205	4.769	18 370	0.920	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.973
270380593	4	3.836	1414.136	2.390	5580	0.911	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.970
391903064	1 2	4.666	1326.494	3.005	760	0.901	0.000	1.000	0.000	0.011	0.989	0.000	0.000	0.964
178242590	4	1.302	1411.139	2.727	15 470	0.990	0.000	1.000	0.000	0.000	0.890	0.110	0.000	0.960
260304277	1 2	0.513	1325.773	1.100	180	0.863	0.000	1.000	0.000	0.000	0.999	0.000	0.001	0.954
234518605	1 2	5.679	1326.532	4.175	62 410	0.859	0.000	1.000	0.000	0.000	0.998	0.002	0.000	0.953
164752991	3	1.222	1386.808	1.225	2930	0.918	0.004	0.969	0.027	0.000	0.970	0.000	0.030	0.952

Notes. 210 rows, see associated CDS table of all TCEs.