IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11797

A Neural Network for Identifying Exoplanets

Varshini L[1], Uday A S[2], Merin Meleet[3]

Student, Information Science and Engineering, RV College of Engineering, Bangalore, India[1]

Student, Information Science and Engineering, RV College of Engineering, Bangalore, India[2]

Assistant Professor, Information Science and Engineering, RV College of Engineering, Bangalore, India[3]

Abstract: The Transiting Exoplanet Survey Satellite (TESS) has now been in operation for slightly more than two years, covering both the Northern and Southern hemispheres once. The TESS team uses the Science Processing Operations Center pipeline and the Quick Look pipeline to generate alerts for follow-up. Combined with other community efforts, over two thousand planet candidates have been discovered, with tens confirmed as planets. We present Udva, our pipeline that is complementary to these approaches. Udva employs a combination of transit detection, supervised machine learning, and detailed vetting to identify a few planet candidates that were missed by previous searches with high confidence. We find shallow transits with a high signal-to-noise ratio (SNR) that may represent more than one transit. Future work will include approaches to enhance stages that have been conservatively abandoned because they lacked one or two datums in order to boost the yield.

Keywords: planetary systems, planets and satellites: detection, techniques: photometric, methods: data analysis

I. INTRODUCTION

The TESS (Transiting Exoplanet Survey Satellite) project, which was put into orbit in 2018, is an all-sky digital survey designed to watch planetary transits of their host stars (Ricker et al. 2014). The TESS spacecraft continuously scans 24 x 96 sectors of the sky for consecutive intervals of 27 days, downloading data once every 13.5 days for additional processing. For context, each sector's data consists of over 20,000 light curves. Less than 1% of them, on average, are likely candidates for further study as planets. In reality, the great majority of false positives (e.g., variable stars, eclipsing binaries, etc.) and/or the effects of equipment systematics outweigh the number of planet candidates.

For categorising light curves, traditional machine learning algorithms including decision trees, k-nearest neighbours, and random forests have been investigated. One method for identifying light curves has been investigated, and it uses convolutional neural networks (CNN), which are efficient at image recognition tasks. These models employ supervised learning with a CNN, with the neural network learning the properties from light curves. To quickly identify targets, these systems use the following procedure: (1) Pre-processing of light curves to search for transits (i.e., periodic dips in the flux) to extract transit parameters such as period and depth, (2) generate phase-folded light curves from the transit parameters, which are then used to train a CNN-based model using labelled data, and (3) make predictions on freshly observed data. This approach has been utilised effectively in practise with Kepler and K2: (1) Shallue and Vanderburg (2018) apply their Astronet deep learning model on Kepler data to select numerous candidates from which they verified two new exoplanets. Ansdell et al. (2018) extend the basic Astronet model to add stellar parameters and show, using Kepler data, that their Exonet model increases model accuracy by including domain information.

In this study, we look at how CNNs may be used to categorise TESS light curves in order to find planet candidates. These current models based on Kepler data provide a suitable starting point. However, the systematics with specific TESS sectors differ greatly from Kepler/K2 (27 days of data against 4 years/81 days of data). As a result, previous deep-learning systems for categorising TESS data, such as Yu et al. (2019) and Osborn et al. (2020), adapt and modify Astronet and Exonet, respectively, to find new possibilities.The main contribution of this study is the construction and assessment of a pipeline based on deep learning for finding potential planets from TESS data. Two components of the construction of our pipeline, Udva, are novel: the input representation of the light curve to the CNN model and the architecture of the model. Utilizing labelled data from real TESS sectors, we train and assess the model. Our model has an AUC score of 0.908. By adding genuine planets to our test collection and evaluating our model, we discover that it recovers 91.9% of the data. We use the data from TESS Sector 6 and Sectors 21–26 and our trained model to generate predictions and identify new candidates.

DOI: 10.17148/IJARCCE.2022.11797

II. LITERATURE SURVEY

A vast survey was conducted over the internet in relevant domains of this project. The domains for this survey included Kepler/K2, TESS mission and several methods for identifying exoplanets.

[1] Sara Cuéllar, Paulo Granados, et al. worked on the article proposing the development of a deep learning system for detecting planetary transits in Kepler Telescope light curves which is based on related work from the literature and enhanced to validation with real light curves. The results demonstrated that the use of synthetic data on the training stage can improve the transit detection performance on real light curves.

[2] Pattana Chintarungruangchai and Ing-Guey Jiang used machine learning techniques with two-dimension convolutional neural networks for detecting exoplanet transits. It was shown that the 2D-CNN-folding-2 model can have rather good accuracy even when the folding period is different from the transit period by 20%.

[3] Abhishek Malik, Ben Moster and Christian Obermeie introduced a new machine learning based technique to detect exoplanets using the transit method. They aimed at exploitation of some methods to improve the conventional algorithm based approach used in astrophysics today to detect exoplanets. On Kepler data, the method was able to predict a planet with an AUC of 0.948.

[4] Yucheng Jin, Lanyi Yang and Chia-En Chiang used the Kepler dataset collected by NASA from the Kepler Space Observatory to conduct supervised learning, which predicts the existence of exoplanet candidates as a three-categorical classification task, using decision tree,random forest, naïve Bayes, and neural network.

[5] Greg Olmschenk, Stela Ishitani Silva, Gioia Rau, et. al built a convolutional neural network, which was trained to identify planetary transit signals and dismiss false positives. To make a prediction for a given light curve, the network requires no prior transit parameters identified using other methods. The network performs inference on a TESS 30-minute cadence light curve in ~5ms on a single GPU, enabling large scale archival searches.

[6] Li-Chin Yeh and Ing-Guey Jiang worked on the photometric light curves of BRITE satellites which were examined through a machine learning technique to investigate whether there are possible exoplan-ets moving around nearby bright stars. The study demonstrates that the CNN models could successfully identify possible transit candidates.

[7] Pawel Pratyush1, Akshata Gangrade worked on detection of transiting exoplanets and habitability determination using machine learning. A new metric Adequate Thermal Adequacy (ATA) score was introduced to classify habitable and non-habitable instances with more robustness. Regression curve was introduced to show Earth Similarity Index (ESI) which was calculated by considering some common attributes.

[8] L. Ofmam, Amir Averbuchch, et.al worked on a novel artificial intelligence technique that uses machine learning methodologies that combines several algorithms and is applied to NASA's TESS dataset to identify exoplanetary candidates. This study demonstrates for the first time the successful application of the particular combined multiple AI/ML-based methodologies to a large astrophysical dataset for rapid automated classification of TCEs.

III. DATA PREPARATION

We make use of TESS light curves that have been analysed and obtained from the Mikulski Archive for Space Telescopes (MAST). These are devoid of instrumental systematics and feature a two-minute cadence throughout a 27-day span. To look for Threshold Crossing Events (TCEs), we use the Pre-Search Data Conditioning Simple Photometry flux (PDCSAP FLUX) time series from each light curve. These are regular flux changes that might be brought on by exoplanets transiting their host star. We employ TCEs from various TESS sectors to train our models, which are subsequently used to find new candidates.

For training labels, we employ two sets of labelled data. Our first labelled data set is created using the TESS TOI catalogue. The disposition for that target as well as supplementary metadata are specified in this catalogue along with the TESS identification for which an alert was created. Some of the TESS labels are, KP stands for planets known before launch, CP for confirmed planets discovered using TESS data, and PC for targets identified as planet candidates.

Using the publicly accessible open source version of TLS, we calculate transit parameters for each dataset utilised in this study. We detrend the PDCSAP_FLUX time series from each TESS light curve file as a preprocessing step. We

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11797

next compute the power spectrum on the detrended data using TLS APIs. In this stage, we enter the stellar limb darkening parameters estimated by Claret (2018) for TESS targets and use the default TLS API values for the remaining parameters. TLS computes the transit parameters corresponding to the best fit for the data based on these inputs. TLS also provides the number of transits, signal-to-noise ratio (SNR), and signal detection efficiency (SDE), which correspond to the peak in the TLS power spectrum. For a light curve to be flagged as a TCE we need (a) at least two transits, (b) an SNR > 7.1, and (c) an SDE > 9.

According to Ansdell et al. (2018), adding more domain features can enhance the performance of deep learning-based classifiers. We therefore add the following extra inputs to our model: (1) stellar parameters, including stellar effective temperature (Teff), stellar log g (log g), stellar radius (R), star mass, brightness, and stellar density; and (2) transit parameters, including transit depth, transit length, Rp/Rs, and mean odd/even transit depth, derived using TLS.

IV. UDVA PIPELINE

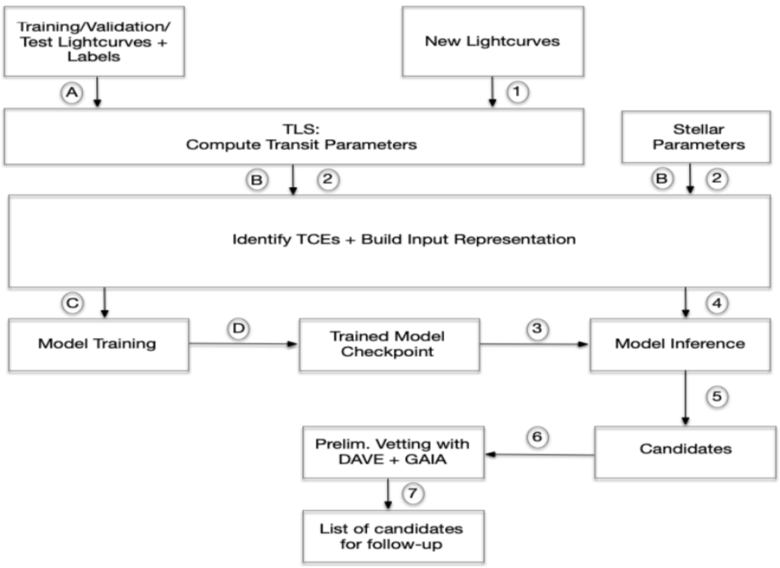
We outline our machine learning model's neural network architecture. Next, we go over our model training and prediction process. After that, we assess model performance.

A. Udva pipeline overview

Figure 1 depicts the training and prediction flows in our pipeline. Steps (A)–(D) comprise the model training path.

A. TLS is used to analyse the incoming light curves and compute transit parameters.
B. The Global/Local/Half-Phase views are constructed by pre-processing the TCEs with the transit parameters. Additionally, the stellar parameters are integrated to create the TCEs' input representation.
C. The input representations are then used for model training.
D. A checkpoint is created at the completion of model training and specifies the model weights.

The prediction path is made up of steps (1) through (4). Steps (1) and (2) provide the input representation for the TCEs in a manner similar to the training path. (Step 3) involves loading the model checkpoint and producing predictions for each input (Step 4). For each TCE, the inference phase gives a prediction, and the PCscore is used to rank the candidates. We use a cutoff on PCscore to choose candidates (Step 5). We perform initial vetting of the candidates (Step 6). The result of the last step is a list of candidates for follow-up (Step 7).



B. Model Architecture

The Udva pipeline is built around a convolutional neural network based deep learning model. Figure 2 depicts the model's architecture.

We base our model on the Astronet architecture (Shallue & Vanderburg 2018), and the convolutional layers for global/local views are analogous. The half-phase view, which includes layers identical to those in the local view, is then added to the model as an extension. Three distinct convolutional columns are input for the Global, Local, and Half-Phase views. The outputs of these convolutional layers are combined with the stellar and transit parameters and then

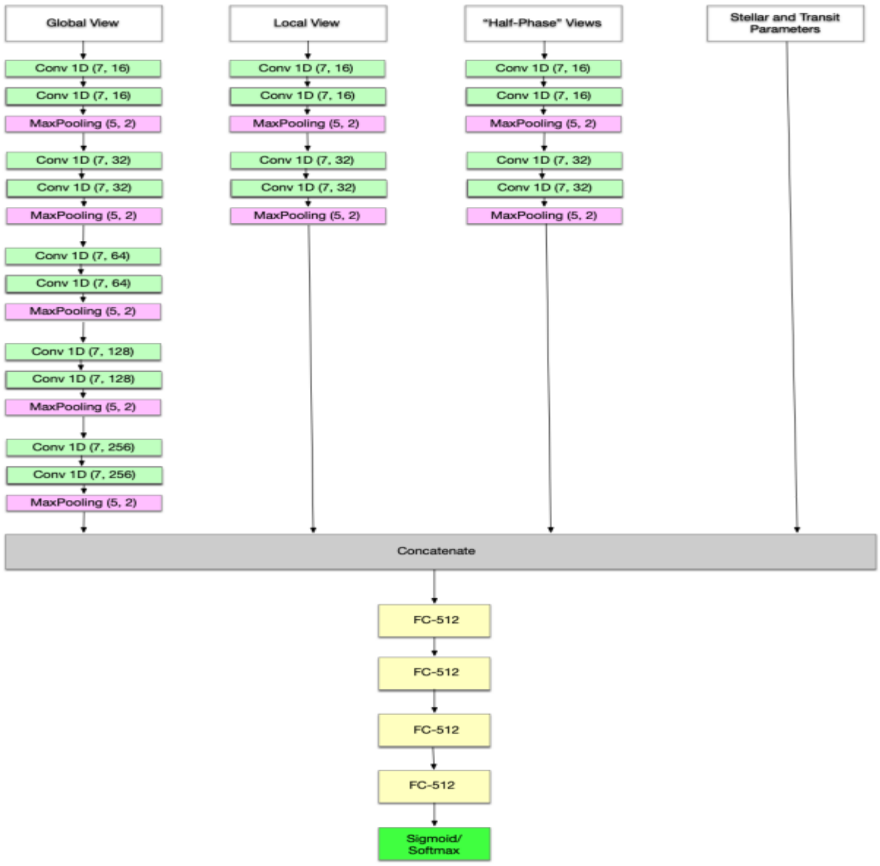
ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⋇ Impact Factor 7.39 ⋇ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11797

transmitted to the fully connected layers, much like the Exonet model (Ansdell et al. 2018). We use the relu activation function.



C. Model Training

We implement our deep learning model using Keras on Tensorflow. We use the labelled datasets to train our model. In Table 1, the labels for the good and bad examples used in the training are defined.

Type	Count		Label	
	DS1	DS2	P	N
Known Planet (KP)	166		✓	
Confirmed Planet (CP)	38		✓	
Planet Candidate (PC)	274	165	✓	
Eclipsing Binary (EB)		711		✓
False Positive (FP)	139			✓
Instrumental Noise (IS)		520		✓
Variable Star (V)		656		✓
Other (O)	1	1		✓
Junk (J)		3917		✓
Total	618	5970	643	5945
Grand Total	6588			

As is customary when training machine learning models, we divided the labelled data randomly into three sets for model training and evaluation: 80% is the training set used for model training, 10% is the validation set used for model validation, and 10% is set aside as the test set used only for model evaluation.

V. CONCLUSION

Udva, a pipeline built entirely from scratch utilising innovative machine learning methods. To start, the technique used to look for transit signals in light curves TLS is different from earlier methods that employed BLS and its variations. TLS is more sensitive to shallow transits brought on by planets that are close to the size of Earth than BLS is. This broadens the pool of potential individuals that may be found through search. The light curve is represented using a new input, which helps the algorithm better discriminate between planetary candidates and false positives brought on by eclipsing binaries. Utilizing information from TESS sectors, the model was trained and assessed. The model has a 0.908 AUC rating. In a test including light curves from known and TESS verified planets, the model successfully locates 91 of the 99 planets.

VI. RESULTS

Our test set consists of 632 TCEs of which 56 TCEs have one of KP, PC, CP labels corresponding to the positive instances. We create a confusion matrix for the test set (see Figure 3) for a threshold of PCscore > 0.7. In the diagram, PC stands for the TCEs corresponding to the positive cases, and FP for the instances that are negative. The model's precision is 88.8 %, recall is 74.3 %, and accuracy is 87.2 %. The PCscore range that our model allocates to TCEs corresponding to distinct classes is an intriguing issue. We also incorporate previously classified PCs from the TOI database in this analysis. Our model successfully recovers 91 of 99 KPs and CPs with PCscore > 0.7. It is desirable to have such a strong recall at a high threshold.

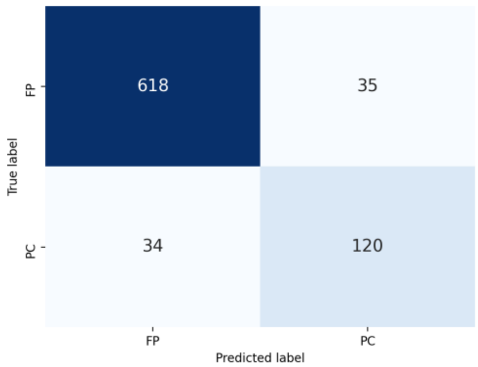


Figure 3: Confusion matrix when PCscore > 0.7 is used to identify planet candidates.

ACKNOWLEDGMENT

We are indebted to, **Prof. Merin Meleet,** Assistant Professor, Dept. of Information Science and Engineering, for her wholehearted support, suggestions and invaluable advice throughout our project work and also helped in the preparation of this thesis. Our sincere thanks to **Dr. Sagar B M,** Professor and Head, Department of Information Science and Engineering, RVCE for his support and encouragement.We express sincere gratitude to our beloved Principal, **Dr. K. N. Subramanya** for his appreciation towards this project work.We thank all the teaching staff and technical staff of the Information Science and Engineering department, RVCE for their help. Lastly, we take this opportunity to thank our family members and friends who provided all the backup support throughout the project work.

REFERENCES

[1]. Shalluel, Christopher J. and Vanderburg Andrew (2019). "Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90".The Astronomical Journal. Volume 155, Number 2, 30 January 2018. DOI: 10.3847/1538-3881/aa9e09

[2]. Maldonado, Miguel J , 2020. "Transiting Exoplanet Discovery Using Machine Learning Techniques:A Survey". Earth Science Informatics. volume 13, pages 573–600, 5 June 2020. DOI:https://doi.org/10.1007/s12145-020-00464-7.

[3]. Batalha, Natalie M. (2018). "Exoplanet populations with NASA's Kepler Mission". Proceedings of the National Academy of Sciences. 111 (35) 12647-12654, September 2014. DOI: 10.1073/-pnas.1304196111.

[4]. Lissauer, Jack J., Dawson, R. I., and Tremaine, S. (2014). "Advances in exoplanet science from Kepler". Nature. 513, 336–344, 2014. DOI: 10.1038/nature13781.

[5]. Saha, Rohan (2019). "Comparing Classification Models on Kepler Data". arXiv:2101.01904v2. DOI:10.13140/RG.2.2.11232.43523.

[6]. Ricker G. R., et al.(2019). " Space Telescopes and Instrumentation 2020": Optical, Infrared, and Millimeter Wave. p. 914320 (arXiv:1406.0151), doi:10.1117/12.2063489

[7]. Christiansen, J. L. (2017). "Planet Detection Metrics": Pixel-Level Transit Injection Tests of Pipeline Detection Efficiency for Data Release 25 (KSCI-19110-001), Tech. rep.

[8]. Coughlin, J. L. (2019). "Planet Detection Metrics": Robovetter Completeness and Effectiveness for Data Release 25 (KSCI-19114-001), Tech. rep.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11797

[9]. Lightkurve Collaboration et al., (2019).“Lightcurve: Kepler and TESS time series analysis in Python”, Astrophysics Source Code Library (ascl:1812.013)

[10]. Vanderburg A., et al.,(2019). “The Astrophysical Journal Supplement Series”, 222, 14 Pipeline Detection Efficiency for Data Release 25 (KSCI-19110-001), Tech. rep

[11]. A. Claret, “A new method to compute limb-darkening coefficients for stellar atmosphere models with spherical symmetry: The space missions tess, Kepler, Corot, and most,” Astronomy & Astrophysics, vol. 618, 2018.

[12]. M. Ansdell, Y. Ioannou, H. P. Osborn, M. Sasdelli, J. C. Smith, D. Caldwell, J. M. Jenkins, C. Räissi, and D. Angerhausen, “Scientific Domain Knowledge Improves Exoplanet Transit classification with Deep Learning,” The Astrophysical Journal, vol. 869, no. 1, 2018.

[13]. L. Yu, A. Vanderburg, C. Huang, C. J. Shallue, I. J. Crossfield, B. S. Gaudi, T. Daylan, A. Dattilo, D. J. Armstrong, G. R. Ricker, R. K. Vanderspek, D. W. Latham, S. Seager, J. Dittmann, J. P. Doty, A. Glidden, and S. N. Quinn, “Identifying exoplanets with deep learning. III. Automated Triage and vetting of tess candidates,” The Astronomical Journal, vol. 158, no. 1, p. 25, 2019.

[14]. H. P. Osborn, M. Ansdell, Y. Ioannou, M. Sasdelli, D. Angerhausen, D. Caldwell, J. M. Jenkins, C. Räissi, and J. C. Smith, “Rapid classification of tess planet candidates with Convolutional Neural Networks,” Astronomy & Astrophysics, vol. 633, 2020.