Bachelor’s Programme in Science and Technology

State-of-the-Art Transiting Exoplanet Detection Algorithms

Ilia Zalesskii

Aalto University

BACHELOR’S THESIS

Aalto University
BACHELOR’S THESIS 2024

State-of-the-Art Transiting Exoplanet Detection Algorithms

Ilia Zalesskii

Thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Technology.
Otaniemi, 26 Apr 2024

Supervisor: professor Maarit Korpi-Lagg
Advisor: docent Ghassem Gozaliasl

Aalto University
School of Science
Bachelor’s Programme in Science and Technology

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Ilia Zalesskii

Title
State-of-the-Art Transiting Exoplanet Detection Algorithms

School School of Science

Degree programme Bachelor’s Programme in Science and Technology

Major Data Science **Code** SCI3095

Supervisor professor Maarit Korpi-Lagg

Advisor docent Ghassem Gozaliasl

Level Bachelor’s thesis **Date** 26 Apr 2024 **Pages** 20+8 **Language** English

Abstract

Exoplanets are planets outside solar system. Some of them may host life or be habitable, and each tells us something new about our own solar system. Most prominent astronomical method of exoplanet detection is currently transit method, which aims to detect regular planetary eclipses of stars (i.e., transits). Transiting exoplanet surveys such as Kepler and TESS observe thousands of stars over yearslong periods, thus producing a substantial amount of data in need of efficient and reliable automated analysis.

This thesis reviews modern algorithms that aim to distinguish true transits from unrelated events in the data. The algorithms are compared by their architecture, performance and applicable missions to determine the state-of-the-art and identify the best approaches. The results indicate that ExoMiner V1.2 (Valizadegan et al., 2023) and Astronet-Triage-v2 (Tey et al., 2023) outperform other models for Kepler and TESS data, accordingly. In general, deep learning methods, such as convolutional neural networks, are the best tools for the problem, with potential for even further improvement from self-attention-based transformer models. This study mainly used qualitative analysis, and further research could focus on quantitative comparison of performance on newest datasets.

Keywords exoplanet astronomy, exoplanets, exoplanet detection methods, convolutional neural networks, transformer models, planets and satellites, machine learning, deep learning

urn https://aaltodoc.aalto.fi

Contents

1. Introduction

Ever since the first planet was found orbiting a star like our Sun, astronomers around the world have been hunting for distant worlds that could host life like on Earth. By 2023, scientists had discovered over 5,000 confirmed exoplanets — i.e., planets outside our solar system (NASA, n.d.), such as LTT 1445Ac (Figure 1.1). Exoplanet research is important for multiple reasons, such as extending our knowledge about the formation of our planetary system, searching for habitable planets (i.e., those with Earth-like conditions and chemical compositions), and looking for life in the universe (e.g., via detecting certain biomarkers in the atmosphere and with information theory (Vannah et al., 2023)).

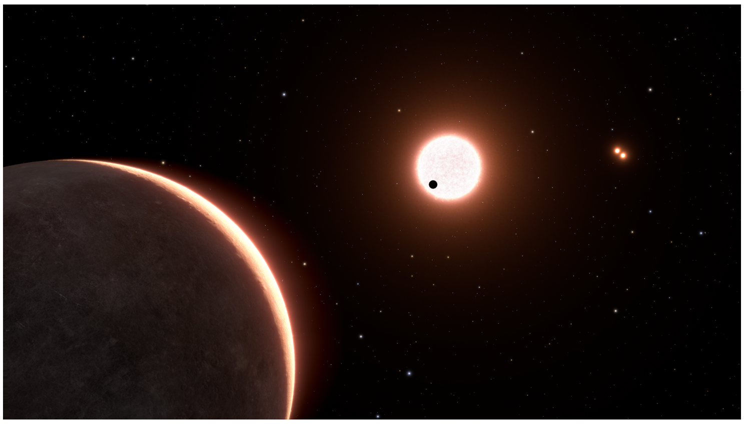


Figure 1.1. Artist's concept of one of the nearest detected Earth-size exoplanets, LTT 1445Ac. The planet can be seen as a black dot in front of the bright sphere. Another planet orbiting the same star, LTT 1445Ab, is in the lower left corner. The star forms a triple system with two red dwarfs that can be seen on the right. The system is located 22 light years from the Sun, which can be seen as a bright dot on the lower right. (NASA et al., 2023)

Purpose-built telescopes, such as the Kepler Space Telescope, have generated a significant amount of data, most of which is accessible through the Mikulski Archive for Space Telescopes (MAST)[1]. Ini-

[1] MAST

tially, this data had to be processed manually by researchers and volunteers with the goal of detecting potential exoplanets and then further analyzing whether the candidate was a planet. However, that proved to be a time-intensive task (e.g., NASA Exoplanet Archive contains observations from over 100 million stars (NASA, n.d.)). Moreover, manual examinations were found to be subject to human error (Pearson et al., 2017).

Several methods of exoplanet detection exist, the most prominent one to date being the transit method. It examines the brightness of a star as a function of time (i.e., its light curve) in attempt to detect the planets passing between some star and the observer. Due to the high amount of noise, weak transit signals, large amount of data, and other challenges (Jara-Maldonado et al., 2020), there is a need for efficient and accurate algorithms for transiting exoplanet detection.

With recent developments in fields of machine learning and artificial intelligence, new powerful methods of light curves analysis have emerged. These include promising results from convolutional neural network approaches (Chintarungruangchai and Jiang, 2019) and Transformer architectures (Salinas et al., 2023). Consequently, the aim of this thesis is to demonstrate the extent of current state of research in the field by conducting a state-of-the-art literature review. This work provides a modern guide to ML for transiting exoplanet research and outlines some algorithms that promise to be useful in further exploration of the sky.

The study is organized as follows. Section 2 outlines most prominent astrophysical methods of exoplanet detection, introduces relevant space missions and describes the light curve data that is used by the algorithms reviewed in this thesis. Section 3 elaborates on the methods used in selection of papers. Then, Section 4 introduces selected few algorithms in detail. Finally, Section 5 provides a comparative analysis of the algorithms covered and Section 6 draws conclusions.

2. Background

The following section provides the necessary astrophysical foundation, describes the relevant observatories and space missions, and explains some of the methods utilized in exoplanet research. Then, it familiarizes reader with the structure of data from transiting exoplanet surveys.

The first telescopes tasked with searching for planets outside our solar system were located in ground-based facilities, usually on high mountains, where light pollution is relatively low, such as the Wide Angle Search for Planets (WASP) survey located on the Canary Islands (Pollacco et al., 2006).

Space-based missions were then also created, because they were unaffected by day-night cycles and atmospheric events, among other reasons. For instance, NASA launched the Kepler Space Telescope in 2009; the primary goal of which was to hunt exoplanets in one section of the Milky Way galaxy. The Kepler mission showed that there are more planets than stars in the universe (NASA, 2019) and surveyed over 500 thousand stars over nine years, including the extended mission K2. The information collected by the telescope is still being used to detect more exoplanets (Jara-Maldonado et al., 2020). The Kepler mission provided over 9,000 Kepler Objects of Interest (KOIs) — stars that show periodic dimming which might be indicative of one or several planets orbiting them (NASA Exoplanet Archive, 2024).

More recently, the Transiting Exoplanet Survey Satellite (TESS) was launched, which is currently active in its extended phase. This telescope was designed to monitor only the closer stars (i.e, up to 300 light-years as opposed to 3,000 of Kepler). However, TESS looks at a much wider angle (over 75% of the sky), while Kepler was fixed on only a small segment during its primary mission (Koch et al., 2010; Ricker et al., 2015). The mission has already collected over 7,000 TESS Objects of Interest (TOIs) and is likely to collect more (NExScI, 2024).

Exoplanets in the scope of this thesis are defined as planets orbiting a star other than the Sun. In this regard, Borucki et al. (2011) separates exoplanets into following classes: Earth-size ($R_p < 1.25R_\oplus$), super-Earth-size ($1.25R_\oplus \leq R_p < 2R_\oplus$), Neptune-size ($2R_\oplus \leq R_p < 6R_\oplus$), and Jupiter-size ($6R_\oplus \leq R_p < 15R_\oplus$). Moreover, an important term for extraterrestrial life research is Habitable Zone (HZ) — i.e., the range of orbit radii where life could form. By some definitions, this range is 0.95 to 1.15 AU (Kasting et al., 1993). Undetected planets in this range could be predicted with the Titius-Bode law (Mousavi-Sadr et al., 2021).

This section further explains the basic physical methods of exoplanet detection, such as the radial velocity method (Section 2.1.2) and transit method (Section 2.1.1). As illustrated by Figure 2.1, the transit method accounts for almost all detections. However, other methods are important for alternative source confirmation and for obtaining further information about the planets. The nature and properties of the transit data collected by the telescopes are described further in Section 2.2.

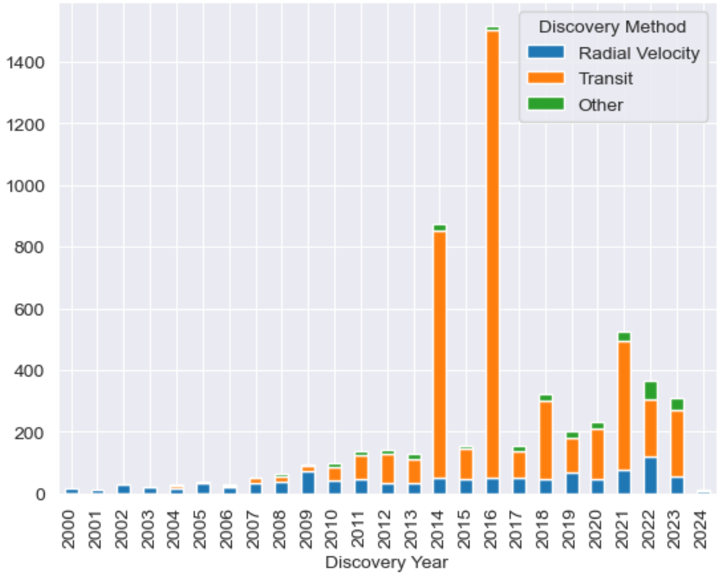


Figure 2.1. Confirmed exoplanet discoveries, grouped by method. Data from NASA (n.d.).

2.1 Exoplanet Detection Methods

2.1.1 Transit

A transit event is a planet passing between an observer and a star. Transits are detected using light curves — i.e., light from a star as a function of time. A typical transit can be characterized by its depth, duration, and period (which corresponds to the period of orbit around the star). When the planet passes in front of a star, the flux that arrives at the observer decreases from the normal value, which is the main evidence of a transit. Moreover, when the planet passes directly behind the star (half a period later), the flux from the system also slightly drops since the observer does not receive the light reflected from the planet; this event is called a secondary eclipse or occultation, illustrated by Figure 2.2.

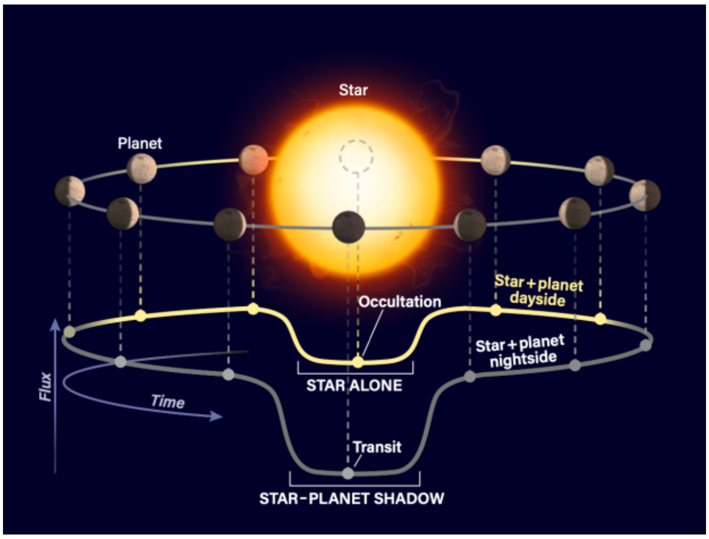


Figure 2.2. A planet transit. Credit: Roen Kelly for Astronomy

The transit method is considered an affordable method of exoplanet detection, as one transit survey allows for the monitoring of thousands of stars simultaneously. As illustrated by Figure 2.1, it is the most effective method of exoplanet discovery to date.

This method can also calculate the ratio between the size of the planet and its host star and orbital period. In some cases, the data can even reveal the stellar mass, stellar radius and orbital semi-major axis (Seager and Mallén-Ornelas, 2003).

The key limitation of the method is that it is restricted to planets whose orbital plane is inclined at a very specific angle to the observer unless the planet is very close to the star (Jara-Maldonado et al., 2020). Furthermore, larger planets have higher chance of being detected with this method.

Finally, special configurations of stars and some other events might produce a similar-looking dip in the light curve (Beky, 2014), which is the reason secondary observations are usually required to confirm the transit.

2.1.2 Radial Velocity

The radial velocity method considers small stellar wobbles caused by the alternating gravitational pull from an orbiting planet. The changes in radial velocity (i.e., velocity in the direction from or to the observer) affect the wavelengths of the absorbed spectrum of the stellar light via the Doppler effect. When the star is moving towards the observer, its absorption lines are blue shifted (towards lower wavelengths); when it is moving outwards, a red shift is observed.

The radial velocity (RV) method accounts for the first few hundred exoplanets ever confirmed, mostly with data from ground-based telescopes (Butler et al., 2006). Moreover, the first-ever exoplanet detection may have happened using the RV method (Mayor and Queloz, 1995). RV method is also commonly utilized to confirm planets that were initially discovered with other methods (such as the transit method) and to get more information about their characteristics, such as estimated mass-radius range (Rogers, 2015).

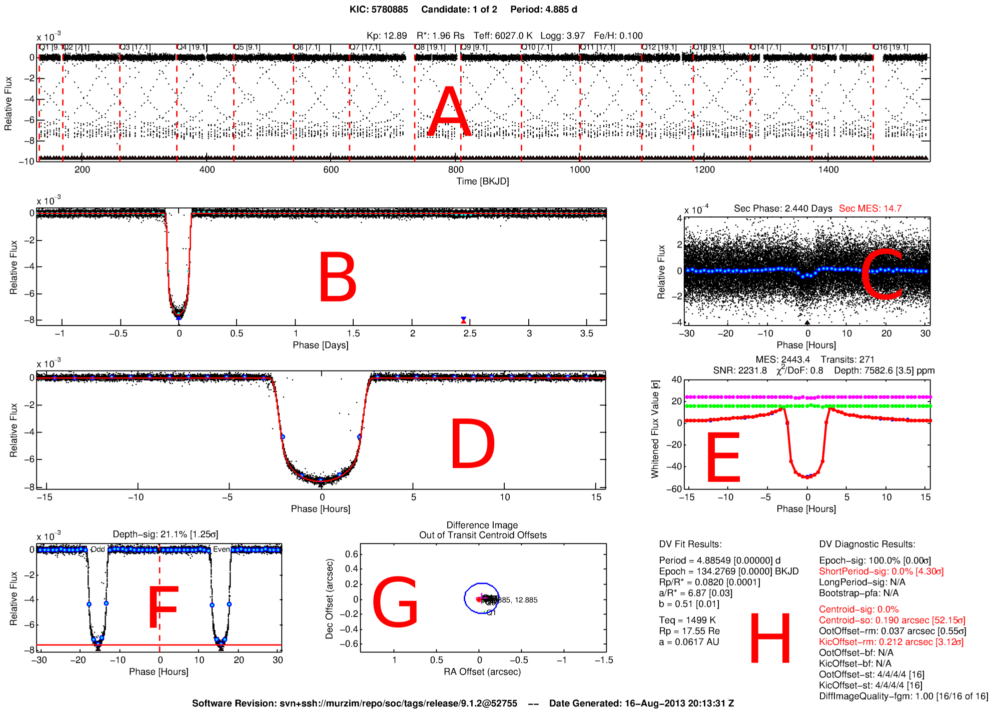
However, the RV method is connected to several challenges. Its main shortcoming is caused by the fact that the exoplanet-caused stellar wobbles are often too small to be detected, which raises dependency on precision of the telescope equipment (Jara-Maldonado et al., 2020). Moreover, the method detects planets with larger mass and smaller orbit radius more often because of a more noticeable gravitational impact. Second, because spectroscopy is pointed at the optical spectrum, the method is biased against stars with cooler temperatures since their light emission has higher wavelengths.

2.1.3 Other Methods of Exoplanet Detection

One of the less used methods is direct imaging, which attempts to obtain images of exoplanets, which may also produce information about their chemical composition and temperature. It requires high-contrast telescopes, because one of its main challenges is to spatially separate the planet from its star. Another method is gravitational microlensing, which exploits the gravitational impact of an exoplanet on the light of distant stars. One of its main issues is connected to rarity and unpredictability of such events when the impact is noticeable. Finally, timing variation and orbital brightness modulation are also sometimes used to detect exoplanets. The list of confirmed planets maintained by NASA includes planets detected by 11 different astrophysical methods (NASA, n.d.).

2.2 Transit data

The minimally processed transit data is an array of high-resolution images from telescopes. As Valizadegan et al. (2022) argues, this data is too high-dimensional to train a Machine Learning (ML) model on, considering the scarcity of labeled transit data. Thus, according to researchers, imaging data is processed through complex pipelines, such as the Kepler Science Processing Pipeline (Jenkins et al., 2010), which infers the light curves, runs basic tests, performs data cleaning, and generates so-called Data Validation (DV) reports on suspected transit events with acceptable signal-to-noise ratio — threshold-crossing events (TCE). A summary page of an example DV report can be seen in Figure 2.3.



KIC: 5780885 Candidate: 1 of 2 Period: 4.885 d

Kp: 12.89 R*: 1.96 Rs Teff: 6027.0 K Logg: 3.97 Fe/H: 0.100

Sec Phase: 2.440 Days Sec MES: 14.7

MES: 2443.4 Transits: 271

SNR: 2231.8 χ^2/DoF: 0.8 Depth: 7582.6 [3.5] ppm

Depth−sig: 21.1% [1.25σ]

Difference Image
Out of Transit Centroid Offsets

DV Fit Results:

Period = 4.88549 [0.00000] d
Epoch = 134.2769 [0.0000] BKJD
Rp/R* = 0.0820 [0.0001]
a/R* = 6.87 [0.03]
b = 0.51 [0.01]

Teq = 1499 K
Rp = 17.55 Re
a = 0.0617 AU

DV Diagnostic Results:

Epoch−sig: 100.0% [0.00σ]
ShortPeriod−sig: 0.0% [4.30σ]
LongPeriod−sig: N/A
Bootstrap−pfa: N/A

Centroid−sig: 0.0%
Centroid−so: 0.190 arcsec [52.15σ]
OotOffset−rm: 0.037 arcsec [0.55σ]
KicOffset−rm: 0.212 arcsec [3.12σ]
OotOffset−bf: N/A
KicOffset−bf: N/A
OotOffset−st: 4/4/4/4 [16]
KicOffset−st: 4/4/4/4 [16]
DiffImageQuality−fgm: 1.00 [16/16 of 16]

Software Revision: svn+ssh://murzim/repo/soc/tags/release/9.1.2@52755 -- Date Generated: 16−Aug−2013 20:13:31 Z

Figure 2.3. A DV summary report corresponding to transit of Kepler-7b, or KOI 97.01. (A) Full-time series plot; (B) Phased full-orbit flux plot; (C) Secondary eclipse plot; (D) Phased transit-only flux plot; (E) Whitened, phased transit-only plot; (F) Odd-even transit plot; (G) Centroid offset plot; (H) DV analysis table (NASA Exoplanet Archive, 2013).

Traditionally, scientists reviewed the DV reports and manually filtered out signals that appeared spurious (Jenkins et al., 2014), after which, the remaining data was released for follow-up studies in the form of KOI/TOI lists. This manual process may be augmented or replaced by algorithms described in further sections.

It is worth noting that due to the limited amount of available labeled transits, synthetically generated data is often used to train

ML models. Armstrong et al. (2021) uses interpolation to allow probability calibration at high precision. Moreover, Mandel and Agol (2002) developed analytic formulas that allow for simulating light curve data, which are used in the batman Python package for modeling exoplanet transits (Kreidberg, 2015). Light curves, KOI/TOI lists, TCEs, and other data for future transiting exoplanet research can be found on the website of NASA Exoplanet Archive[1].

[1] NASA Exoplanet Archive

3. Methodology

The thesis is a literature review of different algorithms of transiting exoplanet detection. The algorithms are split into categories according to their objectives and structure, as in Valizadegan et al. (2022). The first category is expert systems, which are based on if-else clauses and are heavily reliant on domain knowledge incorporated in them. The second category includes generative approaches, which use Bayes' theorem to generate the probability of a certain TCE being a true transit, based on prior assumptions about false positive (FP) scenarios. Finally, the discriminative algorithms category includes approaches that estimate the true transit probability using some underlying function, such as deep neural network.

Due to the nature of expert system and generative algorithms, they may only take scalar values as input. Thus, they are not able to process time series data or images. Depending on the underlying architecture, some discriminative approaches are also limited to scalar values, e.g., Autovetter (Jenkins et al., 2014). However, many recent deep neural network approaches, such as ExoMiner (Valizadegan et al., 2022) utilize the flux time series data in some form.

The papers have been found using the NASA Astrophysics Data System (ADS)[1] and prioritised by normalized citation count. Several papers were also included and prioritised if cited in other high-citation papers such as Valizadegan et al. (2022). Five key papers were chosen and elaborated on in Section 4, and a summarization table with these and five more papers may be found in Section 5.

[1] The SAO/NASA Astrophysics Data System

4. Results

4.1 Expert System Algorithms

The following section describes statistical algorithms that provide foundation of the automated exoplanet detection. These algorithms are designed to work with scalar input characteristics, such as FP test statistics and transit duration.

4.1.1 Robovetter

Robovetter[1] is the algorithm that was applied to automatically generate the KOI list from the final data release of the Kepler Space Telescope (Thompson et al., 2018a; NASA Exoplanet Archive). A ported version of Robovetter called TEC has also been applied to TESS data (Guerrero et al., 2021). The program essentially reproduces the steps of a typical human review of a DV report with if-else conditions. The high-level diagram of the algorithm (Figure 4.1) provides an insight into the basic tests for FPs.

Automation of the manual vetting process reduced cognitive load and human error and made large-data processing possible. However, replicating the manual process exposes this approach to similar bias caused by manually chosen rules and features.

[1] https://github.com/nasa/kepler-robovetter

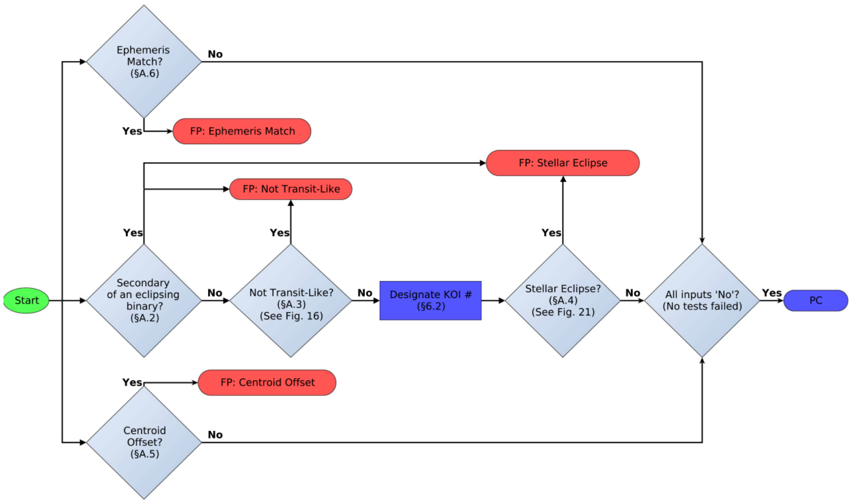


Figure 4.1. Overview flowchart of the Robovetter (Thompson et al., 2018a).

4.2 Generative Algorithms

4.2.1 Vespa

Vespa[2] is an influential library that calculates FP probability for every KOI (Morton et al., 2016). It assumes class priors $p(y = 1)$ and $p(y = 0)$ as well as likelihoods $p(X|y = 0)$ and $p(X|y = 1)$ to compute the posterior probability using the Bayes theorem:

$$p(y = 1|X) = \frac{p(y = 1) * p(X|y = 1)}{p(X)}.$$

X in this context denotes the input features from a DV report, and $p(y = 1)$ is the probability of the event being caused by a planetary transit. The algorithm is considered to confirm KOIs when the posterior probability is over 0.99. Other Bayesian algorithms for statistical exoplanet validation include PASTIS (Díaz et al., 2014), BLENDER (Torres et al., 2015 and TRICERATOPS (Giacalone et al., 2021).

[2] https://github.com/timothydmorton/vespa

4.3 Discriminative Algorithms

Several researchers have implemented machine learning approaches to transiting exoplanet detection. Jenkins et al. (2014) applied random forests in their Autovetter algorithm to identify the most important metrics for exoplanet status. In contrast with Robovetter (Coughlin et al., 2016), this algorithm makes fewer assumptions about the significance of different features. However, Autovetter still depends on the engineered features derived from DV reports and data pipelines.

Recent research has focused on deep learning, a subset of machine learning algorithms that are composed of complex processing layers; this structure allowed them to become state-of-the-art in many domains, such as object recognition and drug discovery (LeCun et al., 2015). Specifically, convolutional neural network (CNN) architectures seem to be most effective at classification of suspected transit events (Figure 4.2).

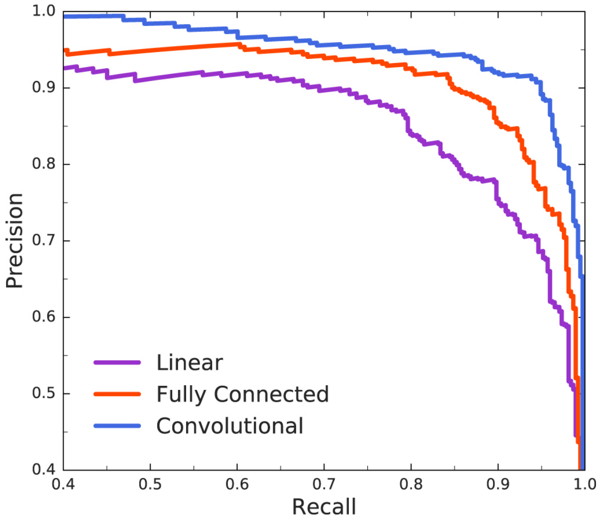


Figure 4.2. Precision vs. recall on a light curve dataset for neural networks with three different architectures. (Shallue and Vanderburg, 2018)

4.3.1 AstroNet

Shallue and Vanderburg (2018) were some of the first researchers to use neural networks and deep learning in exoplanet detection. Their model AstroNet[3] is a one-dimensional CNN. The model considers both local and global views of the phase-folded light curve, which are passed through two disjoint convolutional columns and then

[3] https://github.com/google-research/exoplanet-ml

combined through fully connected layers (Figure 4.3).

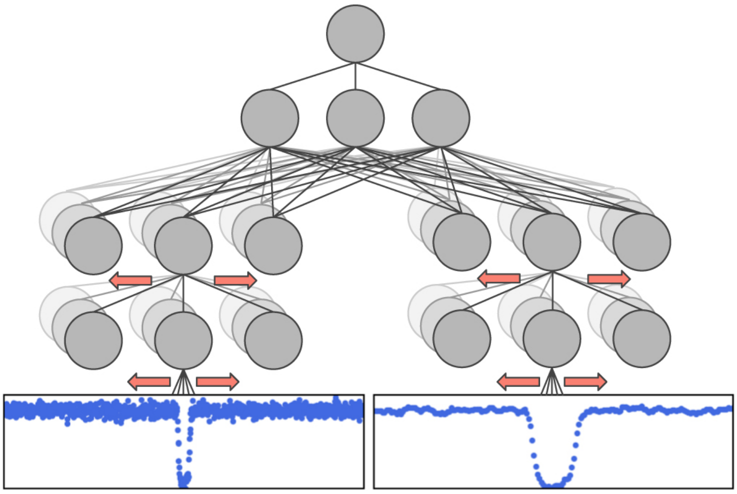


Figure 4.3. Convolutional neural network for light curve classification, with local and global views of the light curve. (Shallue and Vanderburg, 2018)

Researchers demonstrated that their model successfully learned the importance of secondary eclipse and the differences between true transits and FP cases, such as eclipsing binaries. However, their analysis shows that AstroNet still performs worse than Robovetter on simulated light curve data.

4.3.2 ExoMiner

Building on the main ideas of AstroNet, Valizadegan et al. (2022) designed a deep neural network ExoMiner that takes as inputs most of the parameters present in a typical DV report (Figure 4.4).

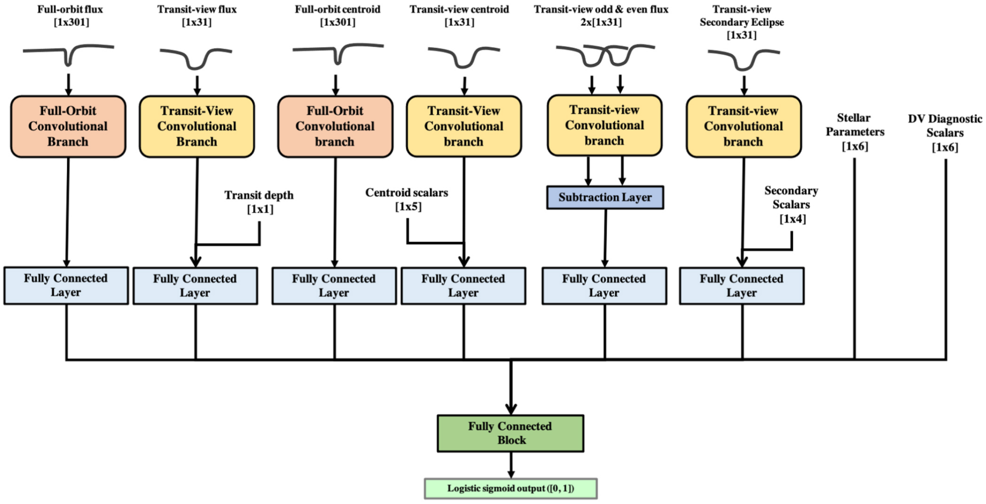

Full-orbit flux [1x301]
Transit-view flux [1x31]
Full-orbit centroid [1x301]
Transit-view centroid [1x31]
Transit-view odd & even flux 2x[1x31]
Transit-view Secondary Eclipse [1x31]
Full-Orbit Convolutional Branch
Transit-View Convolutional Branch
Full-Orbit Convolutional branch
Transit-View Convolutional branch
Transit-view Convolutional branch
Transit-view Convolutional branch
Stellar Parameters [1x6]
DV Diagnostic Scalars [1x6]
Transit depth [1x1]
Centroid scalars [1x5]
Subtraction Layer
Secondary Scalars [1x4]
Fully Connected Layer
Fully Connected Layer
Fully Connected Layer
Fully Connected Layer
Fully Connected Layer
Fully Connected Layer
Fully Connected Block
Logistic sigmoid output ([0, 1])

Figure 4.4. ExoMiner architecture. The model uses local (transit-view) and global (full-orbit) views of time series as well as scalar values. (Valizadegan et al., 2022)

Feature selection and general structure of the model have been informed by domain knowledge, while hyperparameters were optimized using Bayesian Optimization and Hyper-Band optimizer (Falkner et al., 2018).

Furthermore, researchers compared performance of ExoMiner and other classifiers including AstroNet (Shallue and Vanderburg, 2018), Robovetter (Coughlin et al., 2016) and GPC (Armstrong et al., 2021) on the Kepler Q1–Q17 DR25 TCE catalog with 10-fold cross-validation (Valizadegan et al., 2022). The analysis showed that ExoMiner has superior values in precision, recall, and other relevant metrics. For example, the accuracy value was, on average, 0.996, while that of Robovetter was 0.994. Moreover, the model is able to retrieve a significantly higher percentage of all exoplanets at a fixed precision value of 0.99.

Finally, researchers present an explainability framework tailored for ExoMiner based on a branch-occlusion sensitivity technique (Valizadegan et al., 2022). This addresses an important problem in machine learning algorithms for exoplanet detection that is caused by the high cost of misclassification and builds trust in the classifier.

In their later work, Valizadegan et al. (2023) explores the use of diversity boost information, as described in Rowe et al. (2014), to further improve the performance of classifiers mentioned above, including ExoMiner. Other works focused on the prediction of undiscovered exoplanets in multi-planetary systems include Mousavi-Sadr (2024).

4.3.3 Transformer-Based Classifier

Salinas et al. (2023) were the first researchers to apply self-attention mechanism to automatic classification of transit signals. Transformer deep learning models, which are based on the self-attention technique, proved to be successful in other fields, notably natural language processing (Gillioz et al., 2020). However, the architecture is applicable to any sequential data, including light curves. Just as neural networks attempt to mimic the way the human brain works, attention mechanisms try to reproduce the practice of selective focusing on some part of information while ignoring the rest. The approach is thus architecturally different from CNNs such as AstroNet or ExoMiner.

The inputs of the model are similar to those of ExoNet (Ansdell et al., 2018) - local and global views of the transit stacked with

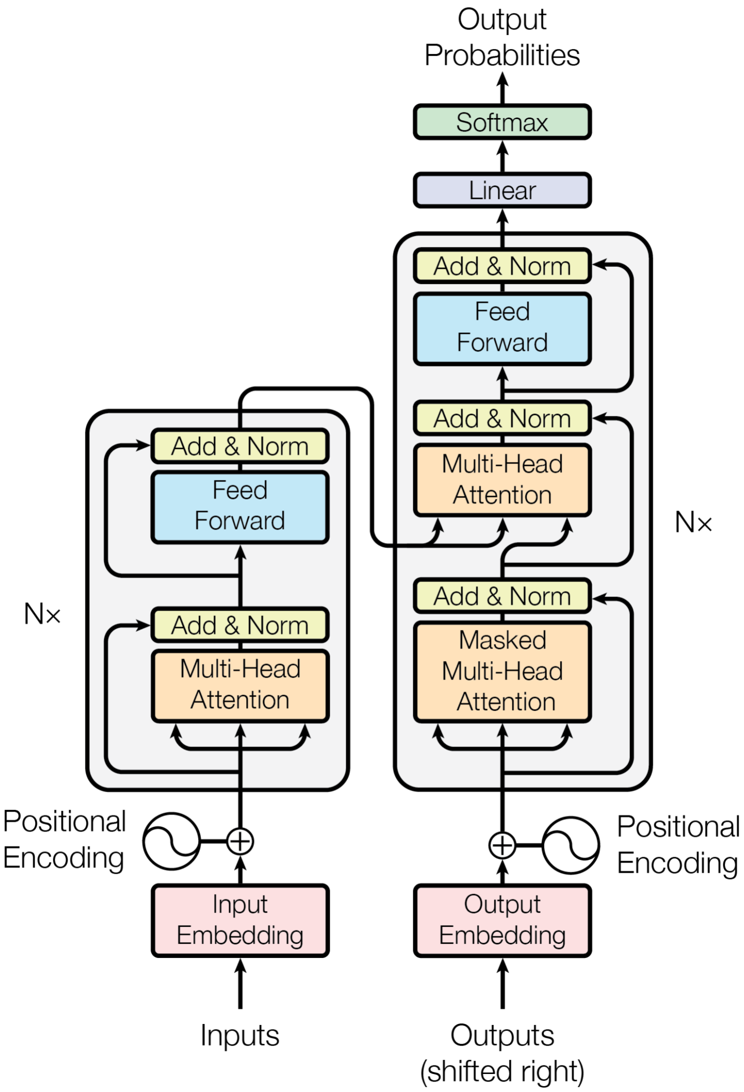


Figure 4.5. General Transformer model architecture. The model consists of encoder (on the left) and decoder (on the right), which use stacked self-attention mechanisms without convolution. (Vaswani et al., 2017)

centroid views, as well as stellar parameters as a separate branch of the model. The model is trained and evaluated on real data from the new TESS telescope, and it achieves results comparable to those of the state-of-the-art CNN solutions. As the research on Transformer models is currently rapidly growing, the technology may potentially perform better than the current state-of-the-art.

5. Discussion

This thesis described some of the most significant research papers in the field of automated exoplanet detection. A detailed comparison of these and several other works is shown on Figure 5.2.

It is worth noting that most works used a unique dataset, which obstructs immediate qualitative comparison of performance of models presented. In some cases, an updated version of the same dataset was released after the paper was published, such as the Kepler Data Release 25 (DR25) KOI catalog that was made public in 2018 (Thompson et al., 2018b). In other cases, the training and/or validation dataset has been based on or enriched with simulated data.

Several research papers compared their algorithm to others by running them on one dataset. For example, Ansdell et al. (2018) and Malik et al. (2021) used Astronet (Shallue and Vanderburg, 2018) as a benchmark. However, one of the most extensive qualitative comparisons of different classifiers to the author's knowledge was done by Valizadegan et al., 2022. Researchers ran six models on the same dataset and compared them across several metrics. Resulting accuracy values are shown on Figure 5.1. The plot demonstrates the incredible progress made in the field in the past decade, and shows that ExoMiner became one of the only models that outperform Robovetter, a rule-based classifier system. Furthermore, one can observe that many state-of-the-art models achieve very high, almost ideal metric values: each algorithm presented on Figure 5.1 has accuracy over 98%. Valizadegan et al. (2022) even claims that the area under receiver operating characteristic curve (AUC) is 1.000 for Exominer, which is the ideal value of the statistic.

Metrics such as AUC, accuracy, precision or recall may be suboptimal to compare the models in the field of exoplanet detection

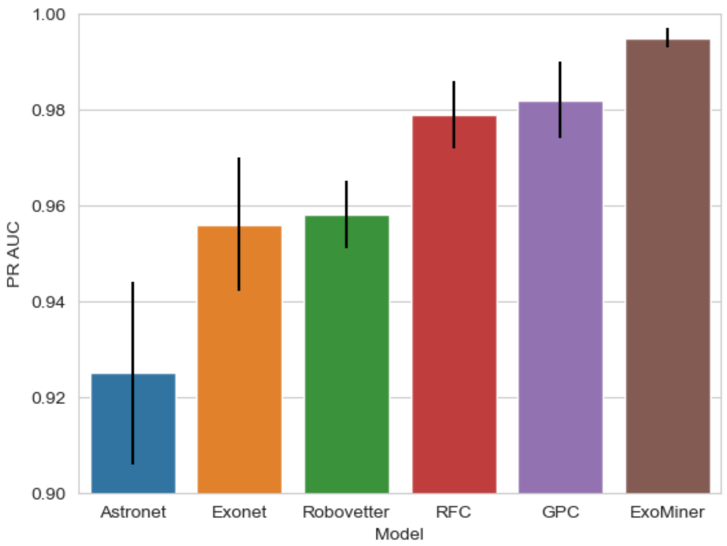
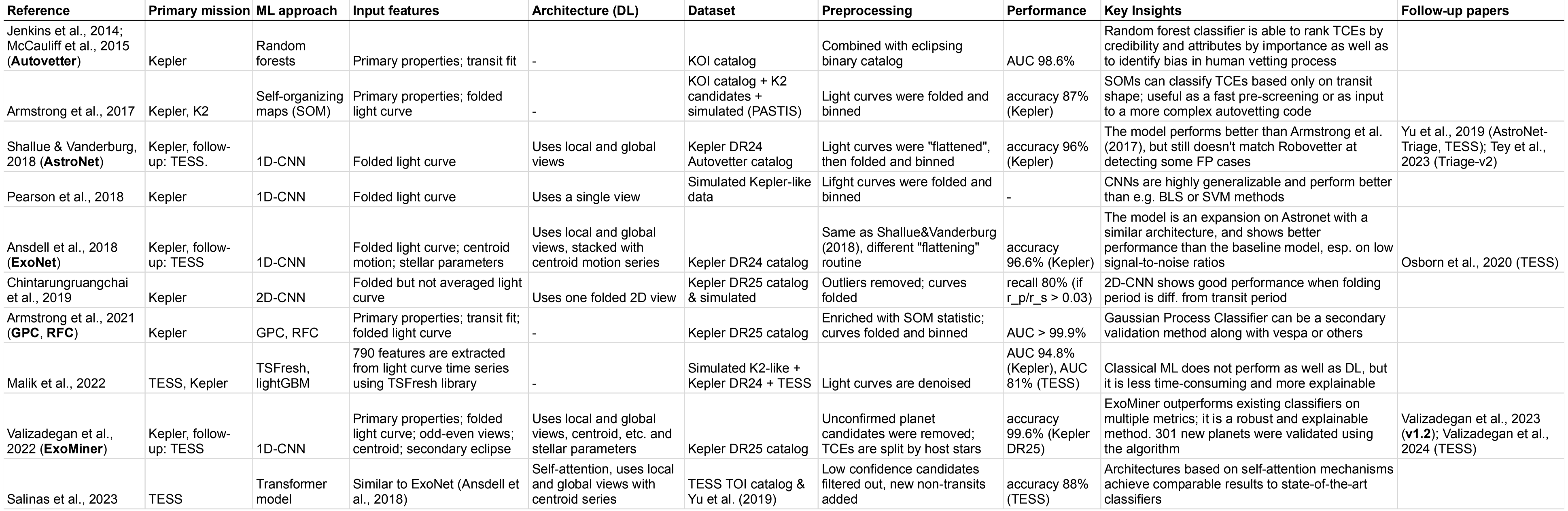


Figure 5.1. Area under precision-recall curve of several models on Kepler DR25 dataset used in Valizadegan et al. (2022). Mean values over folds in a 10-fold cross-validation are shown, black lines denote the standard deviation. **Note**: the Y axis of the graph starts with 0.90, which is already a very high value.

due to relatively low number of available labeled confirmed exoplanets. Since deep learning models usually require high amounts of training data for reliable results, researchers often use simulated light curves, which may prevent generalization on real data. Furthermore, even though, for example, Kepler mission has produced observations of over 100,000 stars (Borucki et al., 2010), only under 2,800 statistically confirmed planets were found by the mission (NASA, n.d.). Because machine learning usually requires a balanced dataset to produce reliable results, the researchers are often forced to consider only a similarly small amount of non-transit TCEs in training the models. Thus, in addition to the aforementioned metrics, many researchers consider aspects of their models such as explainability and training set sensitivity (Valizadegan et al., 2022) and examine the misclassified datapoints case by case (Armstrong et al., 2021).

Deep learning algorithms seem to perform significantly better in comparison with other machine learning or statistical methods, likely because (1) these models are able to almost independently learn feature importance and feature engineering and (2) they receive the light curve time series data instead of some summarizing statistic value. Figure 5.2 shows an extended list of automatic classifiers for exoplanet detection. It could be used in future research for design of new exoplanet detection models or benchmarking.



Reference	Primary mission	ML approach	Input features	Architecture (DL)	Dataset	Preprocessing	Performance	Key Insights	Follow-up papers
Jenkins et al., 2014; McCauliff et al., 2015 (**Autovetter**)	Kepler	Random forests	Primary properties; transit fit	-	KOI catalog	Combined with eclipsing binary catalog	AUC 98.6%	Random forest classifier is able to rank TCEs by credibility and attributes by importance as well as to identify bias in human vetting process	
Armstrong et al., 2017	Kepler, K2	Self-organizing maps (SOM)	Primary properties; folded light curve	-	KOI catalog + K2 candidates + simulated (PASTIS)	Light curves were folded and binned	accuracy 87% (Kepler)	SOMs can classify TCEs based only on transit shape; useful as a fast pre-screening or as input to a more complex autovetting code	
Shallue & Vanderburg, 2018 (**AstroNet**)	Kepler, follow-up: TESS.	1D-CNN	Folded light curve	Uses local and global views	Kepler DR24 Autovetter catalog	Light curves were "flattened", then folded and binned	accuracy 96% (Kepler)	The model performs better than Armstrong et al. (2017), but still doesn't match Robovetter at detecting some FP cases	Yu et al., 2019 (AstroNet-Triage, TESS); Tey et al., 2023 (Triage-v2)
Pearson et al., 2018	Kepler	1D-CNN	Folded light curve	Uses a single view	Simulated Kepler-like data	Lifght curves were folded and binned	-	CNNs are highly generalizable and perform better than e.g. BLS or SVM methods	
Ansdell et al., 2018 (**ExoNet**)	Kepler, follow-up: TESS	1D-CNN	Folded light curve; centroid motion; stellar parameters	Uses local and global views, stacked with centroid motion series	Kepler DR24 catalog	Same as Shallue&Vanderburg (2018), different "flattening" routine	accuracy 96.6% (Kepler)	The model is an expansion on Astronet with a similar architecture, and shows better performance than the baseline model, esp. on low signal-to-noise ratios	Osborn et al., 2020 (TESS)
Chintarungruangchai et al., 2019	Kepler	2D-CNN	Folded but not averaged light curve	Uses one folded 2D view	Kepler DR25 catalog & simulated	Outliers removed; curves folded	recall 80% (if r_p/r_s > 0.03)	2D-CNN shows good performance when folding period is diff. from transit period	
Armstrong et al., 2021 (**GPC**, **RFC**)	Kepler	GPC, RFC	Primary properties; transit fit; folded light curve	-	Kepler DR25 catalog	Enriched with SOM statistic; curves folded and binned	AUC > 99.9%	Gaussian Process Classifier can be a secondary validation method along with vespa or others	
Malik et al., 2022	TESS, Kepler	TSFresh, lightGBM	790 features are extracted from light curve time series using TSFresh library	-	Simulated K2-like + Kepler DR24 + TESS	Light curves are denoised	AUC 94.8% (Kepler), AUC 81% (TESS)	Classical ML does not perform as well as DL, but it is less time-consuming and more explainable	
Valizadegan et al., 2022 (**ExoMiner**)	Kepler, follow-up: TESS	1D-CNN	Primary properties; folded light curve; odd-even views; centroid; secondary eclipse	Uses local and global views, centroid, etc. and stellar parameters	Kepler DR25 catalog	Unconfirmed planet candidates were removed; TCEs are split by host stars	accuracy 99.6% (Kepler DR25)	ExoMiner outperforms existing classifiers on multiple metrics; it is a robust and explainable method. 301 new planets were validated using the algorithm	Valizadegan et al., 2023 (**v1.2**); Valizadegan et al., 2024 (TESS)
Salinas et al., 2023	TESS	Transformer model	Similar to ExoNet (Ansdell et al., 2018)	Self-attention, uses local and global views with centroid series	TESS TOI catalog & Yu et al. (2019)	Low confidence candidates filtered out, new non-transits added	accuracy 88% (TESS)	Architectures based on self-attention mechanisms achieve comparable results to state-of-the-art classifiers	

Figure 5.2. Comparison of selected discriminative algorithms of exoplanet classification. Primary properties include transit depth, period, MES.

6. Conclusion

Exoplanet detection is a complex task that requires understanding of astrophysics, statistics and machine learning. One of the most promising physical methods of exoplanet detection is the transit method, which attempts to find eclipses of observed stars that may signify presence of one or several orbiting planets. This study presented several legacy and state-of-the-art algorithms for classification of transiting exoplanet survey data. It may hopefully be useful to designers of future transiting exoplanet detection algorithms, as new data emerges and new survey missions are launched.

Deep learning techniques currently exhibit the best performance in transit signal classification task. State-of-the-art implementation for Kepler data, ExoMiner V1.2 (Valizadegan et al., 2023), is based on 1D-CNN architecture, however, emerging DL approaches, such as self-attention mechanisms (Salinas et al., 2023) also show high potential. The state-of-the-art for TESS data is Astronet-Triage-v2 (Tey et al., 2023), which is currently used in the MIT Quick-Look Pipeline. ExoMiner is also expected to be soon adapted to TESS data (Valizadegan et al., 2024).

This thesis does not provide an exhaustive list of all transiting exoplanet detection algorithms, focusing only on most influential papers. Moreover, the current work only provides a qualitative comparison of different models. Further research could focus on quantitative comparison, which would require compiling a single dataset that all models would be then trained and evaluated on. Finally, as the field of deep learning is rapidly evolving and transformer models are getting better at sequential data analysis, new models are likely to be worth applying to the problem of classification of suspected transit events.

References

Ansdell, M., Ioannou, Y., Osborn, H. P., Sasdelli, M., Team), (N. F. D. L. E., Smith, J. C., Caldwell, D., Jenkins, J. M., Räissi, C., Angerhausen, D., & Mentors), (N. F. D. L. E. (2018). Scientific domain knowledge improves exoplanet transit classification with deep learning. *The Astrophysical Journal Letters*, *869*(1), L7. https://doi.org/10.3847/2041-8213/aaf23b

Armstrong, D. J., Gamper, J., & Damoulas, T. (2021). Exoplanet validation with machine learning: 50 new validated Kepler planets. *Monthly Notices of the Royal Astronomical Society*, *504*(4), 5327–5344. https://doi.org/10.1093/mnras/staa2498

Beky, B. (2014). *Development and Application of Tools to Characterize Transiting Astrophysical Systems* [Doctoral dissertation, Harvard Smithsonian Center for Astrophysics].

Borucki, W. J., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Dupree, A. K., Gautier, T. N., Geary, J. C., Gilliland, R., Gould, A., Howell, S. B., Jenkins, J. M., Kondo, Y., . . . Prsa, A. (2010). Kepler planet-detection mission: Introduction and first results. *Science*, *327*(5968), 977–980. https://doi.org/10.1126/science.1185402

Borucki, W. J., Koch, D. G., Basri, G., Batalha, N., Brown, T. M., Bryson, S. T., Caldwell, D., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Gautier, T. N., Geary, J. C., Gilliland, R., Gould, A., Howell, S. B., Jenkins, J. M., Latham, D. W., Lissauer, J. J., . . . Still, M. (2011). Characteristics of planetary candidates observed by kepler. ii. analysis of the first four months of data. *The Astrophysical Journal*, *736*(1), 19. https://doi.org/10.1088/0004-637X/736/1/19

Butler, R. P., Wright, J. T., Marcy, G. W., Fischer, D. A., Vogt, S. S., Tinney, C. G., Jones, H. R. A., Carter, B. D., Johnson, J. A., McCarthy, C., & Penny, A. J. (2006). Catalog of Nearby Exoplanets. *The Astrophysical Journal*, *646*(1), 505–522. https://doi.org/10.1086/504701

Chintarungruangchai, P., & Jiang, I.-G. (2019). Detecting exoplanet transits through machine-learning techniques with convolutional neural networks. *Publications of the Astronomical Society of the Pacific*, *131*(1000), 1–15. Retrieved January 25, 2024, from https://www.jstor.org/stable/26660768

Coughlin, J. L., Mullally, F., Thompson, S. E., Rowe, J. F., Burke, C. J., Latham, D. W., Batalha, N. M., Ofir, A., Quarles, B. L., Henze, C. E., Wolfgang, A., Caldwell, D. A., Bryson, S. T., Shporer, A., Catanzarite, J., Akeson, R., Barclay, T., Borucki, W. J., Boyajian, T. S., . . . Zamudio, K. A. (2016). Planetary candidates observed by kepler. vii. the first fully uniform catalog based on the entire 48-month data set (q1–q17 dr24). *The Astrophysical Journal Supplement Series*, *224*(1), 12. https://doi.org/10.3847/0067-0049/224/1/12

Díaz, R. F., Almenara, J. M., Santerne, A., Moutou, C., Lethuillier, A., & Deleuil, M. (2014). PASTIS: Bayesian extrasolar planet validation - I. General framework, models, and performance. *Monthly Notices of the Royal Astronomical Society*, *441*(2), 983–1004. https://doi.org/10.1093/mnras/stu601

Falkner, S., Klein, A., & Hutter, F. (2018). BOHB: Robust and Efficient Hyperparameter Optimization at Scale. *arXiv e-prints*, Article arXiv:1807.01774, arXiv:1807.01774. https://doi.org/10.48550/arXiv.1807.01774

Giacalone, S., Dressing, C. D., Jensen, E. L. N., Collins, K. A., Ricker, G. R., Vanderspek, R., Seager, S., Winn, J. N., Jenkins, J. M., Barclay, T., Barkaoui, K., Cadieux, C., Charbonneau, D., Collins, K. I., Conti, D. M., Doyon, R., Evans, P., Ghachoui, M., Gillon, M., . . . Waite, I. A. (2021). Vetting of 384 TESS Objects of Interest with TRICERATOPS and Statistical Validation of 12 Planet Candidates. *The Astronomical Journal*, *161*(1), Article 24, 24. https://doi.org/10.3847/1538-3881/abc6af

Gillioz, A., Casas, J., Mugellini, E., & Khaled, O. A. (2020). Overview of the transformer-based models for nlp tasks. *2020 15th*

Conference on Computer Science and Information Systems (FedCSIS), 179–183. https://doi.org/10.15439/2020F20

Guerrero, N. M., Seager, S., Huang, C. X., Vanderburg, A., Soto, A. G., Mireles, I., Hesse, K., Fong, W., Glidden, A., Shporer, A., Latham, D. W., Collins, K. A., Quinn, S. N., Burt, J., Dragomir, D., Crossfield, I., Vanderspek, R., Fausnaugh, M., Burke, C. J., . . . Winn, J. N. (2021). The tess objects of interest catalog from the tess prime mission. *The Astrophysical Journal Supplement Series*, *254*(2), 39. https://doi.org/10.3847/1538-4365/abefe1

Jara-Maldonado, M., Alarcon-Aquino, V., Rosas-Romero, R., Starostenko, O., & Ramirez-Cortes, J. M. (2020). Transiting exoplanet discovery using machine learning techniques: A survey. *Earth Science Informatics*, *13*(3), 573–600. https://doi.org/10.1007/s12145-020-00464-7

Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., Twicken, J. D., Bryson, S. T., Quintana, E. V., Clarke, B. D., Li, J., Allen, C., Tenenbaum, P., Wu, H., Klaus, T. C., Middour, C. K., Cote, M. T., McCauliff, S., Girouard, F. R., Gunter, J. P., Wohler, B., Sommers, J., . . . Borucki, W. J. (2010). Overview of the Kepler Science Processing Pipeline. *The Astrophysical Journal Letters*, *713*(2), L87–L91. https://doi.org/10.1088/2041-8205/713/2/L87

Jenkins, J. M., McCauliff, S., Burke, C., Seader, S., Twicken, J., Klaus, T., Sanderfer, D., Srivastava, A., & Haas, M. R. (2014, April). Auto-Vetting Transiting Planet Candidates Identified by the Kepler Pipeline. In N. Haghighipour (Ed.), *Formation, detection, and characterization of extrasolar habitable planets* (pp. 94–99, Vol. 293). https://doi.org/10.1017/S1743921313012611

Kasting, J. F., Whitmire, D. P., & Reynolds, R. T. (1993). Habitable Zones around Main Sequence Stars. *Icarus*, *101*(1), 108–128. https://doi.org/10.1006/icar.1993.1010

Koch, D. G., Borucki, W. J., Basri, G., Batalha, N. M., Brown, T. M., Caldwell, D., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Gautier, I., Thomas N., Geary, J. C., Gilliland, R. L., Gould, A., Jenkins, J., Kondo, Y., Latham, D. W., Lissauer, J. J., Marcy, G., . . . Wu, H. (2010). Kepler Mission Design, Realized Photometric Performance, and Early Science. *The Astrophysical Journal Letters*, *713*(2), L79–L86. https://doi.org/10.1088/2041-8205/713/2/L79

Kreidberg, L. (2015). Batman: Basic transit model calculation in python. *Publications of the Astronomical Society of the Pacific*, *127*(957), 1161. https://doi.org/10.1086/683602

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Malik, A., Moster, B. P., & Obermeier, C. (2021). Exoplanet detection using machine learning. *Monthly Notices of the Royal Astronomical Society*, *513*(4), 5505–5516. https://doi.org/10.1093/mnras/stab3692

Mandel, K., & Agol, E. (2002). Analytic light curves for planetary transit searches. *The Astrophysical Journal*, *580*(2), L171. https://doi.org/10.1086/345520

Mayor, M., & Queloz, D. (1995). A Jupiter-mass companion to a solar-type star. *Nature*, *378*(6555), 355–359. https://doi.org/10.1038/378355a0

Morton, T. D., Bryson, S. T., Coughlin, J. L., Rowe, J. F., Ravichandran, G., Petigura, E. A., Haas, M. R., & Batalha, N. M. (2016). False positive probabilities for all kepler objects of interest: 1284 newly validated planets and 428 likely false positives. *The Astrophysical Journal*, *822*(2), 86. https://doi.org/10.3847/0004-637X/822/2/86

Mousavi-Sadr, M., Gozaliasl, G., & Jassur, D. M. (2021). Exoplanets prediction in multiplanetary systems. *Publications of the Astronomical Society of Australia*, *38*, Article e015, e015. https://doi.org/10.1017/pasa.2021.9

Mousavi-Sadr, M. (2024). Exoplanets Prediction in Multi-Planetary Systems and Determining the Correlation Between the Parameters of Planets and Host Stars Using Artificial Intelligence. *arXiv e-prints*, Article arXiv:2402.17898, arXiv:2402.17898. https://doi.org/10.48550/arXiv.2402.17898

NASA. (n.d.). NASA Exoplanet Archive [[Accessed 19-02-2024]]. https://exoplanetarchive.ipac.caltech.edu/index.html

NASA. (2019). Kepler / K2 - NASA Science [[Accessed 08-02-2024]]. https://science.nasa.gov/mission/kepler/

NASA, ESA, & Leah Hustak (STScl). (2023, November). Nasa's hubble measures the size of the nearest transiting earth-sized planet [[Accessed 07-03-2024]]. https://science.nasa.gov/missions/hubble/nasas-hubble-measures-the-size-of-the-nearest-transiting-earth-sized-planet/

NASA Exoplanet Archive. (2013, November). Kepler Data Validation One-Page Summary Reports [[Accessed 12-02-2024]]. https://exoplanetarchive.ipac.caltech.edu/docs/DVSummaryPageCompanion_q1_q16.html

NASA Exoplanet Archive. (2024). Kepler objects of interest cumulative table. https://doi.org/10.26133/NEA4

NExScI. (2024). Exoplanet follow-up observing program web service. https://doi.org/10.26134/EXOFOP5

Pearson, K. A., Palafox, L., & Griffith, C. A. (2017). Searching for exoplanets using artificial intelligence. *Monthly Notices of the Royal Astronomical Society*, *474*(1), 478–491. https://doi.org/10.1093/mnras/stx2761

Pollacco, D. L., Skillen, I., Cameron, A. C., Christian, D. J., Hellier, C., Irwin, J., Lister, T. A., Street, R. A., West, R. G., Anderson, D., Clarkson, W. I., Deeg, H., Enoch, B., Evans, A., Fitzsimmons, A., Haswell, C. A., Hodgkin, S., Horne, K., Kane, S. R., . . . Wilson, D. M. (2006). The wasp project and the superwasp cameras. *Publications of the Astronomical Society of the Pacific*, *118*(848), 1407. https://doi.org/10.1086/508556

Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. Á., Bean, J. L., Berta-Thompson, Z. K., Brown, T. M., Buchhave, L., Butler, N. R., Butler, R. P., Chaplin, W. J., Charbonneau, D., Christensen-Dalsgaard, J., Clampin, M., Deming, D., Doty, J., De Lee, N., Dressing, C., . . . Villasenor, J. (2015). Transiting Exoplanet Survey Satellite (TESS). *Journal of Astronomical Telescopes, Instruments, and Systems*, *1*, Article 014003, 014003. https://doi.org/10.1117/1.JATIS.1.1.014003

Rogers, L. A. (2015). Most 1.6 Earth-radius Planets are Not Rocky. *The Astrophysical Journal*, *801*(1), Article 41, 41. https://doi.org/10.1088/0004-637X/801/1/41

Rowe, J. F., Bryson, S. T., Marcy, G. W., Lissauer, J. J., Jontof-Hutter, D., Mullally, F., Gilliland, R. L., Issacson, H., Ford, E., Howell, S. B., Borucki, W. J., Haas, M., Huber, D., Steffen, J. H., Thompson, S. E., Quintana, E., Barclay, T., Still, M., Fortney, J., . . . Geary, J. (2014). Validation of kepler's multiple planet candidates. iii. light curve analysis and announcement of hundreds of new multi-planet systems. *The Astrophysical Journal*, *784*(1), 45. https://doi.org/10.1088/0004-637X/784/1/45

Salinas, H., Pichara, K., Brahm, R., Pérez-Galarce, F., & Mery, D. (2023). Distinguishing a planetary transit from false positives: a Transformer-based classification for planetary transit signals. *Monthly Notices of the Royal Astronomical Society*, *522*(3), 3201–3216. https://doi.org/10.1093/mnras/stad1173

Seager, S., & Mallén-Ornelas, G. (2003). A unique solution of planet and star parameters from an extrasolar planet transit light curve. *The Astrophysical Journal*, *585*(2), 1038. https://doi.org/10.1086/346105

Shallue, C. J., & Vanderburg, A. (2018). Identifying exoplanets with deep learning: A five-planet resonant chain around kepler-80 and an eighth planet around kepler-90. *The Astronomical Journal*, *155*(2), 94. https://doi.org/10.3847/1538-3881/aa9e09

Tey, E., Moldovan, D., Kunimoto, M., Huang, C. X., Shporer, A., Daylan, T., Muthukrishna, D., Vanderburg, A., Dattilo, A., Ricker, G. R., & Seager, S. (2023). Identifying Exoplanets with Deep Learning. V. Improved Light-curve Classification for TESS Full-frame Image Observations. *The Astronomical Journal*, *165*(3), Article 95, 95. https://doi.org/10.3847/1538-3881/acad85

Thompson, S. E., Coughlin, J. L., Hoffman, K., Mullally, F., Christiansen, J. L., Burke, C. J., Bryson, S., Batalha, N., Haas, M. R., Catanzarite, J., Rowe, J. F., Barentsen, G., Caldwell, D. A., Clarke, B. D., Jenkins, J. M., Li, J., Latham, D. W., Lissauer, J. J., Mathur, S., . . . Borucki, W. J. (2018a). Planetary Candidates Observed by Kepler. VIII. A Fully Automated Catalog with Measured Completeness and Reliability Based on Data Release 25. *The Astrophysical Journal Supplement Series*, *235*(2), Article 38, 38. https://doi.org/10.3847/1538-4365/aab4f9

Thompson, S. E., Coughlin, J. L., Hoffman, K., Mullally, F., Christiansen, J. L., Burke, C. J., Bryson, S., Batalha, N., Haas, M. R., Catanzarite, J., Rowe, J. F., Barentsen, G., Caldwell, D. A., Clarke, B. D., Jenkins, J. M., Li, J., Latham, D. W., Lissauer, J. J., Mathur, S., . . . Borucki, W. J. (2018b). Planetary Candidates

Observed by Kepler. VIII. A Fully Automated Catalog with Measured Completeness and Reliability Based on Data Release 25. *The Astrophysical Journal Supplement Series*, *235*(2), Article 38, 38. https://doi.org/10.3847/1538-4365/aab4f9

Torres, G., Kipping, D. M., Fressin, F., Caldwell, D. A., Twicken, J. D., Ballard, S., Batalha, N. M., Bryson, S. T., Ciardi, D. R., Henze, C. E., Howell, S. B., Isaacson, H. T., Jenkins, J. M., Muirhead, P. S., Newton, E. R., Petigura, E. A., Barclay, T., Borucki, W. J., Crepp, J. R., . . . Quintana, E. V. (2015). Validation of 12 small kepler transiting planets in the habitable zone. *The Astrophysical Journal*, *800*(2), 99. https://doi.org/10.1088/0004-637X/800/2/99

Valizadegan, H., Martinho, M., Yates, C., Zhong, W., Jenkins, J., Caldwell, D., Twicken, J., & Bryson, S. (2024). Classification of TESS Threshold Crossing Events Using ExoMiner++. *American Astronomical Society Meeting Abstracts*, *56*, Article 421.04, 421.04.

Valizadegan, H., Martinho, M. J. S., Jenkins, J. M., Caldwell, D. A., Twicken, J. D., & Bryson, S. T. (2023). Multiplicity boost of transit signal classifiers: Validation of 69 new exoplanets using the multiplicity boost of exominer. *The Astronomical Journal*, *166*(1), 28. https://doi.org/10.3847/1538-3881/acd344

Valizadegan, H., Martinho, M. J. S., Wilkens, L. S., Jenkins, J. M., Smith, J. C., Caldwell, D. A., Twicken, J. D., Gerum, P. C. L., Walia, N., Hausknecht, K., Lubin, N. Y., Bryson, S. T., & Oza, N. C. (2022). Exominer: A highly accurate and explainable deep learning classifier that validates 301 new exoplanets. *The Astrophysical Journal*, *926*(2), 120. https://doi.org/10.3847/1538-4357/ac4399

Vannah, S., Gleiser, M., & Kaltenegger, L. (2023). *Monthly Notices of the Royal Astronomical Society: Letters*, *528*(1), L4–L9. https://doi.org/10.1093/mnrasl/slad156

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf