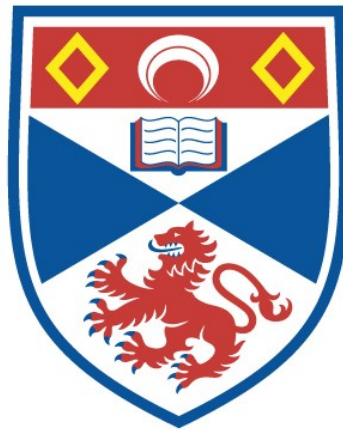# From candidate identification to planet characterization: a machine learning approach

Nicole Schanche

A thesis submitted for the degree of PhD
at the
University of St Andrews

2020

Candidate's declaration

I, Nicole Schanche, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 27,500 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2016.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date   Signature of candidate

 29-Sept-2020

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date   Signature of supervisor

 29-Sept-2020

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Nicole Schanche, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date   Signature of candidate

  29-Sept-2020

Date   Signature of supervisor

  29-Sept-2020

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Nicole Schanche, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date   Signature of candidate

29-Sept-2020

Abstract

Fun will now commence.

- 7 of 9, *Star Trek VOY: Ashes to Ashes*

This thesis is broken into three main sections tracing the steps of the development of a new framework to search for and characterize planets from the WASP survey. While all methods were developed specifically for the WASP project, the principles are easily transferable to any ground or space based survey. In the first part of the thesis, I discuss the development of two machine learning methods, a Random Forest Classifier and a Convolutional Neural Network, that are able to find new exoplanet candidates from WASP archival data and lightcurves. In preparing the training dataset, I also created a standardized catalog of 1,041 false positives from SuperWASP, the northern component of WASP, that were verified with additional observations from other instruments.

The second part of the thesis begins by discussing the results of the machine learning methods. In the analysis of the resulting probabilities, several patterns began to emerge that indicated the differing predictions between the algorithms carry useful information in itself. This realization sparked the development of a new "stacking" framework where the predictions of a number of different machine learning methods are used as the input to a second-level classifier that makes the final prediction. This method is straightforward to implement and demonstrates an improved performance over any individual classifier, making the stacked approach ideal for future large surveys. I use the model to classify and rank more than 100,000 lightcurves in the WASP archive that do not yet have a disposition associated with them and discuss the candidates that are rated most favourably.

Finally, in part three I discuss what to do once a candidate is confirmed to be a planet. In particular, I describe a new MCMC method that combines the likelihood fits of transit and radial velocity data with prior knowledge from several sources including optical and infrared spectrophotometric measurements and the parallax measurements from *Gaia* to constrain the stellar parameters. I apply the method to characterize two new hot Jupiter planets found by the WASP collaboration and confirmed with SOPHIE and *TESS* measurements. WASP-186b is a dense ($4.22 \pm 0.18 M_J$, $1.11 \pm 0.03 R_J$) planet on an eccentric (e=0.33 ± 0.01) 5-day orbit around a mid-F type star. While also in a $\sim$5 day orbit, WASP-187b is puffed up ($0.8 \pm 0.09 M_J$, $1.64 \pm 0.05 R_J$) and orbiting a star that has begun evolving away from the main sequence.

Acknowledgements

Collaboration Statement

This thesis is the result of my own work carried out at the University of St Andrews between September 2016 and July 2020. The text in the chapters has been written entirely by me. All Figures, unless explicitly stated in the Figure caption have been produced by me.

The work in Chapter 2 combines the work of two publications:

"SuperWASP dispositions and false positive catalogue", Schanche, N.; Collier Cameron, A.; Almenara, J. M.; Alsubai, K. A.; Anderson, D. R.; Armstrong, D. J.; Barkaoui, K.; Barros, S. C. C.; Bochiski, J.; Bonomo, A. S.; Bouchy, F.; Brown, D. J. A.; Burdanov, A.; Busuttil, R.; Deleuil, M.; Delrez, L.; Faedi, F.; Gillon, M.; Hay, K.; Hebb, L. Hébrard, G.; Jehin, E.; Kolb, U.; Maxted, P. F. L.; Miller, G.; Nielsen, L. D.; Pollacco, D. L.; Pozuelos, F. J.; Queloz, D.; Relles, H.; Smalley, B.; Triaud, A. H. M. J.; Udry, S.; West, R.; Wheatley, P. J., MNRAS, Volume 488, Issue 4, p.4905-4915, October 2019.

"Machine-learning approaches to exoplanet transit detection and candidate validation in wide-field ground-based surveys", Schanche, N.; Collier Cameron, A.; Hébrard, G.; Nielsen, L.; Triaud, A. H. M. J.; Almenara, J. M.; Alsubai, K. A.; Anderson, D. R.; Armstrong, D. J.; Barros, S. C. C.; Bouchy, F.; Boumis, P.; Brown, D. J. A.; Faedi, F.; Hay, K.; Hebb, L.; Kiefer, F.; Mancini, L.; Maxted, P. F. L.; Palle, E. Pollacco, D. L.; Queloz, D.; Smalley, B.; Udry, S.; West, R.; Wheatley, P. J., Monthly Notices of the Royal Astronomical Society, Volume 483, Issue 4, p.5534-5547, March 2019.

The planet discoveries presented in Chapter 4 have also been submitted to MNRAS for publication.

"WASP-186 and WASP-187: two hot Jupiters discovered bySuperWASP and SOPHIE with additional observations by TESS", N. Schanche; G. Hébrard; A. Collier Cameron; S. Dalal; B. Smalley; T. G. Wilson; I. Boisse; F. Bouchy; D.J.A Brown; O. Demangeon; C.A. Haswell; C. Hellier; U.C. Kolb; T. Lopez; P.F.L. Maxted; D.L. Pollacco; R.G. West; and P.J. Wheatley, MNRAS, accepted September 2020.

The text describing these works was modified from the original papers to fit the narrative of this thesis.

The MCMC code described in Chapter 4 was written primarily by A. Collier-Cameron, and modified by me to incorporate WASP and SOPHIE data. The original analysis of the SOPHIE data used to characterize WASP-186b and WASP-187b was performed by G. Hébrard and S. Dalal with spectral analysis performed by B. Smalley. The isochrone fitting to determine the stellar masses and ages was done by T. Wilson.

All other authors participated in the SuperWASP collaboration and had input in the manuscripts of the papers.

Contents

List of Figures

List of Tables

1

Introduction

There is no such thing as the unknown. Only things temporarily hidden, temporarily not understood.

- Captain Kirk, *Star Trek TOS: The Corbomite Maneuver*

1.1 The History of Transiting Exoplanets

When the search for exoplanets began in earnest in the 1980s and 90s (Campbell et al., 1988; Latham et al., 1989; Marcy & Benitz, 1989), little was known about their prevalence in the galaxy. In fact, when the Hubble spacecraft launched in 1990, there were still no unambiguous detections of exoplanets. Nearly two years later in 1992, the first detection of a planet orbiting another star was reported; however, the host star was a pulsar and not a typical main sequence star (Wolszczan & Frail, 1992). The first planet orbiting a main sequence star, discovered by the variation in measured radial velocity, soon followed in 1995 (Mayor & Queloz, 1995). At the time, this discovery was very unexpected. With a mass of roughly half of that of Jupiter, this planet was orbiting every 4 days. Up to that point, large planets were expected to be much farther away from their host stars, as the gas giants are in our solar system. The discovery of this first "hot Jupiter" led to increased interest in and development of a different discovery technique: planetary transits.

Struve (1952) suggested well before the first confirmed hot Jupiter detection that close-in planets should be detectable by radial velocity measurements and, if oriented correctly, by transits when the planet passes in front of the star. This vision was realized in 2000 with the transit observation of HD 209458b (Charbonneau et al., 2000; Henry et al., 2000), which was already a known radial-velocity planet with a period around 3.5 days. It would take 3 years for the first planet (OGLE-TR-56b with a period of 1.2 days) to be discovered via the transit method, accomplished with data from the Optical Gravitational Lensing Experiment and confirmed as a planet with spectral observations (Konacki et al., 2003). By the time this announcement was made, more than two dozen groups had already begun development of ground-based transit searches (Horne, 2003).

In the years since, many intriguing close-in (P<10 days) hot Jupiter planets have been discovered through transit surveys that show a great diversity amongst systems. For example, the density of the observed planet population varies drastically, with known planets less than 0.06 times as dense as Jupiter (WASP-127b; Lam et al. (2017)) to more than 10 times the density of Jupiter (HAT-P-20b; Bakos et al. (2011)). Massive close-in planets have been found orbiting a range of stars, from M-dwarfs (NGTS-1; (Bayliss et al., 2018a)) to hot, young A stars like WASP-33 (Collier Cameron et al., 2010) and KELT-9 (Gaudi et al., 2017). Ultra-hot planets with even shorter periods were soon discovered, such as WASP-19b (Hebb et al., 2010) with a period of 0.79 days and Kepler-1520/KIC-12557548 (Rappaport et al., 2012), a possibly disintegrating planet showing a comet-like tail with a period of 0.65 days. Planets orbiting so close to their host star were thought to lie in circular orbits due to tidal interactions with the host. However, early on several planets such as CoRoT-16b (e~0.33) (Ollivier et al., 2012) were discovered with significant eccentricities. The discovery of this vast array of objects has provided clues to the formation and migration mechanisms of planetary systems.

1.2 The Search for Hot Jupiters

Because transits are only seen when the system orientation is such that the planet passes in front of its host star from the perspective of the observer, the probability (Pr) of a transit scales with the exoplanet's orbital distance a:

$$Pr = \left(\frac{R_* + R_p}{a}\right)\left(\frac{1 + e\sin\omega}{1 - e^2}\right) \tag{1.1}$$

and is related to the stellar and planetary radii (R_* and R_p) and the shape and orientation of the orbit described by the eccentricity e and argument of periastron ω (Perryman, 2018). This equation is often simplified to $\frac{R_*}{a}$ with the assumption that the orbit is circu-

lar and the star's radius is much greater than the planet's. Furthermore, the light blocked by the planet is tied to the ratio of the planet's radius to the host star's radius:

$$\frac{\Delta F}{F} = \frac{R_p^2}{R_*^2} \tag{1.2}$$

This means that transit searches are biased toward planets with large radii close to their host stars, and therefore ideally suited to detect hot Jupiters.

The earliest estimations for hot Jupiter occurrence rates came from radial velocity surveys, establishing that while hot Jupiters are not unique, they are not ubiquitous either. Marcy et al. (2005) provided an early estimation that $1.2 \pm .2\%$ of FGK stars host hot Jupiters (defined as orbits within 0.1 AU) based on a long term radial velocity survey of over 1000 stars. For the same orbital separation, Cumming et al. (2008) find that $1.5 \pm 0.6\%$ percent of FGK stars host hot Jupiters of a mass greater than $0.3 M_J$. Several years later, results from the California Planet Survey's (CPS) Lick and Keck searches find a similar rate of $1.2 \pm 0.38\%$ (Wright et al., 2012).

Even after correcting for the geometrical probability of observation, the estimations from transit surveys generally have lower estimates. Bayliss & Sackett (2011) cite an occurrence rate of $0.10^{+0.27}_{-0.08}\%$ based on simulations of the detection efficiency of the SuperLupus Survey. The estimated rate found by the Kepler survey is slightly higher, with a value of $0.4 \pm 0.1\%$, although this value just considers GK dwarfs (Howard et al., 2012). The California-Kepler Survey (CKS), a campaign to follow up on Kepler planets and their host stars, performed a detailed analysis of planet occurrence rates, finding hot Jupiters to occur in $0.57^{+0.14}_{-0.12}\%$ of stars (Petigura et al., 2018). The rate extrapolated from the results of the CoRoT mission approach the rates determined from RV surveys at $1 \pm 0.3\%$ (Moutou et al., 2013).

Petigura et al. (2018) emphasize the role that stellar metallicity plays in the distribution of planets of various masses, with higher metallicities tied to an increased number of large planets, although it is uncertain the extent to which this impacts all of the various occurrence rate studies. Bouma et al. (2018) demonstrate that unresolved stellar binaries are a major source of error in estimating occurrence rates, and likely a major cause for the discrepancies seen across estimates.

Regardless, all estimates agree that hot Jupiters are relatively scarce. If we take a rough guideline of 1 in 100 stars hosting hot Jupiters, and further take into account the system orientation, only around 10% of hot Jupiters that exist are expected to exhibit transits. Therefore wide surveys covering tens to hundreds of thousands of stars are

needed to detect a population of hot Jupiters.

Because these planets have deep transits, short periods, and transit durations on the order of a few hours, ground-based surveys such as the Hungarian-made Automated Telescope Network (HATnet; Hartman et al., 2004), the Wide-Angle Search for Planets (WASP; Pollacco et al., 2006), the Kilodegree Extremely Little Telescope (KELT; Pepper et al., 2007), the Qatar Exoplanet Survey (QES; Alsubai et al., 2013), HATSouth (Bakos et al., 2013), and the Next-Generation Transit Survey (NGTS; Wheatley et al., 2018) are well suited to their discovery.

Not limited by atmospheric effects or the day-night cycle, space-based missions such as CoRoT (Convection, Rotation, and planetary Transits; Auvergne et al., 2009), *Kepler* (Borucki et al., 2010), and *Kepler's* extended mission K2 (Howell et al., 2014) have pushed the boundaries further discovering planets with smaller radii and at larger orbital separations. The population of transiting exoplanets now contains more than 350 planets found by ground surveys and nearly 3000 by space based surveys, the distribution of which is shown in Figure 1.1. The ground-based planet discoveries are clustered in the upper left of the plot, demonstrating the effectiveness of these surveys for Jupiter-sized planets in orbits of less than 10 days. The small smattering of orange points below $0.2\ R_J$ are from the 7 planets of the TRAPPIST-1 system (Gillon et al., 2017) and the two planets around LHS 1140 (Dittmann et al., 2017; Ment et al., 2019) discovered by the MEarth project, both programs dedicated to finding small planets around M-dwarf stars.

1.3 Detecting a Transit

Transit surveys operate by taking regular measurements of the brightness of all stars in a given field. After performing calibration and detrending measures, the problem remains of how to actually detect the transit events as the uncertainties associated with the data can be greater than the transit depth, and the transit itself can be buried in the variations caused by stellar activity. Further, ground observations are taken at varying time intervals with breaks in observing due to the daylight hours, weather events, maintenance operations, and seasonal visibility, causing transits to often only be observed in part or missed entirely. The development of algorithms to detect signals in these conditions is of primary importance for the success of these surveys.

An early method proposed to find transits is known as Phase Dispersion Minimization (PDM), originally developed to find periods of unusual variable stars such as RR Lyrae (Lafler & Kinman, 1965; Jurkevich, 1971; Warner & Robinson, 1972; Stellingwerf, 1978). The general principle of this technique is to find the period that has the least observational scatter. However, unlike variable stars, transits are short duration events with long gaps
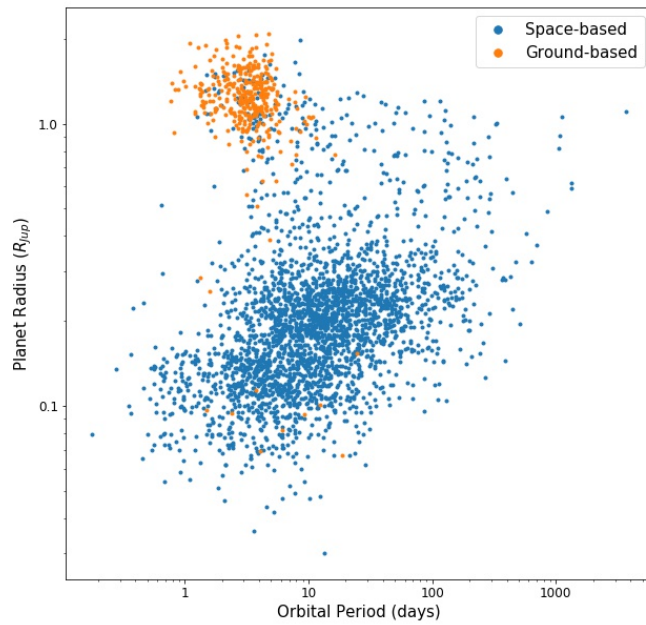
Figure 1.1: Distribution of planets discovered by the transit method as of May 19, 2020. Blue points indicate those found by space-based missions while orange points show those discovered by ground-based surveys.

5

between successive transit dips, with the majority of data points being "constant" (although with stellar variability). New methods specializing in finding this pattern needed to be developed.

Defaÿ et al. (2001) proposed a new Bayesian method for the purpose of detecting exoplanets in the CoRoT data, assuming uniformly sampled data and the presence of only Gaussian white noise, limiting its effectiveness against astrophysical noise. Aigrain & Favata (2002) similarly developed a Bayesian method for the (later cancelled) *Eddington* space telescope. While it similarly assumed white noise, they found their implementation was able to handle small gaps in data on the order of hours, but did not test the performance using longer gaps characteristic of ground observations.

To address this, several new approaches were developed. Schwarzenberg-Czerny & Beaulieu (2006) modified their analysis of variance (AoV) periodogram search, which handles missing data by phase folding and binning the data. Phase binning of the data was also utilized by Kovács et al. (2002) in their Box-Least-Squares algorithm, discussed in greater detail in Chapter 2. This method makes the simplistic assumption that the transit is perfectly box-shaped with a uniform depth and all out-of-transit points have the same baseline value. Cabrera et al. (2012) developed a modification of this approach treating the transit as a parabola rather than a box. Hippke & Heller (2019) proposed the Transit Least Squares (TLS) method variation to take into account the effect of limb darkening with the template optimized to identify small planets, making the method most suited to *Kepler*/K2 and *TESS* rather than ground searches. However, other templates can be created to optimize for other search spaces.

Several further techniques are applicable to both ground and space-based surveys. The matched-filter approach (Jenkins et al., 1996; Doyle et al., 2000; Jenkins et al., 2002; Bordé et al., 2007) compares the observed lightcurves with synthetic models of transits. The wavelet technique, as the name suggests, first performs a wavelet transform, from which the period is then searched. This method is used by the CoRoT mission (Régulo et al., 2007), while *Kepler* employs a wavelet based search using matched-filtering (Jenkins et al., 2010). Caceres et al. (2019b) present a hybrid method, combining a matched comb filter with machine learning (Random Forest) to identify candidates. In the last several years, several groups have been employing machine learning techniques for the primary identification of transit shapes for surveys conducted from both space (Pearson et al., 2018) and ground (Armstrong et al., 2018).

Regardless of the method used, the goal is the same: to identify shallow, periodic dips in the lightcurve. However, all shallow dips are not created equal. Several different sources can give rise to similar signals. Determining the source of the candidate requires
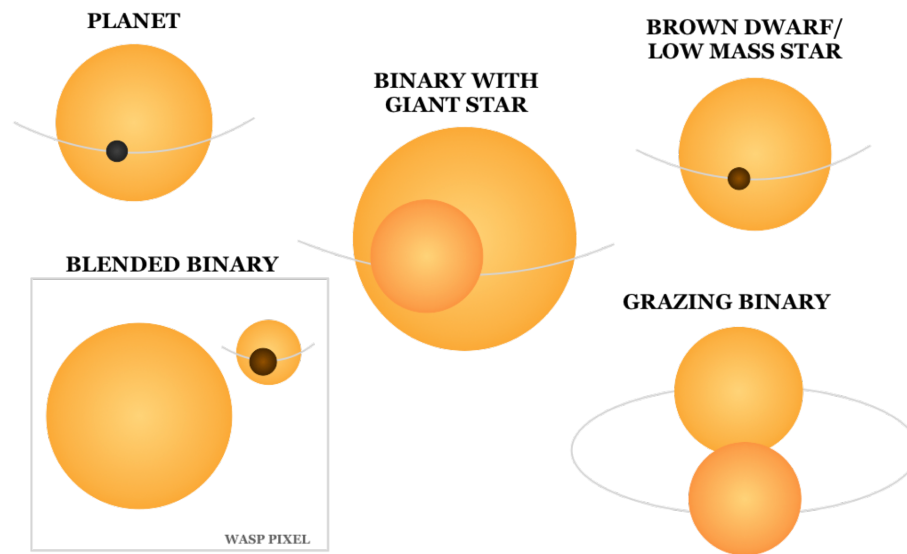
Figure 1.2: Types of astrophysical false positives affecting transit searches. Image based on Figure 1 of Collier Cameron (2012).

the additional step of vetting the data for false positives.

1.4 False Positives

There are several astrophysical sources that can masquerade themselves as a planet, known as false positives (See Fig. 1.2). The main sources include: two main-sequence stars that are oriented in such a way that the transit is only grazing (bottom right of the figure), the transit of a giant star by a main-sequence star (center), or the dilution of the signal from either of the aforementioned scenarios by a third source whose light is blended with the transiting source (bottom left).

Ground-based searches utilizing a "shallow-but-wide" approach such as WASP are particularly susceptible to blending scenarios due to the large pixel size. Evans & Sackett (2010) estimated that such a search strategy will result in false detections at a rate 10 times that of true planet detections, with blending the primarily source. Even spaced-based missions are susceptible to false detections - The *Kepler* catalog reports more than 3,500 certified false positives in their catalog[1], and more the one third of *TESS* planet candidates are expected to be astrophysical false positives (Sullivan et al., 2015).

Even with an undiluted signal, other objects can mimic the signal of a planet transit. Figure 1.3 highlights the difficulty caused by the radius overlap amongst puffy Jupiters, brown dwarfs, and low-mass stars. The lower left of the figure (lower mass and radius range) shows the rocky planets. The black, dark gray, and light grey lines show the theoretical relationship between planets composed of iron, rock, and ice respectively (Fortney

---

[1]https://exoplanetarchive.ipac.caltech.edu
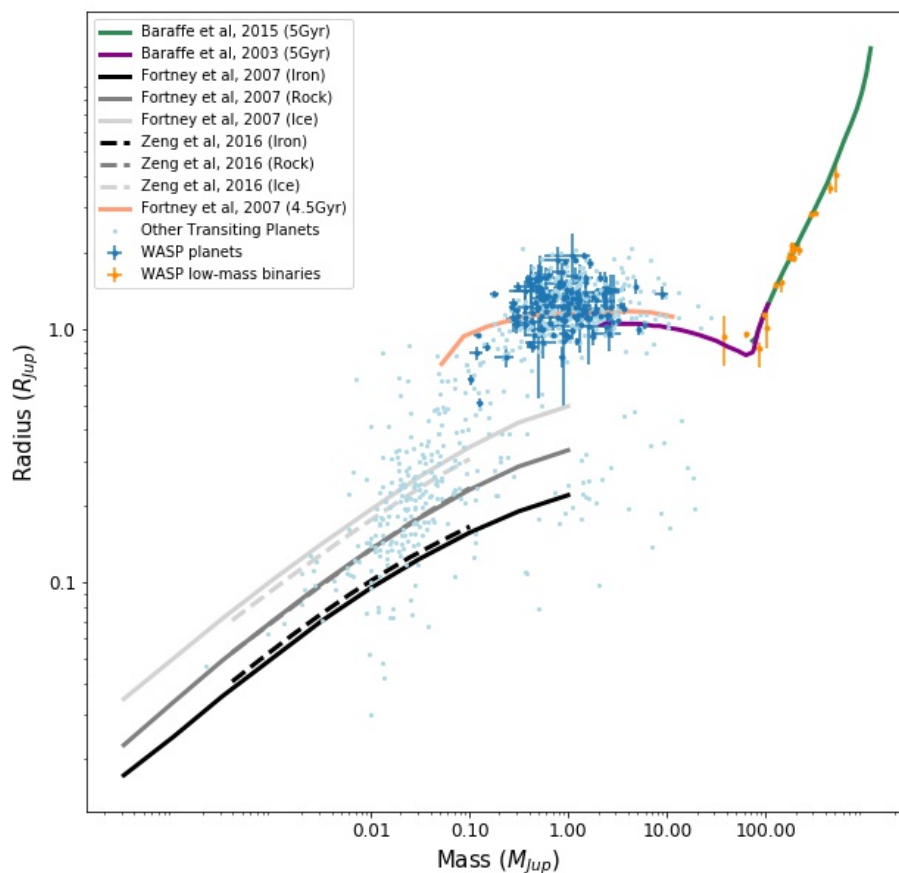
Figure 1.3: Plot showing the mass versus the radius for transiting planets (dark blue for WASP planets, light blue for planets found by other surveys) and low mass eclipsing binaries (orange). Overplotted are theoretical models for icy, rocky, iron, and giant planets as well as low-mass stellar objects.

et al., 2007; Zeng et al., 2016). The relationship with radius and mass increasing together breaks down at a mass of roughly half a Jupiter mass, as can be seen by the theoretical model by Fortney et al. (2007), shown in orange. This model assumes a system age of 4.5 Gyr, a core fraction of 10%, and an orbital distance of 0.02 au. Planets at this mass and greater plateau with a radius of just over that of Jupiter.

The largest planets known to date have an upper limit of $\sim 2R_{Jup}$. Beyond the planetary mass range lie the more massive brown dwarfs, ranging from 13 to $80M_{Jup}$, which are supported by electron-degeneracy pressure and have radii comparable to large planets. Objects with masses above this limit achieve core temperatures high enough to sustain thermonuclear fusion before electron degeneracy sets in and become small stars (See, for

example, Chabrier & Baraffe (2000)). The purple and green lines (Baraffe et al., 2003, 2015) trace out the theoretical relationship in mass and radius for these substellar objects and small stars. The orange points on the plot show the low-mass eclipsing binaries (Triaud et al., 2013; Gómez Maqueo Chew et al., 2014; von Boetticher et al., 2017, 2019; Gill et al., 2019) and brown dwarfs (Triaud et al., 2013; Hodžić et al., 2018) identified by the WASP survey. The nature of these objects was not known until radial velocity measurements were taken showing that the masses were super-planetary.

In addition to the overlapping radii range, close-in planets have daytime temperatures that are similar to small stars and can reflect a large amount of stellar flux (Triaud et al., 2017). Because of this, photometry alone can not distinguish the source unless there is significant ellipsoidal variation evident in the host star from the strong gravitational interaction with the secondary. However, ellipsoidal variation is not always a telltale sign of a brown dwarf or small star; WASP-18b is a high mass planet in a very close orbit, producing significant ellipsoidal variation in its host star (Shporer et al., 2017).

Several techniques have been developed over time to reduce the false alarm rate. For example, candidates from the WASP survey undergo numerous checks (Collier Cameron et al., 2006; Christian et al., 2006) for their transit depth, signal-to-noise and signal-to-red noise ratio, ellipsoidal variation induced by a massive companion, blending in the target pixels through aperture photometry, and stellar radius (either via the relationship between the proper motion and J-H color estimates or more recently *Gaia* parallax measures). Even with all of these checks, it is still not always possible to distinguish genuine planets. Figure 1.4 shows real false positives from the WASP survey, including a non-astrophysical source (bottom right), due to systematics in the WASP data. All samples show a reasonable U-shaped transit at a plausible depth, and were only revealed as false positives after additional observations.

The difficulty in identifying genuine planets has been reflected in several different ground-based surveys. Of 41 sources that pass all the aforementioned WASP threshold criteria, Christian et al. (2006) analyzed the 12 planetary candidates deemed most promising from the first season of SuperWASP North data. Of them, it is now known that only two are confirmed planets, six are eclipsing binaries, two are low mass stars, one is caused by blending with a nearby eclipsing binary, and one is a non-astrophysical false positive. Bayliss et al. (2018b) reported similar findings from K2 observations of HATSouth candidates. They were able to find 1 planet, 3 remained candidates but required confirmation of mass, and 18 candidates were found to be eclipsing binaries or blends. O'Donovan et al. (2006) provided a cautionary tale, where a false positive was only identified after noting the different transit depths by different wavelength photometry, indicating that the
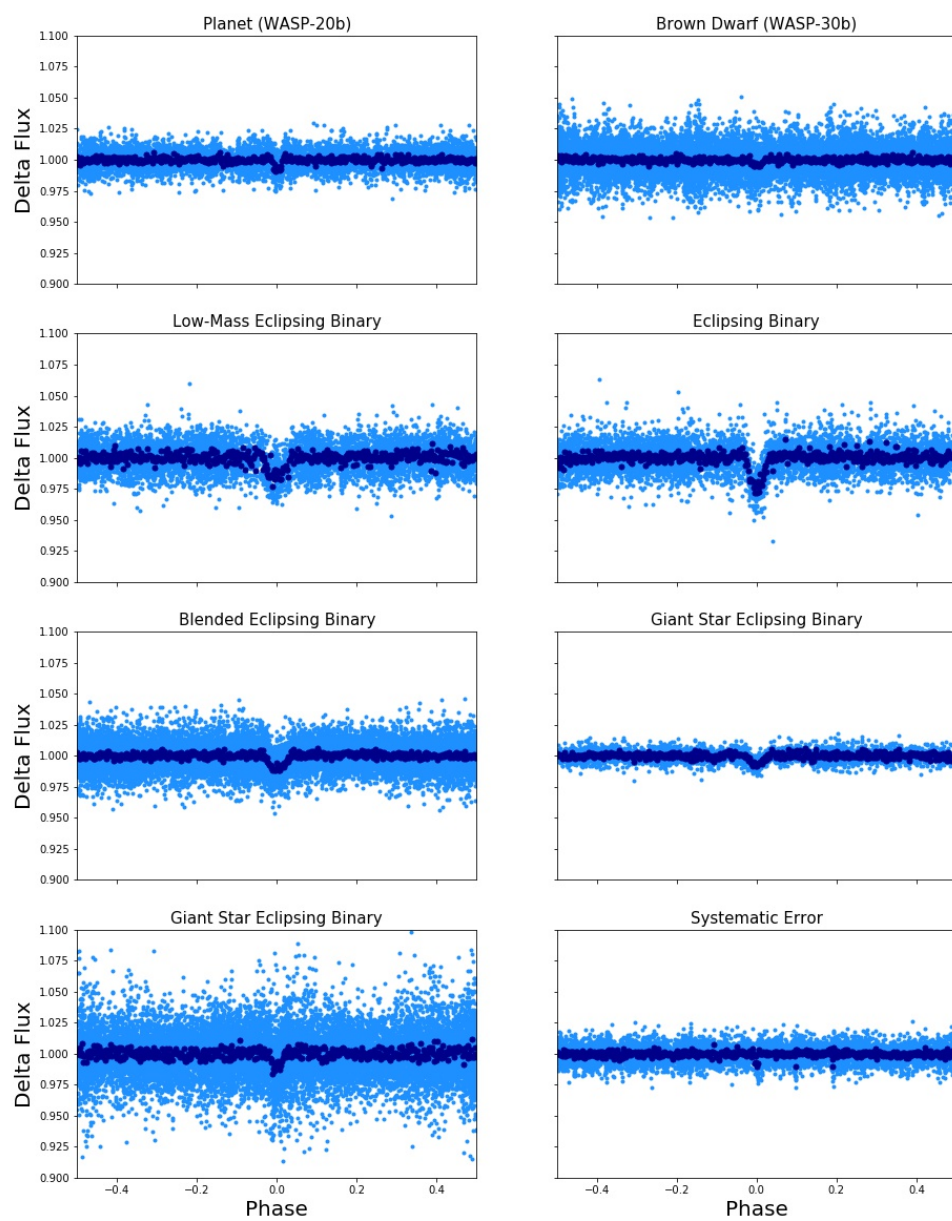
Figure 1.4: Various types of false positives seen in WASP data that mimic the signal from a planet transit.

"planet" they discovered was in fact due to a blend.

These findings only represent small samples from ground surveys. However, survey-long records have begun to be published, sharing the false positive results so that other groups need not spend follow-up time on the same objects. The KELT team presented a catalog of 1,128 false positives, while there are 26 confirmed planets (Collins et al., 2018) for an overall success rate of near 2%. Schanche et al. (2019a) reported a total of 1,041 false positives in the SuperWASP north survey in contrast to the 54 known planets, giving an overall success rate of ∼5%. This work includes objects reported by the EBLM project (Triaud et al., 2017), which aims to characterize low-mass stellar companions originally found in the WASP survey.

1.5 Thesis Outline

Clearly the process of identifying an exoplanet has many complications. Each survey generates tens to hundreds of thousands of lightcurves that need to be processed and checked for transits. While many lightcurves can be quickly eliminated through automated thresholding of various features, thousands of events are left for manual inspection. The human eyeballing step is a time intensive task that is subject to variations in individual assessments.

In this thesis, I present a machine learning method to automatically vet lightcurves from the WASP survey to identify planet candidates while minimizing the number of potential false positives. Chapter 2 describes the WASP survey and the data set used for training the machine learning algorithms, including information about the WASP false positive catalog. The chapter also describes the Random Forest Classifier and Convolutional Neural Networks used to make preliminary classifications for candidates. Chapter 3 describes the performance of these two methods, motivating the development of a "stacking" method, in which the results of multiple machine learning models are combined and used for a second-level classifier. We apply the stacked model to over 100,000 Super-WASP lightcurves and discuss the resulting top 12 planetary candidates. In chapter 4 we review how transit and radial velocity data are used to understand planetary systems, and describe an MCMC method that is able to fit both datasets simultaneously while also incorporating prior knowledge of the stellar angular diameter and distance from optical and infrared spectrophotometry and parallax measurements from *Gaia*. This method is used to characterize two new exoplanets, WASP-186b and WASP-187b. Finally, we address how modern space-based missions are revolutionizing the way follow-up programs for ground surveys are conducted. We conclude with a summary of the works presented here, and address how stacking models can be used to increase the planet yield for transit

searches such as *TESS* and *PLATO*.

2

Methods

"When a child is taught, it's programmed with simple instructions, and at some point, if its mind develops properly, it exceeds the sum of what it was taught, thinks independently."

- Dr. Daystrom, Star Trek TOS: The Ultimate Computer

In the last decade, machine learning has become an increasingly utilized tool in astronomy. While the language around machine learning can be mystifying, in reality it is a very simple concept. In essence, machine learning is just using knowledge about known data to make some statement or prediction about new data. For example, fitting a line to a set of data points is a simple type of machine learning, with the form "given x, what is y?" Of course many real-world situations there are hundreds or thousands of features with complex associations that are too complicated to be adequately modeled with a simple line fit. These increasingly large data sets led to interest in a data-driven approach, in which the model can adapt to the data and identify patterns without needing explicit human intervention. It is for these reasons that the field of machine learning has exploded.

There are two broad categories of machine learning - supervised and unsupervised. Supervised machine learning requires a ground-truth "training" dataset which has a large

number of examples labeled as the type of thing that you are interested in ("Y"). For prediction purposes, the descriptive features of the training dataset ("X") are used to fit a model to maximize correct output labels, while in regression the goal is to use the features to predict a continuous output value. In unsupervised learning, the labels are not known. Instead, the algorithm uses the features to group the data in some way in order to make inferences about the data. In the case of identifying planets in the WASP dataset, we have a specific output label we want to obtain (planet or other false positive). This problem falls squarely in the realm of supervised machine learning.

This chapter will discuss the data produced by the WASP collaboration in the context of planet detection. Using the archive of all transit candidates and the extensive catalog of false positives already in existence, including the SuperWASP false positive catalog, we implement two different types of machine learning, Random Forest Classifiers and Convolutional Neural Networks, to identify new planet candidates in previously unlabeled data.

2.1 The WASP Dataset

The Wide Angle Search for Planets (WASP; Pollacco et al., 2006) project has been ground-breaking in the field of exoplanets. Since its first observing season in 2004, nearly 200 planets have been discovered, the most of any ground-based transit survey. The WASP consortium operates two telescope sites, one in each hemisphere. SuperWASP, the northern component at the Observatorio del Roque de los Muchachos on La Palma and WASP-S, the southern observatory at the Sutherland station of the South African Astronomical Observatory were constructed as low-cost observatories, and as such the telescopes are made up of 8 commercial Canon 200mm f/1.8 cameras mounted together to cover 482 square degrees of sky with each exposure. WASP was designed to have a "shallow-but-wide" approach to search for exoplanets, meaning that it is targeting stars that are bright ($V < \sim 13$) across the entire sky. This is important as the stars targeted by WASP are bright enough for further characterization by radial velocity surveys. By combining information for millions of stars over thousands to tens of thousands of observations, WASP has beaten down the noise in the data to identify close-in hot Jupiter planets.

WASP's raw image data are corrected using bias frames, thermal dark-current exposures, and flat field exposures (Collier Cameron et al., 2006). The lightcurves are then extracted for all stars in the USNO-B1.0 catalog (Monet et al., 2003) that have a red band magnitude less than 15. Each pixel on WASP's CCD is large, covering an area of sky of $\sim$13.5 arcseconds. In order to help detect whether a star's light has been blended with other nearby stars, the lightcurves are extracted from three apertures of 2.5, 3.5, and 4.5

pixels surrounding the target star. The ratio of the brightness in each ring can be a good indicator of whether the signal is a product of blending with a nearby star. Additional trends in the photometry, such as extinction effects, color response, and system zero point are removed in the WASP pipeline by tying each frame to standard stars in the field and relating their derived flux to the the TYCHO-2 data (Høg et al., 2000).

The noise described above is uncorrelated white noise. However, there are also low level systematic sources of noise, termed red noise. This red noise has been shown to be extremely detrimental to transit detection (Pont et al., 2006) as the systematic noise is hard to remove and tends to vary on the order of a few hours, the same as the transit duration of a typical hot Jupiter planet. To try to correct for the red noise, the systematics removal algorithm of Tamuz et al. (2005) is applied to the WASP data. This is an iterative approach to correct smoothly varying systematic effects as a function of time. This is effective in removing the most significant sources of red noise that affect all stars in the field; however, some sources of noise still remain in the final data product.

Once the instrumental effects are corrected for, the transit search begins. This is done using a modification of the Box-Least Squares (BLS) method (Kovács et al., 2002) that uses a measure of χ^2 rather than the signal residue, as described by Collier Cameron et al. (2006) and outlined below. The BLS method is a brute-force algorithm that iterates through a grid of periods, transit durations, and transit epochs in search of the strongest box-shaped signal, a simplification of the shape of a planetary transit. The simple model consists of datapoints inside the possible transit (l) and those outside of the transit. The mean value and associated variance inside these regions (L for low and H for high) is defined by

$$L = \frac{s}{r}, \quad \text{Var}(L) = \frac{1}{r} \quad H = \frac{-s}{t-r}, \quad \text{Var}(H) = \frac{1}{t-r} \tag{2.1}$$

where

$$t = \sum_i w_i, \quad s = \sum_{i \in l} x_i w_i, \quad r = \sum_{i \in l} w_i \tag{2.2}$$

with w_i defined as the inverse-variance weights and x_i as the observation values corrected with the optimal average value. The depth of the transit can then be estimated as

$$\delta = L - H = \frac{st}{r(t-r)}, \quad \text{Var}(\delta) = \frac{t}{r(t-r)} \tag{2.3}$$

The signal-to-noise ratio of the transit depth is then given by

$$\text{S/N} = s\sqrt{\frac{t}{r(t-r)}} \tag{2.4}$$

Assuming the noise is white, the improvement of fit as compared to a constant lightcurve is

$$\Delta\chi^2 = \frac{s^2 t}{r(t-r)} \tag{2.5}$$

At each trial period, the best $\Delta\chi^2$ value for all trial epochs and transit durations is saved, with the overall largest amplitude value being interpreted as the best-fitting transit.

Since the BLS method is computationally expensive and because transit signatures are smaller at larger transit periods and are therefore harder to distinguish from noise, the WASP search grid only goes out to 16 days; in practice signals greater than 10 days tend to be spurious leading to the exclusive discovery of very close-in planets.

The lightcurve folded on the best-fitting period from the BLS algorithm forms the basis of the human vetting. The lightcurve for each star is accessible to the WASP team in an online repository. In addition to the folded lightcurve and the BLS periodogram, information about the star pulled from other publicly available sources, such as the magnitude, mass, and radius estimates, and information about the potential secondary object based on the BLS fit, such as the period, transit duration, and object radius, is provided to aid the user.

An example of the webpage generated for each object is shown in Figure 2.1. The different data processing runs can be accessed by the links shown in the upper right of webpage. The two plots featured prominently in the center of the page represent the lightcurve folded on the best-fitting period (all data in black, binned data shifted up for visibility in blue; left) and the periodogram from the BLS run (right). Above the plots is a table with stellar parameters scraped from the NOMAD catalog. Below the plots are tables listing system parameters determined by the Hunter catalog (in gray) and an MCMC model to refine the estimates (in blue). Users are also able to generate other relevant information from the links in the upper left. For example, "Thumbnails" produces a series of 3 images of the star in different wavelengths to check for visible blends. From the information provided, users can flag the star for further study or reject it as a false positive. Users are also able to contribute comments or upload additional data from follow-up efforts.

1SWASPJ144746.57+010353.9

Field = J144746, Cam = 300, H_run = ORFG_TAMTFA

SW Vt=12.8272 | Pts_gd=22612 | TSTART=2008-03-08 02:35:20
| TSTOP=2009-07-25 20:52:07 | P_{min}=0.35 | P_{max}=10
Thumbnails | VSI2.0 | Aperture Blends? | Nearby SW objects? | RPM | RPM(new) |
Blends
Param. Fit | Phase predictor | Transit scheduler | Hunt1star i | SWBLS | SWFOLD

Alternative Fields/Cams i
J144746 300 22612 (ORFG_TAMTFA)
J144746 300 22593 (ORFG_TAMUZ)
J144746 300 22696 (ORFG_TFA)
MF1444+0221 300 6419
(ORION_TAMTFA)
OF1434+0143 300 39685
(ORCA_TAMTFA)
SW1438-0246 141 5355 (10OR_TAMTFA)
SW1441+0129 141 3922 (08LNAT2)
SW1441+0129 141 3922 (08NAT2)
SW1447+0314 226 2692 (08NAT2)
SW1447+0314 226 2692 (08SNAT2)
SW1447+0314 226 5766
(09OR_TAMTFA)
SW1448+0559 143 5315
(09OR_TAMTFA)
SW1448+0559 143 5443
(10OR_TAMTFA)
WASP37B 300 64012 (ORKP_TAMUZ)
WASP37B 300 65903 (ORKP_TFA)

i	Teff JH	Teff VK	R_{star} JH	R_{star} VK	V mag	J mag	V-K	J-H	MU_RA (mas/yr)	MU_DEC (mas/yr)	RPMJ	RPMJ diff	Giant?	Dw:Gi	Dil. V	Dil. R
	5799 = G3	5737 = G4	1.02	0.99	12.75	11.50	1.66	0.32	-23.2 (± 5.4)	22.8 (± 5.4)	4.06	-7.64	0	70:0	0 %	0 %

Period: 3.577446 Epoch i: 4533.695 Re-plot

Pk1 | **Pk2** | **Pk3** | **Pk4** | **Pk5** | MCMC | Updated
Px2 | P/2 | Zoom | Transits | Opti-fold
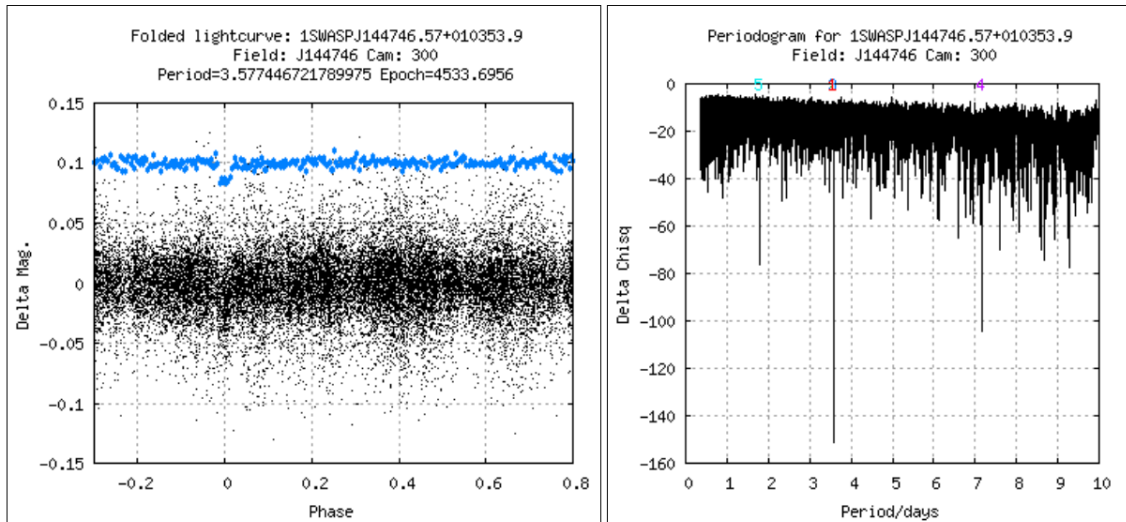
P1=**3.577** P2=**3.577** P3=**3.577** P4=**7.154** P5=**1.789**
harmonics on | lime line = selected period P
magenta lines = harmonics = {0.5, 1.5, 2, 3, 4} x P
Detrended Data | Full archive LC

i	Period	Epoch	Width (hr)	Depth (mag)	Del Chisq	S/N Red	N_tr	Ellip Var	Ellip S/N	Frac. in Tr.	SN Anti	R_{Pl} VK	R_{Pl} JH	sig det eff i	H_match	Clmp	All
	3.577446721789975	4533.6956	2.9354	-0.0143	264.5117	-12.8389	17	0.0004	1.8499	0.0185	6.2430	1.01	1.04	45.5463	70	0	

i	Period	Epoch	Width (hr)	Depth	Impact	R_{star}	M_{star}	R_{Pl}	Prob pl	Prob MS	Prob imp	Chisq_cs	Chisq_ucs	Q	All	
	3.5773986 ±	4809.15484 ±	3.0565	0.0121	0.377	1.059	1.015	1.132	0.759	0.616	0.905	22605.4	22605.2	0.24	Plot 1	Plot 2

Figure 2.1: Example of an entry in the online WASP archive showing the information available for WASP-37b.

The team of human vetters can access the lightcurves by camera field or by a query where thresholds on any combination of features can be set manually. For each star, the observer can either disposition the object as an eclipsing binary (EB), blended star (Blend), eclipsing low-mass companion (EBLM), variable star (V), reject the star for other reasons such as systematic errors (X), or leave the star unlabeled. If the lightcurve shows signs of having a planetary transit, it can be flagged with an AA, A, B, or C depending on the level of confidence and the practicality for follow-up. For example a star in a crowded field may be downgraded from A to B or C because follow-up observations would be more difficult. If the observer is uncertain, they can flag the object with a D to generate more discussion and trigger further processing of the data. It is important to note that all of these flags can be assigned without any additional follow-up information, and are therefore prone to errors in the dispositions.

Depending on telescope availability and target visibility, the targets flagged "A" or "B" are observed again either with photometry to improve the precision of the transit ephemeris, check for nearby stars contributing to blending, and check the color dependence of the transit depth, or with spectroscopy to check for binary stars and get a mass estimate on the secondary object.

In the past decade, the WASP team has dispositioned many thousands of objects, yet many thousands still remain unexplored. This scenario is a natural starting point for the application of machine learning with the goal of optimizing the use of follow-up facilities.

2.2 WASP False Positives

In order to optimize follow-up studies, it is important not just to correctly identify transit signals but to reduce the number of false positives. The most interesting objects in the training dataset, therefore, are the ones that "tricked" the human observer into performing follow-up observations. In many cases (such as EBLMs where the radius overlaps with the planetary regime), the only way to correctly disposition the object is through radial velocity observations. However, there may be subtle features that make it possible to better classify other types of objects and thereby improve the follow-up success rate. Because of this, we put great care in creating a list of convincing WASP false positives to be included in the training data.

With the successful launch of the Transiting Exoplanet Survey Satellite (TESS; Ricker et al., 2015a), it has become important for previous follow-up attempts to be made public. While the TESS mission has enlisted the help of a large number of observatories for follow-up work, TESS is producing candidates faster than they can be observed with the facilities available, necessitating candidate ranking and prioritization. Following in the footsteps of

Figure 2.2: Locations of the observatories used for follow-up of SuperWASP targets. The stars at the Roque de los Muchachos Observatory and South African Astronomical Observatory denote the location of SuperWASP and WASP-South respectively.

the KELT catalog of false positives (Collins et al., 2018), we have made the WASP catalog of false positives publicly available (Schanche et al., 2019a) so that follow-up efforts are not duplicated. The false positive labels can also be incorporated into further machine learning efforts by other surveys that observe the same fields.

To create the false positive catalog, we visually inspected all objects in the northern SuperWASP field for which follow-up observations were reported. The full list of observatories that are involved in validating WASP candidates can be seen in Figure 2.2, with their contribution to the effort shown in Figure 2.3. The majority (891) of catalog dispositions make use of spectroscopic data, while 315 have additional photometric data. Note that this includes all observations attempted. In many cases, events such as weather may have prevented useful observations for disposition purposes.

We then went through all 1,041 objects in the catalog to ensure that the correct disposition was given. This was especially important for objects dispositioned early in the WASP project, as labelling choices were not as consistent across observers, with a particularly high level of inconsistency between EB, EBLM, and Blend categories. In some cases, a star in the WASP database is labeled as rejected without further information (denoted as "RAF"), in which case we updated the flag to represent the reason for the rejection.

The general process used for the classifications is described in Fig. 2.4. In many cases, there are not sufficient observations to completely go through the decision tree to make final labels. In those circumstances, only branches containing known information are used. For example, in the case where only spectroscopic observations are available, we bypass
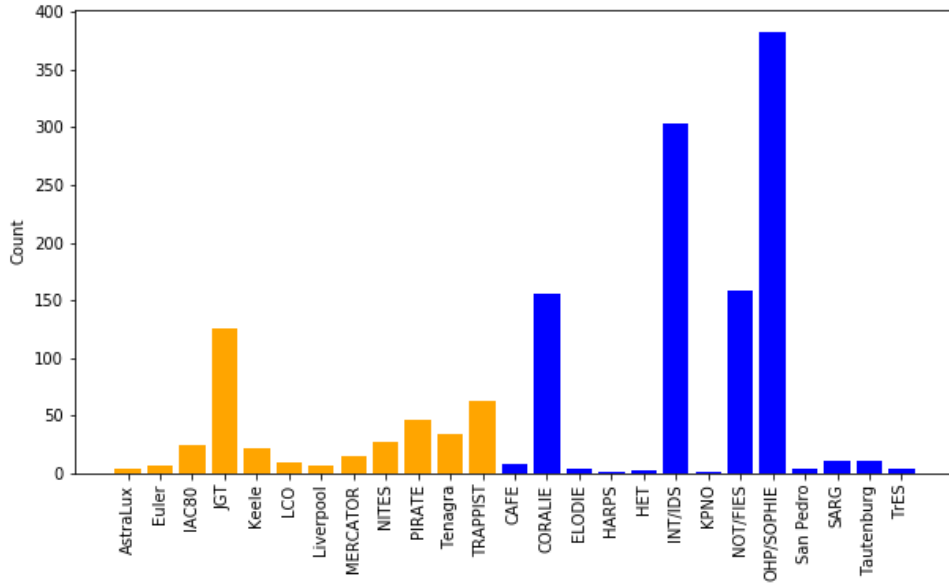
Figure 2.3: Contribution by individual observatories in follow-up observations for SuperWASP candidates. Orange bars represent photometric and blue bars represent spectroscopic observations.

the branch pertaining to aperture and photometric blends.

In addition to validating the primary disposition, we also standardized the application of follow-up flags. These flags can include information from spectroscopic follow up, such as single or double lined eclipsing binary (SB1 or SB2) or line-bisector variations (LB). The follow-up flag can also add further information about the star, such as whether the star is rotating rapidly (RR) as defined by a FWHM > 8 km s^{-1} or that the primary star in the system is a giant (G). In the event that the giant status was deduced prior to follow-up by looking at the reduced proper motion of the star, the star is not included in the catalog. However, in some cases, the giant status of the star was not known until spectral data were obtained, in which case we do include the star in the catalog. The recent Gaia data release has now greatly reduced the risk for spending follow-up effort on evolved stars, so it is unlikely that this category of false positive will pose a problem in future efforts. Finally the "O" flag stands for "other", meaning that none of the follow-up flags provided captures why the object was categorized as it was. When possible, further information about the classification is included in the final catalog.

The catalog highlights several interesting trends in the population of stars that have been observed. One such trend is the magnitude sensitivity of SuperWASP. Figure 2.5 shows that the vast majority of stars observed with WASP have V-band magnitudes less than 14, even though the input catalog contains all stars brighter than a magnitude of 15. The reason for this is twofold. First, brighter objects have better signal to noise, with
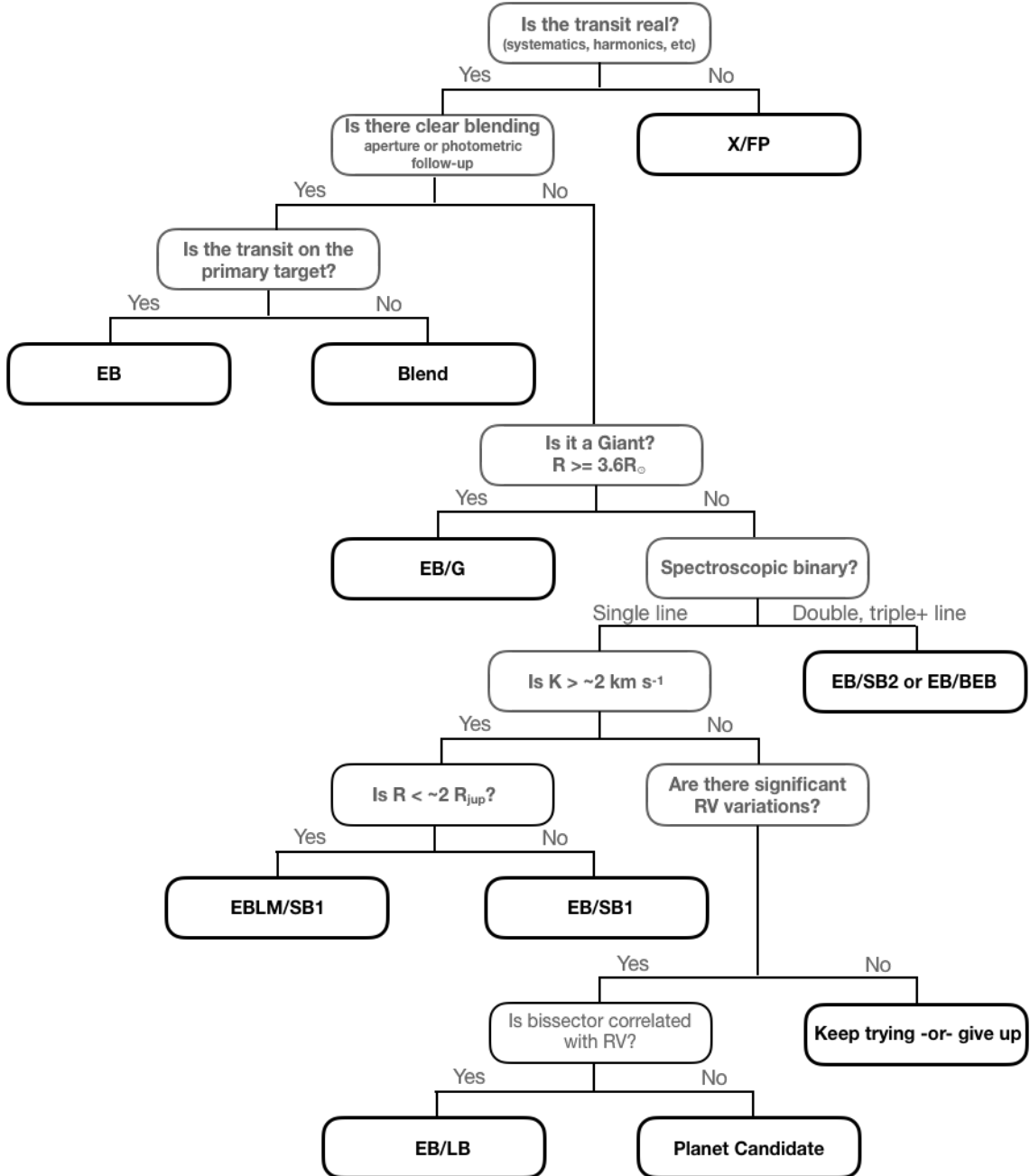
Figure 2.4: The decision tree describing the characteristics used to disposition objects after follow-up observations in the WASP FP catalog. Nodes describing decision criteria are in gray, while final classifications are shown in black.
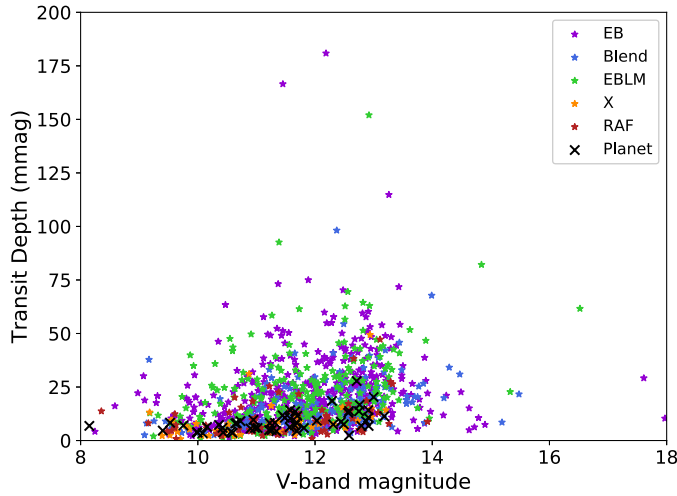
Figure 2.5: Comparison of the V-band magnitude and transit depth for all objects in the Super-WASP false positive catalog. Only planets found by either SuperWASP or a joint discovery with WASP-South are overplotted for context.

objects brighter than a V magnitude of 9.4 having precision around 0.004 magnitudes. By a V magnitude of 11.5, the precision drops to .01 mags (Pollacco et al., 2006). Secondly, follow-up observations, and especially sensitive spectroscopic observations, are magnitude limited, so brighter stars are treated preferentially for follow-up attempts.

Figure 2.6 shows the relationship between the transit period in days to the depth of the transit in magnitude for each type of false positive. It is clearly shown that the majority of planets detectable with WASP have transit depths less than 25mmag, and a depth greater than this threshold is strongly suggestive of an eclipsing binary or low mass eclipsing object. The gap of objects with periods of one day is the result of the day/night cycle on Earth. This cycle creates artifacts in the SuperWASP data making identifying real transits in the signature impossible, and so signals at this period are automatically filtered out.

In figure 2.7, we see that the majority of planets that have been detected orbit around stars with temperatures in the range of 5,000-6,500K. This observational bias reflects that fact that the sensitivity of WASP favors stars in the F and G range (Bentley, 2009). It is also notable that many stars with estimated secondary radii that are much larger than would be expected for a planet were targets for follow-up observations. This is due largely to the pixel size of SuperWASP, leading to blending of light from multiple stars in the aperture diluting the transit depth in the WASP data.

Like WASP, TESS has a large pixel size and therefore is likely to suffer from blending. We are therefore interested in exploring the impact that blending has on the false positive
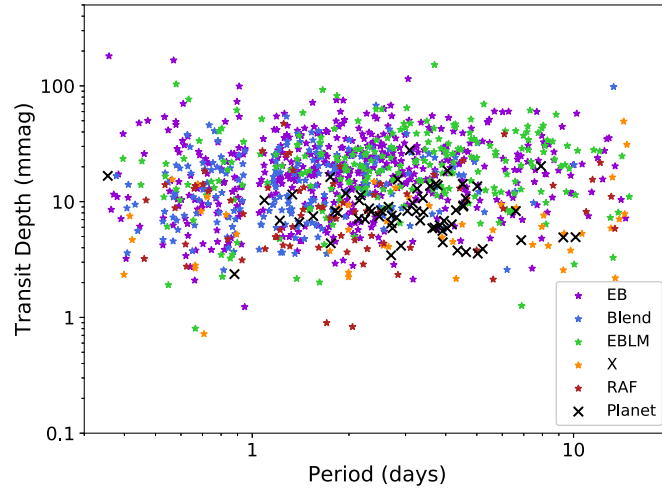
Figure 2.6: Period in days and depth in magnitude, both scaled logarithmically for clarity, for all objects in the SuperWASP false positive catalog. Planets found by either SuperWASP or a joint discovery with WASP-South are overplotted for context.
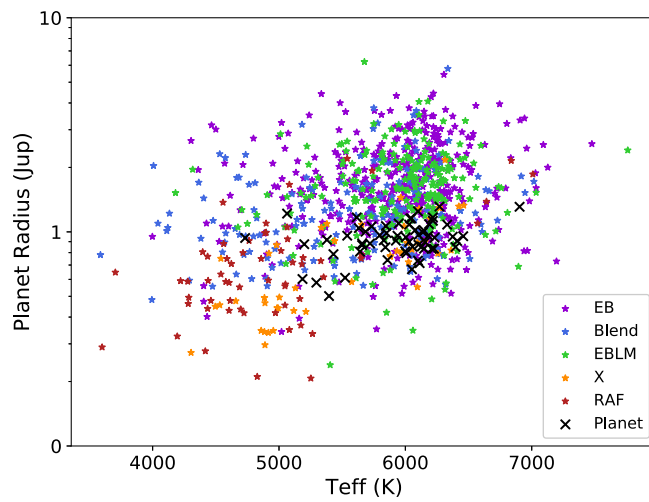
Figure 2.7: Comparison of the relationship between host star temperature and estimated secondary radius.
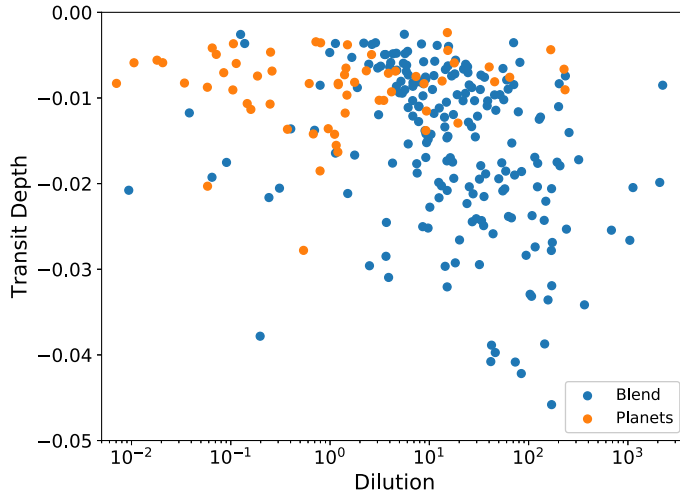
Figure 2.8: Comparison of the relationship between dilution in the 3.5 pixel aperture and transit depth for planets and blends.

rate to better inform decisions on the use of telescope time. In order to try to quantify the effect of blended light, we use *Gaia* photometry to make a rough prediction on the dilution of light from the primary target. We search the *Gaia* DR2 database (Gaia Collaboration et al., 2016, 2018) for all objects within the middle WASP aperture of 3.5 pixels ($\sim$48") in radius surrounding the target star. We then combine the relative magnitudes of the surrounding stars to the target star as an estimate of the total dilution. Higher dilution values reflect increased light from other sources within the aperture. Note that this dilution factor differs from that generally used by WASP which measures flux rather than magnitudes, but the two values are correlated. The large scale of values shown in Figure 2.8 are normalized before being used in the the machine learning dataset, compacting the range of values. There is a clear tendency for the successful detection of planets to occur when few other stars are present to dilute the target light. While this effect is expected, it does demonstrate the trade-off of completeness and success rate when observing in crowded fields. This effect is most apparent when comparing to the population of blends (as can be seen in figure 2.8), although a similar trend is seen when comparing planets to eclipsing binaries and other low mass eclipsing companions. This demonstrates another way which *Gaia* data can be used to help with candidate prioritization.

2.3 Machine Learning with WASP

While there are a wide range of machine learning methods that have been developed, the algorithms we selected were chosen to function in the way a human observer makes decisions on selected lightcurves. As noted above, all objects in the WASP catalog contain

information about the star and the potential secondary object from the BLS algorithm (referred to as the features) as well as class labels for each object. We used this information as input to a Random Forest Classifier to make predictions on the nature of the observed transit, analagous to the way a human would look at the system properties and decide whether they could reasonably represent a planetary system. In addition to the descriptive feature data, a human observer also has access to the lightcurve data itself. This visual data can be interpreted with Convolutional Neural Networks. We combine the results of both methods in order to identify and rank new WASP candidates.

The training and test datasets were composed of all items in the WASP archive labelled as eclipsing binaries or blended stars, planets, variable stars, or otherwise rejected as transits as defined above. This includes all objects in the false positive catalog as well as objects in the southern sky and those labelled without additional follow-up. We trained the algorithms as a multi-class problem with 4 output categories: EB/Blend, P, V, and X. Blends and binaries are included as one class because blends are in fact binary systems that have been diluted by light from another star. Therefore we would expect the shape of the lightcurves to be similar, albeit with a shallower transit signal. EBLMs were not included in training at this stage, although we do test the final algorithm's performance on these objects.

2.3.1 Random Forest Classification

Random Forest Classifiers (RFCs; Breiman, 2001) are made up of several distinct decision trees working together as an ensemble. Before discussing the RFC as a whole, first it is important to understand the constituent decision trees.

In essence, a decision tree is a classifier that takes as input a training dataset containing numerical descriptors (features) corresponding to assigned labels. The decision tree attempts to sort the input data by passing the features through a series of splits, or "branches". At the end of the branches is the "leaf", where a classification is assigned.

The criteria for how the splits are made can vary, but most commonly the Gini impurity score is used, and this is what we implemented in this work. Formally, the Gini impurity can be written as

$$G = \sum_{i=1}^{m} f_k(1 - f_k) \tag{2.6}$$

Where m is the number of classes in the dataset and f_k is the fractional number of samples that belongs in class k. An impurity measure of 0 would imply a perfect split,

with all samples in the node belonging to the same class. The algorithm iterates over each feature and and finds the best split corresponding to the lowest weighted Gini impurity, a combination of Gini scores for both sides of a split, for that feature. The overall lowest impurity is the final split at that branch. This process is repeated for each branch in the tree. The goal is to create "pure" nodes where the final leaf contains items of a single type.

There are many advantages to a decision tree. For example, because the branch split is made based on a single feature, there is no issue with having data of different types (categorical features mixed with continuous features). Another advantage related to the single-feature splits is that decision trees are very good at ignoring irrelevant features as they will not produce the lowest impurity. This is great news in data exploration where the best or most useful features are unknown. In addition, the calculations are very straightforward and simple to compute, so trees are able to handle a large number of features, run quickly, and create trees with decision splits that are easy to trace and interpret.

All of these advantages makes it sound like decision trees are a great final classifier. Unfortunately, decision trees fail at one very important point: their predictive power. The splits are made to optimize the training data, but they are highly sensitive to the exact data points used in the training process. This means that they suffer from high variance, also known as overfitting. That is to say, single decision trees do not generalize well to new datasets, and therefore do not make great classifiers on their own.

RFCs were developed to harness the great advantages of decision trees while eliminating the overfitting problem. They do this by combining the results from many individual decision trees to make a "best vote" prediction. The key to this method is to make sure that each tree learns something different about the data. This is accomplished through a "bagging" (**b**ootstrap **agg**regat**ing**) method in which the training data are broken into a specified number of sub-training sets. Each sub-training set is bootstrapped, meaning that it draws random values from the training dataset, with replacement. A separate decision tree is then trained on each boostrapped sample. The final prediction is the aggregation of the predictions from all of the individual trees.

In order to further reduce the variance, an additional layer of randomness is added to the trees of the RFC. Rather than giving each decision tree in the forest access to the all of the features to calculate the best split at a branch, a tree in the random forest is only given a random sub-sample of the full feature set. In implementation, this is usually $\sqrt{k}$ where k is the total number of features. At each split in each tree, a different sub-sample of features is chosen.

In recent years, RFCs have begun to be used widely in the astronomy community for predicting solar flare eruptions (Liu et al., 2017; Florios et al., 2018; Domijan et al., 2019) and filament eruptions (Barnes et al., 2017; Aggarwal et al., 2018), classifying quasars (Carrasco et al., 2015), variable stars (D'Isanto et al., 2016; Dubath et al., 2011; Masci et al., 2014), X-ray variables (Lo et al., 2014), and supernovae (Revsbech et al., 2018). A new variant, Probabilistic Random Forests, has even been developed to handle the informative errors in astronomical measurements (Reis et al., 2019).

In the field of exoplanets, RFCs have been used to identify planetary candidates primarily for space-based missions such as Kepler (McCauliff et al., 2015; Caceres et al., 2019b; Mislis et al., 2016). RFCs can also be used for regression, rather than classification. This technique has been used by Ulmer-Moll et al. (2019) to model the mass-radius relationship of exoplanets, and by Márquez-Neila et al. (2018) and Fisher et al. (2020) to perform atmospheric retrievals.

Training Dataset for RFC

An essential part of any type of machine learning is creating a dataset with useful predictive features. Several different approaches have been taken to do this in terms of exoplanet transit detection. McCauliff et al. (2015) used a large set (>200) features that are derived from wavelet analysis, transit model fitting, centroid motion, and other tests of the lightcurve. Mislis et al. (2016) take a simpler approach, using only 4 features describing the distribution of datapoints in the lightcurve of simulated transits in a Kepler-like sample. Caceres et al. (2019a) combine physical descriptions of the system, such as stellar mass, radius, and temperature, properties of the lightcurve, and results from their autogressive technique to train the algorithm. Our approach is similar to that of Caceres et al. (2019a), but including information on the candidate system derived from the BLS method. A full list of our training features can be found in Table 2.1.

An initial transit width, depth, period, epoch of mid-transit, and radius are estimated from the BLS. Stellar features such as the mass, radius, and effective temperature are found by the method described by Collier Cameron et al. (2007), in which the effective temperature is estimated from a linear fit to the 2MASS $J - H$ color index. The stellar radius is then calculated from a polynomial fit to the temperature/radius relation for main-sequence stars tabulated in appendix B1 of Gray (1992), with the mass following from a power-law approximation to the main-sequence mass-radius relation, $M_* \propto R_*^{5/4}$. A more rigorous fit to the transit profile yields the impact parameter and the ratio of the stellar radius to the orbital separation, and hence an estimate of the stellar density. Markov-chain Monte Carlo (MCMC) runs are performed to sample the posterior probability distributions

Table 2.1: Features used by the RFC. Starred features are those added to the dataset, while the rest were taken directly from the WASP database. The efficacy of many of these measures for false-positive identification is discussed in detail by Collier Cameron et al. (2006)

Feature Name	Description
clump_idx	Measure of the number of objects in the same field with similar period and epoch
dchi_P*	The $\Delta\chi^2$ value at the best-fit period from the BLS method.
dchi_P_vs_med*	The ratio of $\Delta\chi^2$ at the best-fit period to median value.
dchisq_mr	Measure of the change in the χ^2 when MCMC algorithm imposes a main-sequence (MS) prior for mass and radius.
delta_Gaia*	stellar radius from MCMC - Gaia dr2 radius divided by Gaia dr2 radius
delta_m*	The difference between the mass calculated by J-H and the MCMC mass.
delta_r*	The difference between the radius calculated by J-H and the MCMC mass.
depth	The depth of the predicted transit from Hunter.
depth_to_width*	Ratio of the Hunter depth and width measures.
epoch	Epoch of the predicted transit from Hunter (HJD-2450000.0)
impact_par	impact parameter estimated from MCMC algorithm.
jmag-hmag	Color index, J magnitude - H magnitude.
kurtosis*	Measure of the shape of the dip for in-transit data points.
mstar_jh	Mass of the star, from the J-H radius*(1/0.8).
mstar_mcmc	Stellar mass determined from MCMC analysis.
near_int*	Measure of nearness to integer day periods, abs(mod(P+0.5,1.0)-0.5).
npts_good	Number of good points in the given lightcurve.
npts_intrans	Number of datapoints that occur inside the transit.
ntrans	Number of observed transits.
period	Detected period by Hunter? in seconds.
rm_ratio*	Ratio of the MCMC derived stellar radius to mass.
rplanet_mcmc	Radius of the planet, from MCMC analysis.
rpmj	Reduced proper motion in the J-band (RPMJ=Jmag+5*log10(mu)).
rpmj_diff	Distance from DWs curve separating giants from dwarfs.
rstar_jh	Radius of the star derived from the J-H color measure.
rstar_mcmc	Radius of the star determined from MCMC analysis
sde	Signal Detection Efficiency from the BLS.
skewness*	Measure of the asymmetry of the flux distribution of data points in transit.
sn_ellipse	Signal to noise of the ellipsoidal variation.
sn_red	Signal to red noise.
teff_jh	Stellar effective temperature, from J-H color measure.
trans_ratio	Measure of the quality of data points (data points in transit/total good points)/transit width.
vmag	Cataloged V magnitude.
width	Width of the determined transit in hours.

of the stellar and planetary radii and orbital inclination. The MCMC scheme uses optional Bayesian priors to impose a main-sequence mass and radius appropriate to the stellar effective temperature. Note that the results and predictions would change if the precise radius were used instead, particularly if the star has evolved off of the main sequence.

In addition to the provided information, we add several new features to capture more abstract or relational information, such as the ratio of transit depth to width and the skewness of the distribution of the magnitudes found within the transit event. The latter is a possible discriminator between 'U' shaped central transits of a small planet across a much larger star, and shallow 'V' shaped eclipses of grazing stellar binaries. The new high precision distance calculations released by *Gaia* DR2 are used to measure the deviation of the estimated main sequence radius calculated as above and the measured radius. In total, 34 features are included in the dataset.

Before training, the full dataset containing the star name, descriptive features, and disposition is split randomly into a training dataset and a test dataset. In total there are 4,697 training cases and 2,314 testing samples. Prior to running the classifiers, all of the features of the training dataset are median centered and scaled in order to reduce the dynamic range of individual features and to improve performance of the classifier. The scaling parameters are retained so that they can be applied to subsequent datasets to which the classification is applied, including the testing dataset.

RFC customization

RFCs are straightforward models with a small number of possible tunable parameters. The main features that can be adjusted are the number of random features introduced at each branch to calculate the split, the number of splits that a given tree can make, and the total number of trees in the forest. The customization of the RFC we implement is described below. The evaluation was performed by a 10-fold cross validation, where the training data were randomly split into 10 groups with 10 models trained, each withholding and testing on a different fold of the data.

By default, at each branch $\sqrt{k}$ features are available to find the best split. We tested this value by testing a range from 1 to 17 (half of the total features), and found that performance did improve up to $\sqrt{34} \approx 6$, and adding more features did not aid performance (See Fig. 2.9).

Trees can be grown until the final leaves all contain a single class, or the growth of the tree can be truncated. The deeper the tree, the more information from the training data goes into the decisions. However, the trade-off is in overfitting. We tested several depths (see Fig. 2.10) and found that the deeper the tree the better the performance. However
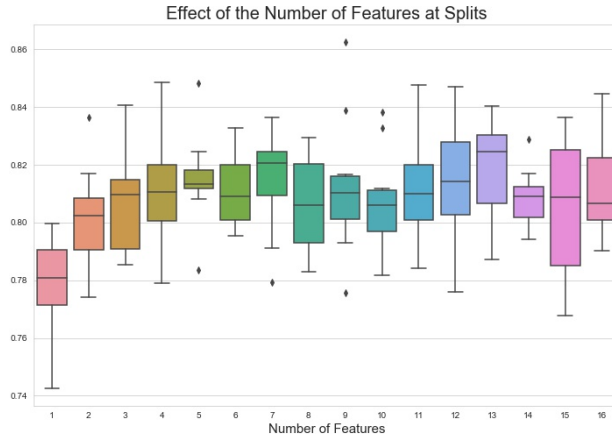
Figure 2.9: Box-and-whisker plot showing the effect of the number of random features available for each split in the decision trees making up the Random Forest. The y-axis in this and following box-and-whisker plots shows the f1-score, described in more detail in Section 3.3. The parameters for tree depth and number of trees is set at the default value.

we found a depth of 6 was sufficient to reduce overfitting.

The final tuning parameter for the random forest is the number of trees. Initially increasing the number of trees improves performance rapidly, but soon the benefit for each additional tree tapers off (see Fig. 2.11). When the performance plateaus, adding additional trees increases computation time without adding information content or improving predictions. We therefore tested forests from 1 to 500 trees in steps of 20 and found that performance did not improve beyond 200 trees.

Because the decision trees test the importance of available features at each split of the tree, the RFC can analyze these values for feature ranking. The results of such an analysis using our training dataset are shown in Fig. 2.12. This can be used to gain insight into the decision making process that the classifier has developed, which can inform further analysis. For example, the detected period was the strongest indicator. This can be explained in large part because false planet detections arising from diurnal systematics tend to have orbital periods close to multiples of one sidereal day due to the day/night cycle present in Earth-based observations. This indicator would likely not play as significant a role in a space-based survey unaffected by the day/night cycle. The width or duration of the transit, estimated radius of the planet, the $\Delta\chi^2$ value (a product of the BLS search) of the object at the best-fit period, and the number of transits of the object round out the top 5 features in prediction.

The features that had relatively little impact on the overall prediction related largely to stellar properties, including the magnitude and radius of the star. This shows that there is no strong preference for a certain size star to host a particular type of object in our
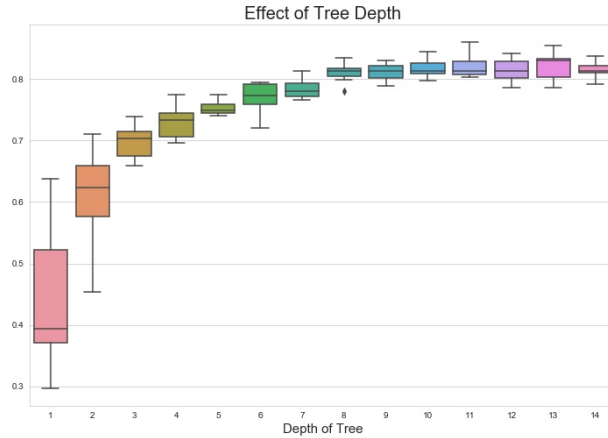
Figure 2.10: Box-and-whisker plot showing the effect of the maximum depth (number of splits) of the decision trees making up the Random Forest. The number of features to split on and the number of trees are set at the default value.
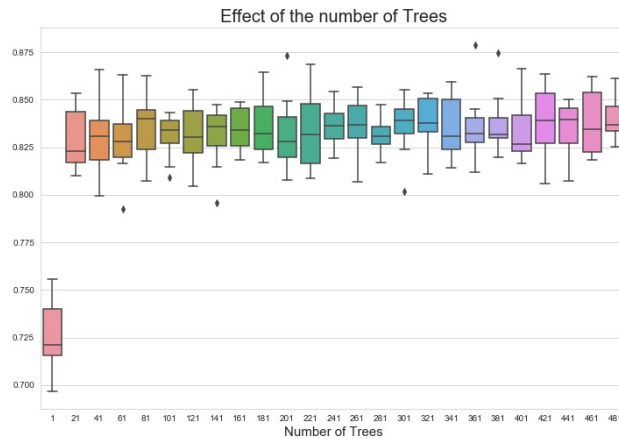
Figure 2.11: Box-and-whisker plot showing the effect of the number of individual decision trees that make up the forest. After an initial large gain in performance, the improvement quickly flattens. We therefore set the final number of trees to 250, above which the time of computation outweighs the gains in performance. The number of features available for each split and the maximum depth are set to the default values.

31
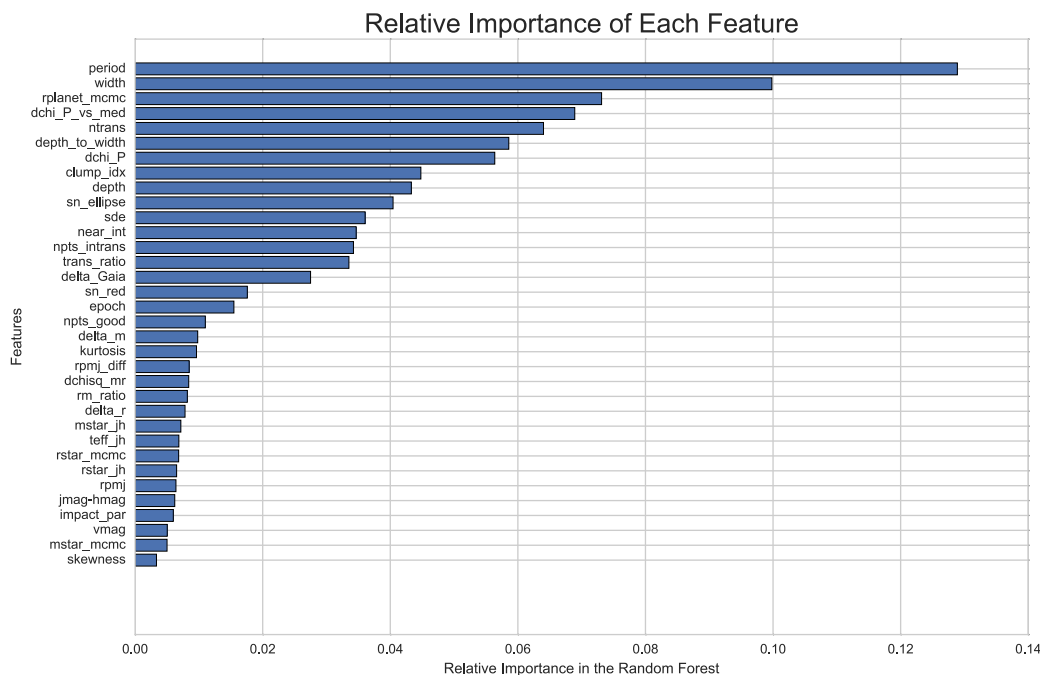
Figure 2.12: Ranked list of the effectiveness of each of the features in making correct classifications of the training dataset for the Random Forest Classifier.

sample, as the apparent magnitude range to which WASP is most sensitive is dominated by F and G stars (Bentley, 2009). The lowest ranked feature is the skewness, showing that the asymmetry of datapoints falling within the best-fit transit is not sufficiently capturing the transit shape information.

2.3.2 Convolutional Neural Networks

The RFC takes into account only numerical descriptors of the dataset. However, an important consideration for a WASP observer looking at the online database is the two images provided (See Figure 2.1). These are, first, an image of the lightcurve folded on the best-fit period, and second a plot showing the $\Delta\chi^2$ periodogram from the BLS run. The observer can often rule out candidates based on these two plots alone. While the features used to train the RFC attempt to capture the information contained in the plots, they are not complete. It is therefore desirable to learn characteristics of the different classes from the lightcurve itself. In order to do this, we implement a Convolutional Neural Network (CNN).

Convolutional neural networks got their start with the concept of the neocognitron (Fukushima, 1980), although this early method was missing the distinctive backpropagation element (described below) that is widely used today. The first time a CNN using
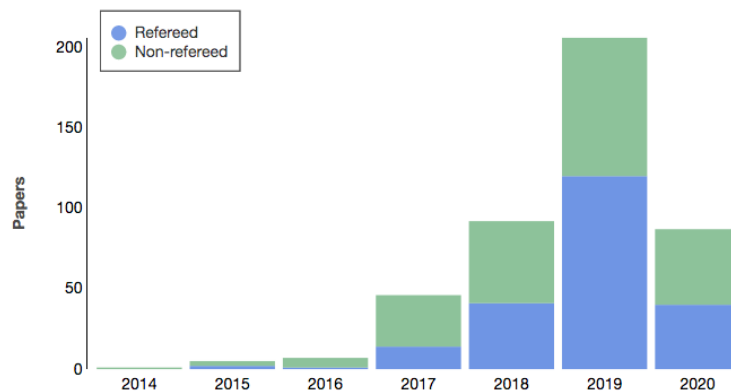
Figure 2.13: Number of papers mentioning Convolutional Neural Networks in the SAO/NASA Astrophysics Data System by year, as of May, 2020.

backpropagation appeared in literature was in LeCun et al. (1990) for the application of recognizing numbers written by hand. However, it wasn't until 2012 that CNNs made their big break into wide use when a CNN drastically outperformed previous attempts at image recognition in the ImageNet Large-Scale Visual Recognition Challenge (Krizhevsky et al., 2012), an annual computer-vision competition. The spread of CNNs has also been aided by the use of GPUs, which drastically speed up the training time.

In the field of astronomy, CNNs have quickly gained popularity and have been employed in a wide range of uses. In an application close to the CNN roots of handwritten number identification, Zheng et al. (2016) used CNNs to interpret handwritten notations on historical sunspot drawings from Yunnan Observatory. CNNs have been used in the identification of objects from large surveys, such as separating stars from galaxies in Sloan Digital Sky Survey and the Canada-France-Hawaii Telescope Lensing Survey (Kim & Brunner, 2017), identifying pulsars in survey data (Zhu et al., 2014), finding and mapping craters on the lunar surface (Silburt et al., 2019), and identifying transients in difference images from the High cadence Transient Survey (Cabrera-Vives et al., 2017). In the competition setting, CNNs were the most effective model for identifying strong lensing sources in simulated ground and space survey data (Schaefer et al., 2018) and in predicting ratings for the Galaxy Zoo's Galaxy Challenge (Dieleman et al., 2015). Recently, CNNs have been used in the detection of meteors by several groups to high (>99%) accuracy (Gural, 2019; Cecil & Campbell-Brown, 2020).

In the search for planet transits, Pearson et al. (2018) used simulated Kepler-like data to identify planet transits and found that a 1-dimensional CNN outperformed other methods of detection including a BLS approach. Similarly, Zucker & Giryes (2018) used simulated Kepler-like data with an improved red noise contribution. Chintarungruangchai & Jiang (2019) devised a way to find Kepler transits that does not rely on knowledge of

the period by constructing a 2D map of the data folded on a range of arbitrary periods, with the resulting map used as input to the CNN.

Shallue & Vanderburg (2018) set the standard in transit detection with Kepler data using a CNN taking both full phase-folded lightcurves (termed the "global view") and zoomed-in views of the transit event ("local view"). They then applied the network to Kepler lightcurves of known multi-planet systems and were able to find additional planets in two systems. Ansdell et al. (2018) expanded this method and improved the accuracy and precision of Shallue & Vanderburg (2018) by adding additional time series data representing the centroid shift of the lightcurve and domain knowledge about the star into the architecture of the CNN.

This combination of lightcurve, centroid, and domain data was used by Chaushev et al. (2019) for application to the ground-based NGTS to aid with candidate vetting. This study tested the use of real and simulated data and found that the fully-simulated data did not adequately replicate the real data when it came to prediction on real data. This highlights the importance of careful construction of the training data. A further insight by the authors was that the CNN was surprisingly robust to incorrectly labeled objects. In fact, performance remained relatively high up to a nearly 50% misclassification rate in the simulated training data. This is good news for the WASP survey, as the dispositions contain errors and biases.

This network structure was also used by Osborn et al. (2020) for the application on TESS data, although the simulated data were not as reliable when using real data for training, reinforcing the need for accurate systematic information for the given survey when generating training and testing sets. Yu et al. (2019) take a different two-step approach to finding TESS candidates, first using the local and global view of phase-folded data to do an initial search for transit candidates, then adding the local view of the most likely secondary eclipse and a measure of the blend likelihood to perform a more in depth vetting of candidates. Both methods will help in the identification of new planet candidates for the TESS survey.

Convolution

A CNN can be thought of as having two stages. The first stage is the Convolutional step, which aims to process the image in a way that picks up spatially associated traits in the input data that are distinct to their class through the combination of convolutional layers and pooling layers. The second is a fully connected neural network.

The convolutional layers as the name suggests convolve the input of the layer with a filter (also called a kernel). The input to the nth layer is a stack of K matrices $\boldsymbol{x_{n-1}}^{(k)}$

where $k = 1,...,$K. The output of the layer, then, is a stack of L layers, known as feature maps, written as $\boldsymbol{x_n}^{(l)}$ where $l = 1,...,$L. In order to produce the output, each value in the input layer is convolved with the filters, added to a bias value for the feature map l, and then passed through a non-linear function:

$$\boldsymbol{x_n}^{(l)} = f \left(\sum_{k=1}^{K} \boldsymbol{W_n}^{(k,l)} * \boldsymbol{x_{n-1}}^{(k)} + \boldsymbol{b_n}^{(l)} \right) \tag{2.7}$$

where the filters are $\boldsymbol{W_n}^{(k,l)}$, the bias is $\boldsymbol{b_n}^{(l)}$, $*$ is the discrete cross-correlation operation, or more commonly called the convolution, and f is the activation function. The values of the filter and the bias terms are initially randomly assigned, but are adjusted during the learning process.

The purpose of the convolutional layer is to identify and highlight local features in the data. The role of the pooling layer, then, is to merge similar features together making the algorithm less sensitive to the position of a feature while also reducing the size of the data. This is accomplished by taking successive subsamples of the data and performing some function to create a single data point to represent that subsample. Most often, as is the case for our implementation, that function is simply taking the maximum value in the subsample. However other choices, such as taking the average value of each subsample, are possible.

The convolutional and pooling layers are applied alternatingly in several steps, leading to successively more abstract representations of the original data. Following the final pooling step, the processed data are flattened into a single 1-dimensional array. This array is used as the input to the fully-connected neural network.

Fully Connected Neural Networks

A neural network is a type of machine learning algorithm that connects the input data to the final output classification or regression through a series of interconnected layers (See Fig. 2.14). The left-most row in this example represents the input layer, followed by 2 hidden layers, and finally an output layer. The layers are comprised of units called *neurons* that each contain a numerical value. The values of every neuron in a given layer are connected to each neuron in the following layer through a weighted transformation, with the exact parameters of that transformation determined during the training process. Each neuron in the output layer contains a value representing the likelihood that the example falls in that class.

Formally, the output of layer n, x_n is given by
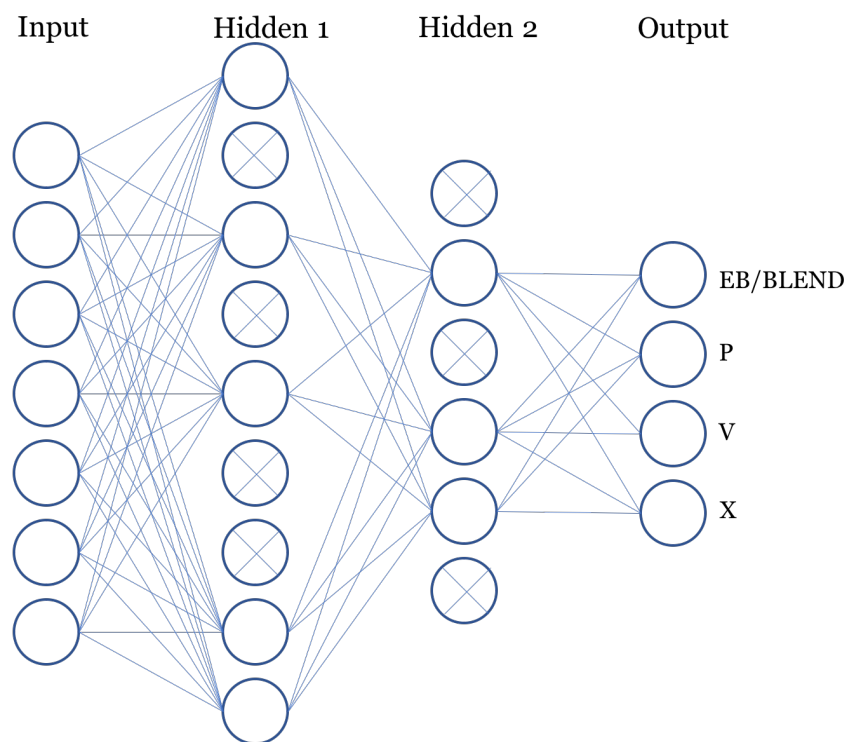
Figure 2.14: Visual representation of a neural network scheme, where circles represent individual neurons. In this example, the layers progress from left to right. Circles with crosses through them represent dropped neurons, described further in text.
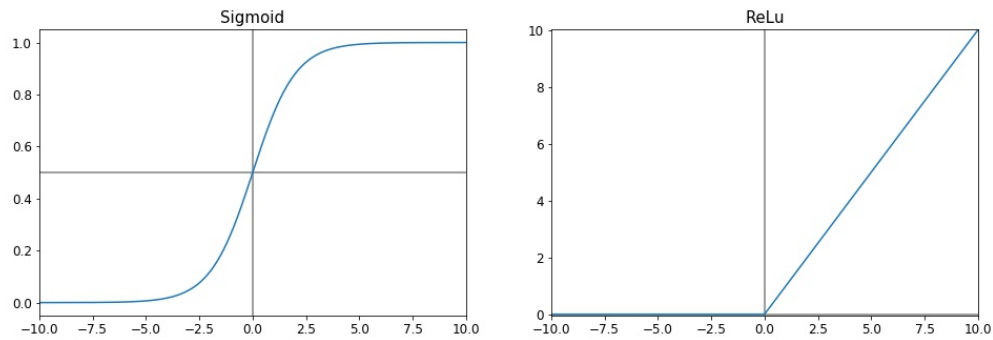
Figure 2.15: Graphs showing the form of the sigmoid (left) and ReLu (right) activation functions.

$$x_n = f(\boldsymbol{W_n} \cdot \boldsymbol{x_{n-1}} + b_n) \tag{2.8}$$

where $\boldsymbol{W_n}$ is the matrix of weights, b_n is the bias, and f represents the activation function. The values for $\boldsymbol{W_n}$ and b_n are initialized randomly and adjusted during training, whereas the activation function is specified by the user.

The format of this transformation is very similar to that of the convolutional layer, but the weights are applied to each individual neuron rather than a block of data. Using the KERAS package[1], we tested a random normal, a truncated normal distribution within limits as specified by LeCun (LeCun et al., 1998), and a uniform distribution within limits specified by He (He et al., 2015) initialization of the weights and found the greatest performance with He uniform variance scaling initializer.

The activation function serves as a filter for the network to determine what information is useful and what can be ignored, making it possible to learn complex relationships in the data. For all layers except the final layer leading to the prediction, we used the Rectified Linear Unit (ReLU) activation function (Nair & Hinton, 2010), which has been shown to speed up the training process and negate the need for any type of pre-training in the network (Glorot et al., 2010). This simple function takes the form f(x) = max(0,x), meaning that if the value in the neuron is less than 0 it will be ignored. For the final layer, we instead implement a sigmoid function, $f(x) = \dfrac{1}{1 + e^{-z}}$. The reason for this is discussed in more detail in Section 3.1.3. A graphical representation of the two functions is shown in Figure 2.15.

The initial performance of the network is naturally poor, given that the weights and biases are randomly assigned. In order to improve the network, the weights and biases are adjusted by minimizing a cost function, which is a measure of how close the prediction is to the desired result. As there are multiple output classes of data (EB/Blend, P, V, and

---

[1] https://keras.io/

X), we use categorical cross-entropy as the loss function. The categorical cross-entropy is written as

$$L(y, \hat{y}) = -\sum_{n=0}^{N} \sum_{k=0}^{K} y_{nk} \log \hat{y}_{nk} \qquad (2.9)$$

where K is the number of output classes, N is the number of training examples, $\hat{y}$ is the prediction of the network, and y is the one-hot encoded label. One hot encoding is when the output label is represented as an an array of length K where the array index position representing the category is given a value of 1 and all other positions are given a value of 0. For example, for a target that is a planet with a label of 1, the one-hot encoded label would be [0,1,0,0] while a variable star with a label of 2 would be [0,0,1,0].

Once the loss function is computed, the information is propagated backwards through the network to change the weights and reduce error during the next pass through the network. This is an iterative process, with small adjustments made after each pass through the network until a minimum error (or maximum number of iterations) is reached. We implement the Adamax optimizer to control the adjustments of the parameters (Kingma & Ba, 2014). The maximum change in the weights allowed by Adamax at each iteration is controlled by the learning rate, which can be tuned to different values for individual datasets, with a learning rate of 0.001 implemented after testing a range of values.

Calculating the gradient that determines the adjustment is memory intensive and slow, so rather than calculating the gradient on the full training dataset, we implement batch learning. This means that the gradient is calculated for a limited number of samples, the number specified by the batch size. After a forward and backward pass of the first batch, the weights are adjusted and the next batch is used. This repeats until all of the samples are used. The forward and backward pass of all batches constitutes one epoch. We use a 3-fold cross validation and perform a grid search over several batch sizes and epochs to test the performance and find that using a batch size of 20 samples run over 225 epochs is a good balance of speed and performance.

Finally, we incorporate neuron dropouts during the training stage, where a fraction of the neuron outputs are set to 0 (Srivastava et al., 2014). The dropped neurons are randomly chosen at each pass through the network. By randomly removing neighboring neurons, the network is forced to not become dependent on the presence of the other neurons in the layer. This has been shown to be effective in preventing overfitting. The total dropout percent is determined experimentally, and here we find 40% to be effective. A schematic of the final CNN architecture can be seen in Fig. 2.16.
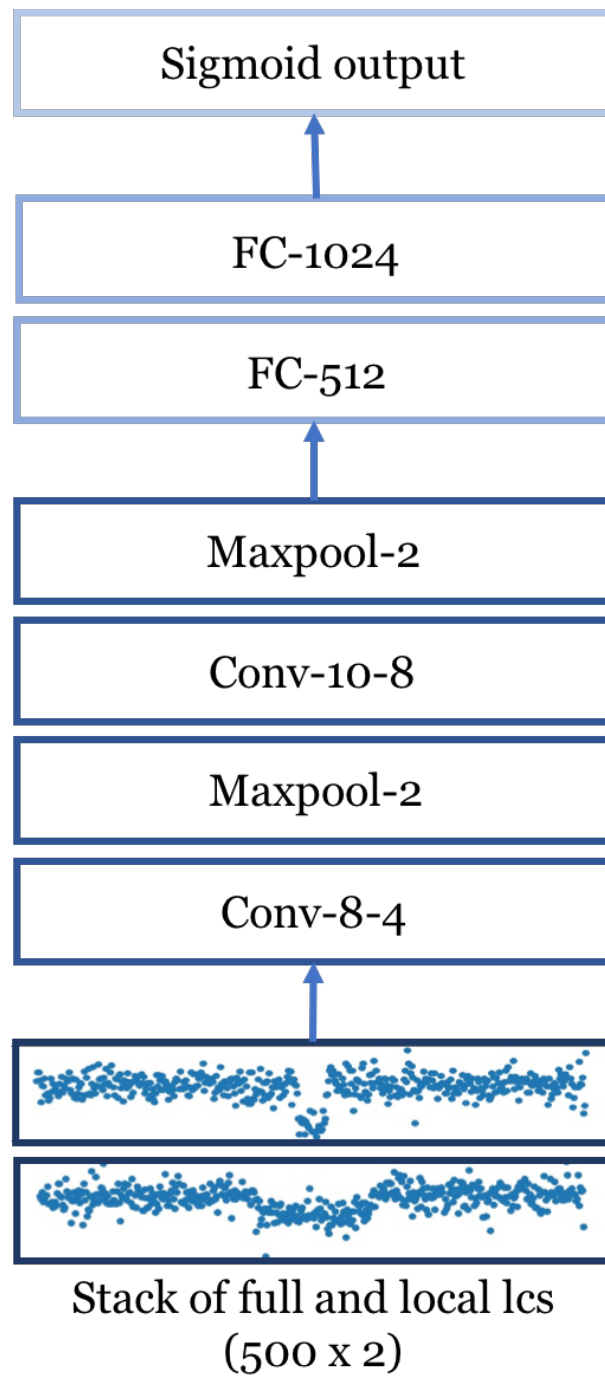
Figure 2.16: Representation of the final CNN architecture. The network proceeds from bottom to top. The input is the binned lightcurve data. Convolutional layers are represented by Conv-<number of filters>-<kernel size>, and Max Pooling layers are indicated by Maxpool-<Pool size>. The fully connected layers are denoted by FC-<number of neurons>. Not shown are the dropouts, which are introduced at each Max Pooling layer and each fully connected layer.
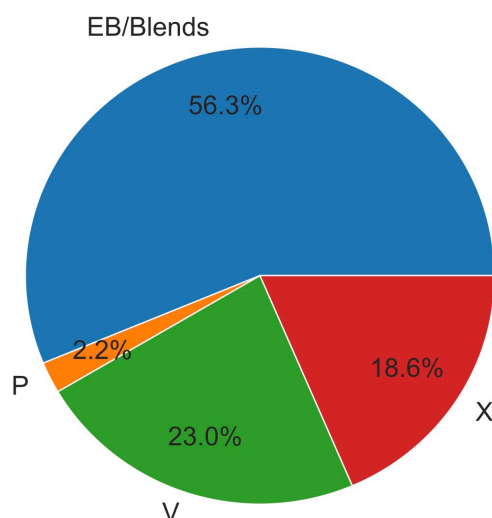
Figure 2.17: Relative number of samples in the WASP archive for each class used in training the RFC and CNN.

2.3.3 Unbalanced Classes

The WASP dataset used for training and testing is highly unbalanced, as the desired target (planets) are much less common than the other categories (Fig. 2.17). In machine learning, this presents a challenge, as classifiers that always predict the majority class do perform well by many metrics, but do not actually help with identifying the class of interest. To try to compensate for this for the RFC, we added additional datapoints using the Synthetic Minority Over-sampling Technique (SMOTE; Chawla et al., 2002). This technique creates synthetic datapoints for the minority classes that lie between existing datapoints with some added random variation. The synthetic data are added only to the training data, and the test dataset remains the same. We add SMOTE samples of P, V, and X categories to make the total sample for each class even. The addition of SMOTE data marginally increased the number of planets retrieved from the test data from 44 to 45, but also increased the number of non-planets given a planet classification, increasing false positives from 125 to 137.

For the CNN, we took a different approach by injecting artificial planets to the training data. We added synthetic transits to existing WASP light curves that showed no transit signal or other variability to capture realistic sampling and typical patterns of correlated and uncorrelated noise. To select appropriate lightcurves, we began with a sample of

objects classified as "X" in the Hunter catalog and measured the RMS against the V-magnitude. Those objects that fell more than 1σ below the best fit to the data were selected, as they show the least amount of variation. This left a total of 848 light curves.

The planetary signal added to the WASP data was created using the BATMAN package for Python (Kreidberg, 2015). The stellar mass, radius, and effective temperature were set using the known values for the star itself, while the planetary properties were generated randomly with the following distributions.

The period was randomly selected to be a value uniformly located in log space between 0.5 and 12 days. The mass of the planet follows the same lognormal distribution used in Collier Cameron & Jardine (2018), with a mean of 0.046 and a sigma value of .315. Following Kepler's third law, the semi-major axis a can be found for the period P, stellar mass M_*, and planet mass M_p using

$$a = \left(\frac{P^2 G (M_* + M_p)}{4\pi^2} \right)^{\frac{1}{3}} \qquad (2.10)$$

The radius of the planet is dependent on both the mass of the planet and the equilibrium temperature. We use a cubic polynomial in log mass and a linear term in log effective temperature to approximate the planetary radius, using coefficients derived from a fit to the sample of hot Jupiters studied by Collier Cameron & Jardine (2018):

$$\begin{aligned} \log \left(\frac{R_p}{R_{\text{Jup}}} \right) &= c_0 + c_1 \log \left(\frac{M_p}{0.94 M_{\text{Jup}}} \right) + c_2 \log \left(\frac{M_p}{0.94 M_{\text{Jup}}} \right)^2 \\ &\quad + c_3 \log \left(\frac{M_p}{0.94 M_{\text{Jup}}} \right)^3 + c_4 \log \left(\frac{T_{\text{eql}}}{1471 \text{ K}} \right)^4 , \end{aligned} \qquad (2.11)$$

where $c_0 = 0.1195$, $C_1 = -0.0577$, $c_2 = -0.1954$, $c_3 = 0.1188$, $c_4 = 0.5223$, and $T_{eql} = T_{eff} \left(\frac{R_S}{2a} \right)^{\frac{1}{2}}$

As we are looking only for close-in planets, we make the simplification that the orbit has been circularized and set all eccentricities to 0. Finally, the inclination was calculated by first randomly picking an impact parameter, b, between 0 and 1. The inclination was then calculated by

$$i = \cos^{-1} \left(\frac{b R_*}{a} \right) \qquad (2.12)$$

The transit model was generated and added to one of the selected WASP lightcurves
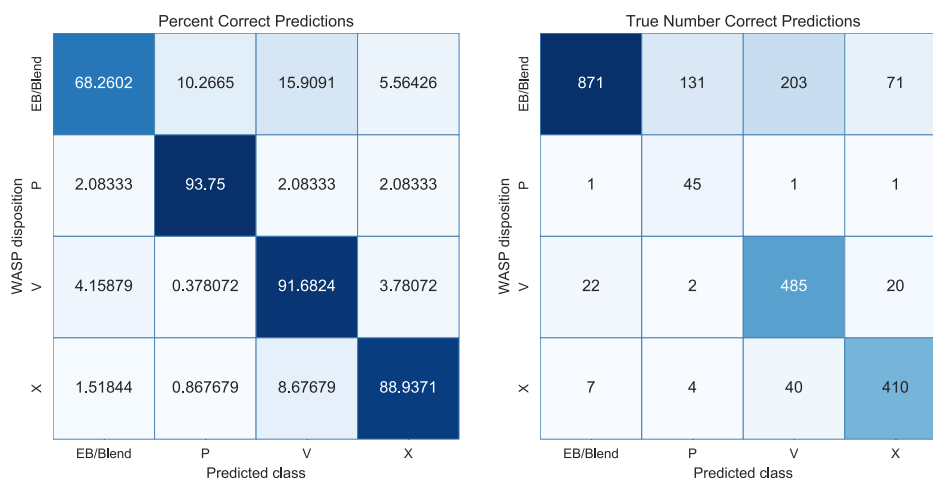
Figure 2.18: Confusion matrix showing the results of the RFC on the test set using the training dataset containing synthetic datapoints generated through SMOTE sampling.

after folding the WASP data on the assigned period. While some of these new planets were too small to be visible, and others were much larger than would be expected, we chose to include them all in order to push the boundaries of the parameter space that the CNN is sensitive to so as not to exclude potentially interesting, though unusual, objects.

2.4 Initial Results

The results of the RFC trained with the addition of SMOTE datapoints are shown in Fig. 2.18. The horizontal axis indicates what the algorithm predicts and the y-axis displays the human labeled class, which we assume to be accurate. Correct predictions fall on a diagonal line from upper left to lower right. The plot on the left shows the results as a fraction of lightcurves that fall into that bin. However, since the number of samples in each class varies, a more practical depiction is shown in the right plot, which shows the actual number of lightcurves for each category. ∼94% of confirmed planets are recovered from the dataset, while more than 10% of EB/Blends were misclassified as planets. Since there are far more binary systems recorded than there are planets, this quickly turns into a large number of lightcurves incorrectly flagged for interest, which translates to many hours of wasted follow-up time. For our testing dataset, there are 45 correct planet identifications and 137 that are false positives. This means that if all objects flagged as planets are followed up, we would expect about 25% of them to be planets.

To try to improve upon this false positive rate, we trained two CNN models. The first uses as input only the full lightcurve that has been phase folded and binned. The CNN achieves an overall accuracy (correct predictions divided by total lightcurves analyzed) of

around 82% when applied to the test dataset. While the fraction of correctly identified planets is lower than the RF (88% as opposed to 94%), the CNN performs much better in classifying eclipsing binaries and blends in terms of the percent of false positives with only ∼5% of EB/Blends being labeled as planets, as opposed to 10% for the RF. The CNN therefore has an overall better performance for follow-up efficiency.

The second CNN algorithm also includes the local transit information, using an approach similar to that of Shallue & Vanderburg (2018). The local information is comprised of the data centered on the transit and only containing the data 1.5 transit durations before and after the transit event, standardizing the transit width across events. The effect of this is to provide greater detail and emphasis on the shape of the transit event itself in order to understand the subtle shape difference between a typical planet and an eclipsing binary system. The full lightcurve and the local view are stacked and passed together into the CNN. In this case, the overall accuracy (83%) remained roughly the same, but the total percentage of planets found increased to (94%). The trade-off is a slight increase of the number of EB/Blends being labeled as planets. A full comparison of both methods can be seen in Fig. 2.19. Note that for both CNN networks, we use the artificially injected planets into both the training and test sets, so the numbers reported are for a combination of the real and artificial planets.

It is important to note that the way missing data were handled for both the full lightcurve and the local lightcurve made a large difference in the final performance. When binning the data, the full dataset was evenly split in 500 equal phase steps and all of the datapoints within those phase steps were averaged. In some cases, for example when the lightcurve was folded over an integer day, there were gaps in the phase ranges in which no data were present. Since the CNN can not handle missing data in the input string, a value needs to be inserted. We tried inserting either a nonsense value, in this case 0.1 which is far above any real datapoint, or repeating the last good value. In some cases there were several phase bins in a row that were missing data, causing a small section of the lightcurve to be flat.

After trying both options, we found that by far the best performance was obtained when inserting the nonsense value into the full lightcurve and repeating the last good value into the local lightcurve. This makes sense, as the full lightcurve gives a broader view of the star's lightcurve and is likely to have regular gaps in the data when it is folded on a bad period. The algorithm was able to identify that pattern and reject it. The local data, however, have fewer total datapoints because they cover a smaller total range of phases, and therefore are more likely to randomly have missing data. The algorithm appears to not be able to distinguish lightcurves missing data because of a systematic issue with the

Figure 2.19: Confusion matrices showing the results of the CNN using only lightcurves folded on the best-fit period (top) and with the addition of the local transit information (bottom) as input. The axes are interpreted the same as in Fig. 2.18.

data and those missing data simply because they lack enough observations during the transit.

The results and implications of both the RFC and CNNs will be described in more detain in Chapter 3.

2.4.1 Method limitations

There are several caveats to our study. One note of caution is the underlying training dataset. The training data were obtained by combining the entries of a number of WASP team members over the course of many years. This leads to two main problems. First, different team members may label the same lightcurve differently based on their interpretation. Blends and binaries for example can be used differently by different users. We attempted to control for this by manually inspecting a random sample of objects and updating flags when possible to maintain consistency across all fields. The "X" category is also notably inconsistent, with objects that were rejected as planets for many different reasons, including blends and binaries, being given the same label.

The second issue comes from the fact that the classification began before all current data were available. After the first few WASP observing seasons, classifications were made based on the limited data available at that time. When further observations were added in the following seasons, the shape of the lightcurve may have changed and more (or less) transit-like shapes became obvious. However, since the candidate was already rejected, it was never re-visited and updated. Several examples like this were found by looking through the incorrect classifications, and remained uncorrected in our training data. Regardless of these problems, the algorithms were robust and were able to make reasonable predictions even with small variations in the training.

Finally, we rely on the BLS algorithm to provide an accurate best-fit period. This is especially important for the CNN, which only has the lightcurve folded on that period as input. The CNN is therefore not equipped in its present form to handle possible incorrect periods due to aliases or harmonics. It would be possible to augment the code to also include other information for the CNN, such as the data folded on half of the period and twice the period, either in a stack or as a separate entry, to try to identify planets in the data that were found at the wrong period.

3

Ranking

*Since your actions did not produce the desired result, the only advisable so-
lution would be to reexamine your decision making process and look for errors.*

- Data, *Star Trek TNG: Data's Day*

3.1 Results and Trends

The RFC and CNN models described in Chapter 2 were both able to correctly classify the
vast majority of planets. However, the number of false positives generated suggests that
there is still room for improvement in the classification. In this chapter we discuss the
results in more detail and explore the performance of the classifiers on EBLMs. We also
discuss the probabilities, not just the classifications, associated with the models. Finally,
we examine how the probabilities themselves can be used to create a stronger classifier and
apply the principle with a stacked ML model. We apply the stacked model to more than
100,000 unlabeled lightcurves in the northern WASP fields and rank the objects according
to their final class probabilities.

3.1.1 Follow-up Success Rate

Currently, radial velocity follow up of WASP targets takes place primarily using the CORALIE spectrograph at the La Silla Observatory in Chile (Queloz et al., 2000) for southern targets and the Spectrographe pour l'Observation des Phénomènes des Intérieurs stellaires et des Exoplanètes (SOPHIE; Perruchot et al., 2011) at the Haute-Provence Observatory located in France for targets in the north.

The project records of the WASP follow-up program with CORALIE through July 2018 document that 1,234 candidates have been observed and dispositioned. Of those, 150 (12%) have been classified as planets (2 of which are the brown dwarfs WASP-30 and WASP-128), 713 (58%) are binaries or blends, 225 (18%) were low mass eclipsing binaries, and the remaining 146 (12%) were rejected for other reasons, including 60 because the stars turned out to be inflated giants. The SOPHIE follow-up effort has a similar success rate. Of the 568 total candidates dispositioned, 53 (9%) are planets, 323 (57%) are blends or binaries, 116 (20%) were low-mass eclipsing binaries, 72 (13%) were rejected for other reasons including being a giant star, and 4 (1%) were variable stars.

As a comparison, for our RFC, 182 objects were classified as planets, with 45 true positives and 137 false positives, indicating a success rate of 25%. The CNN with the full lightcurve showed even better results, with 81% estimated follow-up accuracy, and when the local lightcurve data was added 75% of the objects flagged as planets were true positives.

This value seems like a striking improvement, but we note that this result overestimates the true success rate. The training and test samples are composed of all samples from the false positives categories, including those which were very clearly not planets without the need for follow-up. These lightcurves are easy classifications to make and bolster the number of reported correct true negatives.

In addition, the test data do not incorporate eclipsing binaries with low-mass stellar companions, which closely resemble planets in their lightcurves and derived features. When we applied the RFC to 399 low-mass eclipsing binary systems, 92 (23%) were classified as planets. The CNN with only the full lightcurve also returned 64 as planets, partially overlapping with the RFC predictions. Adding the local lightcurve information made the CNN more likely to identify EBLMs as planets with 95 (24%) being labeled as planets. In total, 29 EBLM systems were labelled as planets by all 3 models. On the other hand, in 244 cases the 3 models unanimously agreed the object was not a planet. Figure 3.1 shows a subset of the EBLMs unanimously labeled as planets and those unanimously rejected. It is clear from the images that the algorithms penalize deep transits and transits with
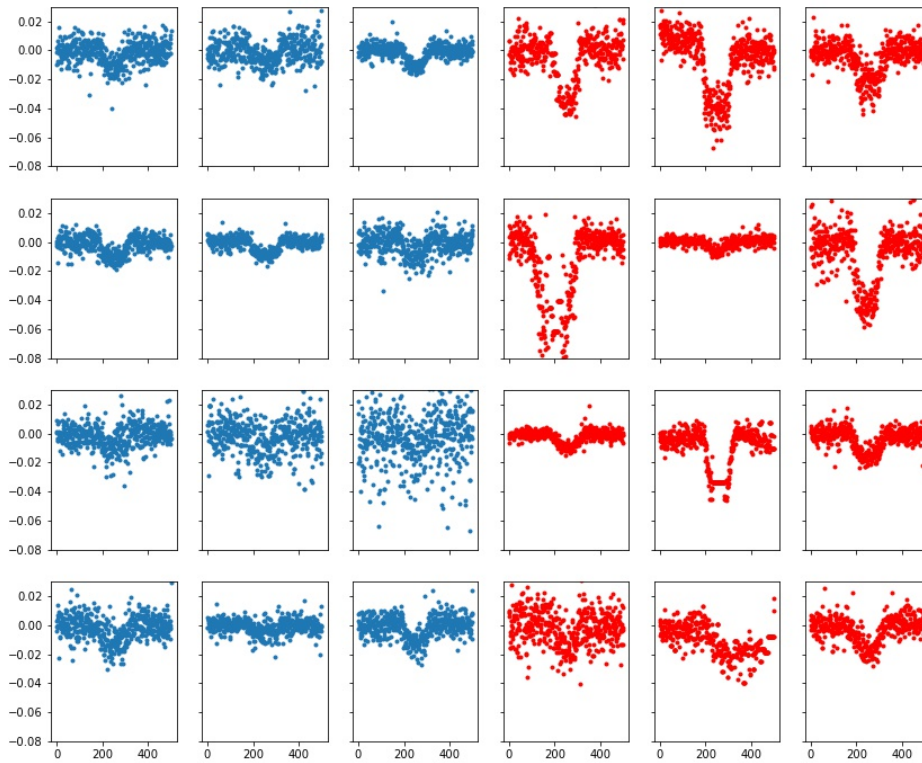
Figure 3.1: Local view of a random sample of EBLMs unanimously labeled as planets (blue) or non-planets (red) by the initial RFC and CNNs.

a strong V shape. The lightcurves that were identified as planets are all shallow with a general U-shape, identical to the transit shape expected of planets. This confirms that the algorithms are sensitive to features we would expect to be important for classification. It also highlights that even if the classifiers performed perfectly, we expect to have many objects that turn out to be low mass stars after follow-up as the transit data alone is not enough to distinguish the objects, lowering the success rate.

3.1.2 Results Discussion

The classifications in this study have a lower accuracy than those reported in the Kepler studies, which range from around 87% (Armstrong et al., 2017) up to almost 98% (Ansdell et al., 2018). This is to be expected as WASP data are unevenly sampled and have larger uncertainties, making definitive identifications impossible with WASP data alone. WASP's large photometric aperture (48 arcsec) also makes convincing blends more common. Nevertheless, the machine learning algorithms were able to correctly identify

∼90% of planets in the testing dataset and operate much faster than human observers (less than 1 minute to train the RFC and around 20 minutes to train the CNNs, and less than a minute to apply to new datasets on a MacBook Pro with 3.1 GHz Intel Core i5) and produce more consistent results. The fast training time means that as new data are added after observing seasons or new data reductions, it is not necessary to manually look at each lightcurve again to see the new data unearthed a new transit signature. Rather, the entire dataset can be quickly re-run through the algorithms to obtain new observing targets.

When looking at the results of the RFC and CNNs, the percentage of correct predictions for the different approaches is consistent, with ∼ 90% of planets being correctly identified. However, when manually checking the original lightcurves for both true positives and false negatives, clear patterns in the different machine learning methods begin to emerge.

The RFC uses features that are derived from the fitted light-curve parameters and external catalog information, but the lightcurves themselves are not included. This logically leads to candidates that have typical characteristics of known exoplanets to emerge. However, because the WASP data can be very noisy and have large data gaps, there are many occasions where the derived "best fit" planet features fall into the known distribution, but upon inspection of the lightcurves it is clear that there is no periodic transiting signal. Examples of true positive and false negative classifications for the RFC are shown in Fig. 3.2. Looking at the samples in the off-diagonal boxes provide insight into how the RFC makes its decisions and what the common failure modes are. For example, SW1832+53 was labeled as an X in the archive, but the RF predicted it was a variable object. This classification was made early in WASP's history, and a clearer picture of the lightcurve has since been established. While an X classification generally means that there is no planetary signal which is true in this case, a better classification for the object would be to label it as a variable lightcurve, which is what the random forest does. SW0826+35 is another interesting object. It was labeled as a variable star in the archive, but the uneven depths indicates it may actually be an eclipsing binary, consistent with the machine classification. For the purposes of planet identification this is not an issue, but if a network was developed to identify contact binaries, the distinction is important. The final object of special note is SW0146+02, verified as WASP-76b. This planet's transit is particularly deep contributing to incorrect EB/Blend label given by the RFC. This shows that this algorithm is sensitive to the depth of transit despite the classic "U" shape of the event.

In general, the main contributors of false positives for the RFC are the blended star (rather than the eclipsing binary) component of the EB/Blend category. In many cases,
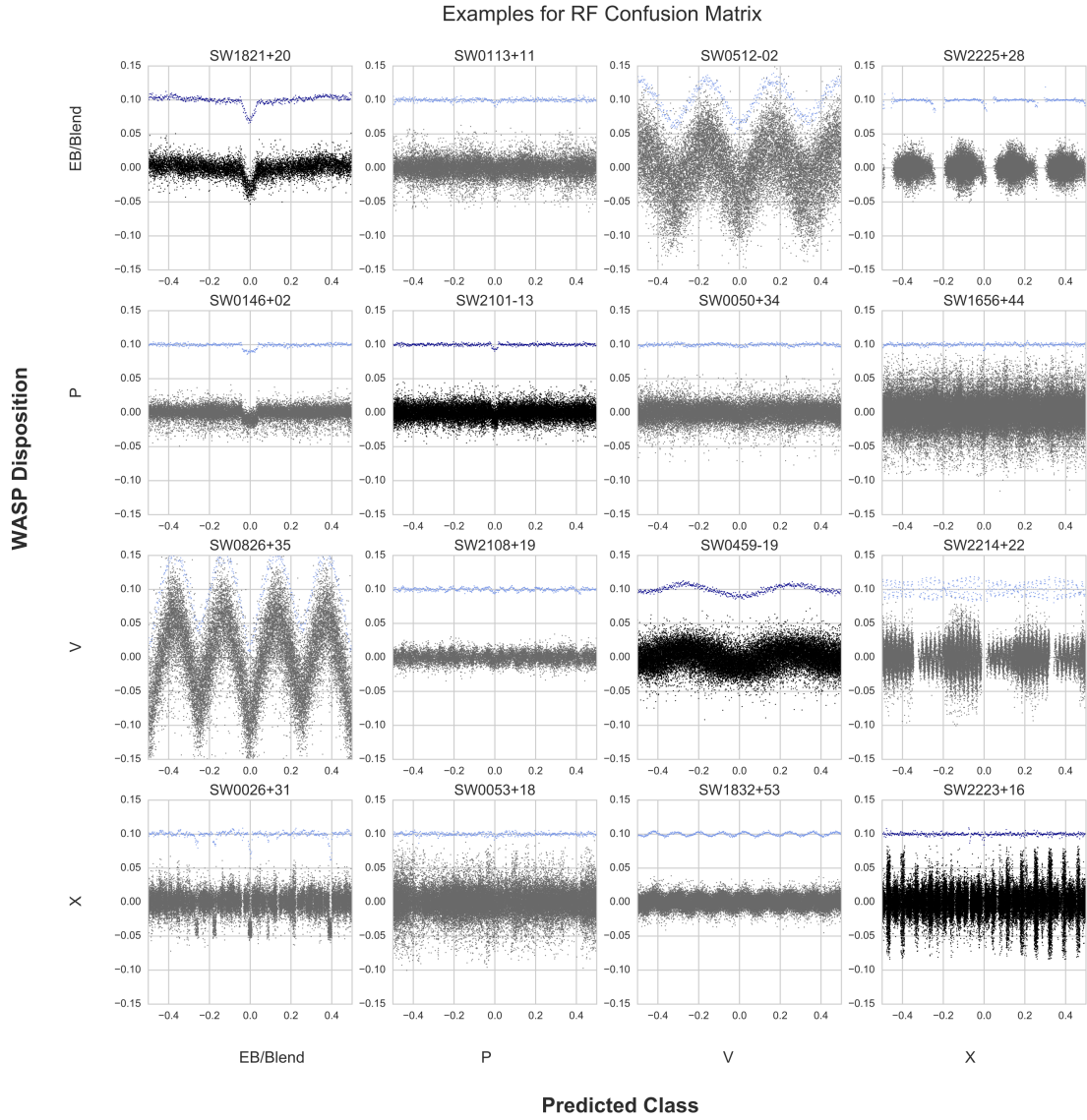
Figure 3.2: Confusion matrix of RFC results showing examples of lightcurves selected from samples that fall into each category, chosen to represent typical failure modes. Lightcurves along the diagonal, shown in black, were correctly classified by the RFC. Off diagonal boxes, shown in gray, were incorrectly identified, with the true classification shown on the vertical axis and the predicted classification shown on the horizontal axis.

the blended stars look very similar to planets by their numerical descriptors; in particular the depth of the transit and the distribution of transit durations look very plausible. Without looking at the nearby stars, these objects are very hard to discern.

The CNN has a fundamentally different method of identifying transits. As described in section 2.3.2, the CNN is not provided with derived data, but rather has direct access to the magnitude data folded on the best-fit period. In this case, the algorithm is essentially trying to pattern-match to find the correct shape for a transit. Accordingly, looking at the true positives and false negatives (examples shown in Fig. 3.3) for this subsection of data shows a different failure mechanism for wrongly-identified planets. SW1308-39 is an example where near-integer day (in this case near 11 days) effects can look like transit signals when the data are phase binned. The CNN did miss several true planets, such as SW1303-41 (WASP-174b; Temple et al. 2018), as the dip is very small with a noisy lightcurve. SW1521-20 is an example of a planet found in a different survey (EPIC 249622103; David et al. 2018) with the signal not being visible in the WASP data. As in the RFC, the overlap in categories in the human classifications is evident in the CNN results. For example, SW1924-33 was labeled by a WASP team member as an X because it does not contain a planetary transit, but it clearly does show a transit event and therefore could be instead classified as an eclipsing binary. While this is considered a wrong classification in the algorithm evaluation, in practice it is an acceptable output.

In many cases, light curves will look like planets, but when other information, such as the depth of the transit, is known it becomes clear that the object is more likely an eclipsing binary or other false positive. In addition, fainter objects tend to have noisier data and more sporadic signals, which can sometimes look like a transit signal when the data are binned down to 500 data points. Finally, the drift of stars across the CCD during each night can lead to systematic disturbances that are consistent at the beginning or end of each night in some (but not all) target stars. Since this effect is specific to each star, it is not always corrected by decorrelation. This can lead to the lightcurve having a clear drop in magnitude at regular intervals, and the gaps in the data can appear transit-like to the CNN. Interestingly, this last problem is far more prevalent when fewer neurons in the ANN are used. Increasing the neurons to 512 and 1024 in our final configuration nearly eliminates the problem, although a few cases do remain.

3.1.3 ML Probabilities

In practice, the machine learning results will be used in combination with expert opinion in order to select the most scientifically compelling targets for follow-up. For example, the area surrounding the star might be crowded with other stars making follow up observations
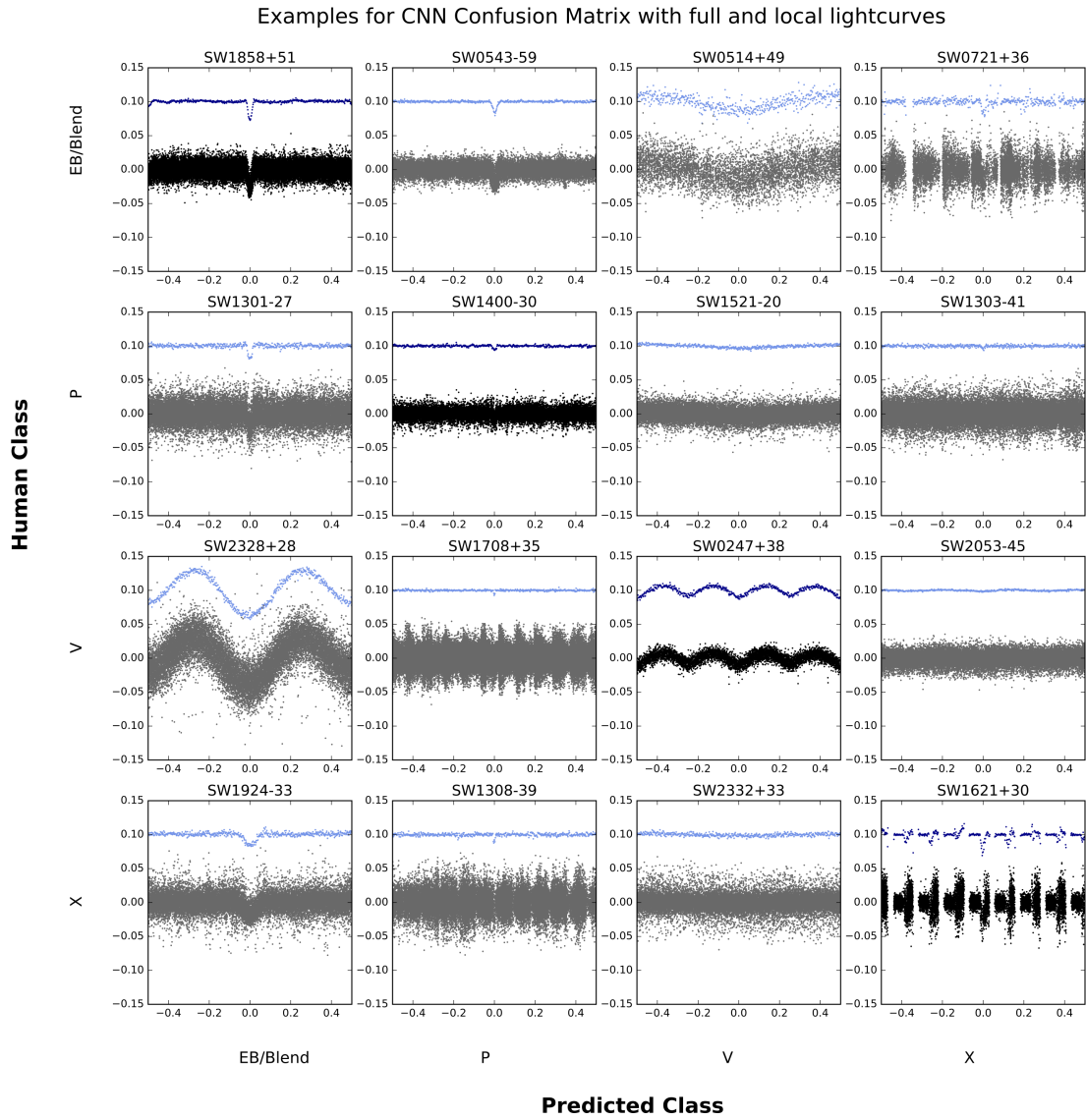
Figure 3.3: Same as for Fig. 3.2, but for the CNN using the local and full binned lightcurve.

difficult. In several cases, a lightcurve looks promising, but another star within WASP's pixel resolution has already been labeled as a Blend (often through follow-up) and the label did not propagate to the surrounding lightcurves. These are easy to identify manually, but that information is not available for the machine learning algorithms to use. Therefore it is still essential that targets selected with machine learning are curated by a human user for practical observation. Another factor only taken partially into account by the RFC is the improvement in the knowledge of stellar parallaxes, and hence radius estimates, made possible by the first and second data releases of the *Gaia* mission (Gaia Collaboration et al., 2016, 2018). Knowledge of the stellar radius, and therefore the radius of the transiting object, aids in distinguishing dwarfs from giants, and eliminates an entire subclass of blended eclipsing binaries at a single stroke.

The confusion matrices shown in 2.4 represent the hard classifications of the RFC and CNN. However, simply assigning a class label does not help in terms of ranking the quality of the candidates to prioritize further observations.

The final output of the RFC, and the output used in the confusion matrices shown previously, can be a simple prediction for the class, assigning a class label for the sample based on the majority vote by the trees in the forest. More interestingly for our purposes, however, is returning instead a probability that the sample falls in each class. The RFC probability is the average of the class probability for each tree in the forest, which for a given tree is measured by the fraction of samples of each class that are in the same leaf as the test sample. The classification is still made by assigning the class with the highest probability as the label, but the output value can be used as an additional measure of confidence that the RFC has in the label.

The probabilities for the CNN predictions are a product of the choice of the sigmoid activation function for the final layer. This is because the sigmoid function produces continuous output values in the range of 0 to 1 for each output neuron. As we are performing a classification task with multiple output classes, the output of the sigmoid function can be interpreted as a confidence level that a given input belongs in each class. Values close to 0 mean low confidence, while values approaching 1 indicate near certainty.

We emphasize that while in machine learning we generally refer to the output of the RFC and CNN as a probability or likelihood, this is not a probability in a statistical sense. A more accurate description is an output score where high numbers mean a higher confidence in the prediction. This is a result of the biases of the algorithms. For example, the RFC makes a probability calculation based on the output of each tree. This means that scores near 0 and 1 are very unlikely as that would mean either none or all of the trees performed perfectly. This pushes the output probabilities towards middling values.

3.1.4 Initial Ranking

The predictions and probability estimates made by the RFC and CNNs used in Chapter 2 were used as a starting point for ranking of planet candidates. An initial assessment was done by visual inspection, with lightcurves being checked if either 1) all 3 algorithms predicted a planetary class or 2) any of the algorithms gave a probability estimate of .4 or greater.

The RFC and CNN each perform best on a specific subset of planets. After manually going through the candidates, several patterns started to emerge that emphasized this point. One frequent failure mode occurred when the RFC labelled an object X and gave a star a very low probability as a planet, but both CNN models labelled it as a planet with high probabilities. In every instance when this dichotomy occurred, the "transit" signal turned out to be a systematic error due to folding the lightcurve at a near integer day. The RFC learned that periods near a day are generally not caused by genuine signals, but the CNN is oblivious to that fact and is fooled by the transit-like shape. After noticing several cases where the performance of the different algorithms could inform better decisions, it became clear that the predictions themselves could be used to better classify the data and obtain a more robust list of planetary candidates.

3.2 Stacking Machine Learning Methods

Many studies focus on comparing the performance of different types of classifiers for a problem and using the best model for implementation (Dubath et al., 2011; Masci et al., 2014; Zhu et al., 2014; du Buisson et al., 2015; McCauliff et al., 2015; Florios et al., 2018; Pearson et al., 2018; Alibert & Venturini, 2019; Domijan et al., 2019; Marton et al., 2019). However, the clear connection between the class predictions and probabilities for the WASP candidates indicates that training a new classifier incorporating this information would increase the performance of the predictor and reduce the time needed to manually examine the resulting candidates, a similar conclusion to that of D'Isanto et al. (2016).

Combining machine learning methods is common practice in the machine learning competition setting on platforms such as Kaggle[1] and the KDD Cup[2]. The Netflix Grand Prize of $1,000,000, awarded to the team that was best able to predict user ratings, was won by a blend of hundreds of different models (Toscher & Jahrer, 2009), bringing attention to the method. While the winning model had excellent performance, it was deemed too complicated to put into practical use. In astronomy, the Galaxy Zoo project created the Galaxy Challenge to model galaxy classifications made by the previous crowdsourcing

---

[1]https://www.kaggle.com/
[2]https://www.kdd.org/kdd-cup
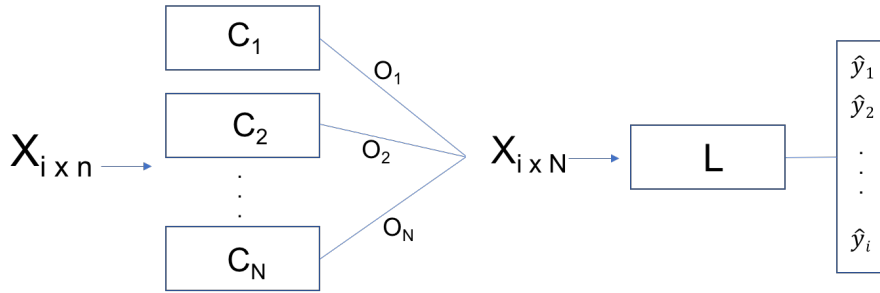
Figure 3.4: Basic flow diagram for a stacked model with i observations with n features. There are N models that make up the first level of classifier. The final classifier L makes the final predictions of the model.

efforts. The winning model averaged the results of 17 variations of a CNN model (Dieleman et al., 2015).

The importance of the combination of machine learning algorithms has also been noted by others (e.g. Morii et al., 2016; D'Isanto et al., 2016; Kodikara & McHenry, 2020), and will be an important framework for upcoming large-data surveys. There are several possible ways in which the results can be combined. For example, the lightcurve data and the domain data could be combined and analyzed by a single model, as done by Ansdell et al. (2018); Chaushev et al. (2019); Osborn et al. (2020). Another approach is to combine the predictions of several individual models as an ensemble. This approach is conceptually similar to the RFC, in which the individual predictions of many imperfect trees are combined to make a better classifier. In a similar vein, Shallue & Vanderburg (2018) used the average of 10 CNN models with different random initializations for their final model.

However, we want to incorporate the predictions and probabilities themselves into a new blanket model. This idea led to the development of a stacking method (Wolpert, 1992). The basic idea of stacking is that a series of individual models $C_1, C_2, ...C_N$ are fit on the data x_i and (for supervised methods) labels y_i, producing a series of predictions $O_1, O_2, ...O_N$. The outputs of all first-level models are then used as inputs to a single second-level classifier L which makes the final prediction $\hat{y}$. See Figure 3.4 for a flow diagram describing the process.

As before, the data are broken up into a training dataset and a testing dataset. The predictions are used as a second-level training set, but in order to get predictions on the training data, several sub-models must be created. To do this, the training data set is further broken down into s sections, with each sub-model trained on s-1 sections of data. The predictions (either a classification or a probability score depending on the model) for the final testing section is kept aside for the next stage. This process repeats s times using a different holdout section until every entry in the training set has a prediction

associated with it. The entire original training set is then used (without the predictions that were added in the previous step) to train a new set of models to make classifications and predictions for the second-level test set.

The second level training and testing dataset is now made up of the predictions as features, and the new second-level model is trained using only this information. The final model will in essence learn that if all models agree on a classification, it should also make that same prediction. However, if patterns emerge, such as the low RFC/high CNN pattern described above, the model will recognize the pattern and adjust the classification accordingly. Increasing the number of models will generally improve performance, provided that the models are uncorrelated with each other (Breiman, 1996). This method was explored by Zitlau et al. (2016) for predicting galaxy redshifts. Although the input models were all variations of the same model rather than many different types of models as we do, they found that stacking improved the performance over any individual model, with the magnitude of the gains being inversely related to the performance of the lower-level models. The revised data sets, models, and stacking architecture used in our application are described in more detail in the following sections.

3.2.1 Refined Training Set

In combination with the new stacking approach, we also make several key changes to the training data. The primary change is that we eliminated the use of artificial data either through SMOTE sampling or the injection of artificial lightcurves. This decision was driven by the findings by Chaushev et al. (2019); Osborn et al. (2020) and others in which the performance suffered when training on simulated data. Additionally, we note that while the CNN testing dataset showed that 88 and 89% of planets were recovered in the test set, only 52 and 43% of the genuine planet transit signals were found, highlighting that there is a clear difference between the genuine transits and those artificially created.

Instead, we bolster the planet category by including a select group of EBLMs. This is for two reasons: 1) many EBLMs are indistinguishable in their transit signal because the radii range of EBLMs and planets overlap, and therefore are interesting and worthwhile objects to follow up (See Fig. 3.1); and 2) studies such as Chaushev et al. (2019) and Reis et al. (2019) have shown that machine learning can be robust to misclassifications in the training data, and therefore may be able to separate the true imposters from those that are in fact out of the planetary parameter space. We therefore include all EBLMs that were classified as planets by either the RF or CNNs described in Chapter 2.

The second change applied to the numerical data sets is an improvement in the features. The data from *Gaia* DR2 are better incorporated, using the stellar radius value from

Table 3.1: Features used by the RFC, LR, KNN, and SVC.

Feature Name	Description
clump_idx	Measure of the objects in the same field with similar period and epoch
dchisq_mr	Measure of the change in the χ^2 when MCMC algorithm imposes a main-sequence (MS) prior for mass and radius.
depth	The depth of the BLS fit transit from Hunter.
dilution_v	Estimate percent dilution from nearby stars.
kurtosis	Measure of the shape of the local lightcurve.
near_int	Measure of nearness to integer day periods, abs(mod(P+0.5,1.0)-0.5).
npts_intrans	Number of datapoints that occur inside the transit.
ntrans	Number of observed transits.
period	Detected period by the WASP pipeline
Rplanet	Radius of the planet, calculated using depth and Rstar.
Rstar	Radius of the star from *Gaia*, MCMC, or J-H color measure (see text).
sde	Signal Detection Efficiency from the BLS.
skew	Measure of the asymmetry of the flux distribution in the local view.
sn_ellipse	Signal to noise of the ellipsoidal variation.
sn_red	Signal to red noise.
teff_jh	Stellar effective temperature, from J-H color measure.
trans_ratio	Measure of the quality of data points (data points in transit/total good points)/transit width.
vmag	Cataloged V magnitude.
width	Width of the determined transit in hours.

the survey whenever available. If a *Gaia* radius presented by Andrae et al. (2018) is unavailable, the radius estimate from the initial WASP MCMC fitting run is kept, and in small number of cases where the MCMC data are unavailable, a radius estimate from the J-H color is used. This single radius value is then used to recalculate the estimated planetary radius for all objects. As the depth (δ) is unchanged, the updated R_p is found using

$$R_p = \sqrt{\delta} * R_* \tag{3.1}$$

Additionally, we use the feature ranking results from the initial implementation to remove some of the uninformative features. This is essential because we will incorporate different ML methods that are not as robust to bad features as the RFC. The stellar mass estimates are removed entirely, as they are found by a simple radius-mass relationship, and are therefore highly correlated values that do not add additional information to the system. Finally, the skew and kurtosis values were recalculated based on the binned local data, rather than the entire WASP lightcurve for the object, in an effort to better capture the shape differences. The final list of parameters used in this data set is found in Table 3.1.

Table 3.2: Values for the tuning parameters for the models used for stacking.

Model	model parameters
RF	ntrees=250, nfeatures=6, nbranches=15, class_weight='balanced_subsample'
KNN	nneighbors=3, weights='distance'
LR	c=100, tol=0.0001, max_iter=400, class_weight='balanced'
SVC	c=1000, tol=0.0001, class_weight='balanced'
CNN	See Figure 3.5
Stacked LR	c=100, tol=0.0001, max_iter=400, class_weight='balanced'

Finally, we change the structure of the CNN to incorporate the periodogram binned in $\log P$ so that the presence or absence of harmonics in the periodogram can help to determine if the signal is from an astrophysical source or is due to systematics. The local lightcurve and the periodogram are scaled between the values of 0 and 1 prior to the convolutional steps.

The structure of the CNN implemented in this second phase of ML is substantially altered from the original CNN. We bin the full lightcurve to 500 datapoints as before, stack that image with the binned periodogram, and run three iterations of two convolutional layers, a max pooling layer, and a dropout layer. We found that in many cases, binning the local lightcurve to 500 datapoints using WASP data led to many lightcurves with large stretches of null data, so we reduce the binning to 250. As such, we treat the local lightcurve separately and apply two iterations of convolutional, max pooling, and dropout layers. The data from both inputs is then flattened and combined before entering the fully connected neural network. As before, we use the ReLu activation function for all layers except the final output layer, where a sigmoid function is used instead. The final structure can be seen in Figure 3.5.

The more models that are stacked, the better the performance of the final algorithm is likely to be. We therefore implement several different machine learning algorithms, utilizing the SCIKIT-LEARN[3] library in Python. We briefly describe these new methods here. The tuning parameters for all algorithms were determined using 3-fold cross validation over a grid of parameter values scored using a weighted f1 score. The final values for the models are shown in Table 3.2.

3.2.2 Logistic Regression

The logistic regression (LR) function is frequently used to assign a probability to a class label. The LR can be thought of as a simplified neural network without any hidden layers. As this is a multi-class problem, we use a variant called multinomial logistic regression

---

[3]https://scikit-learn.org/
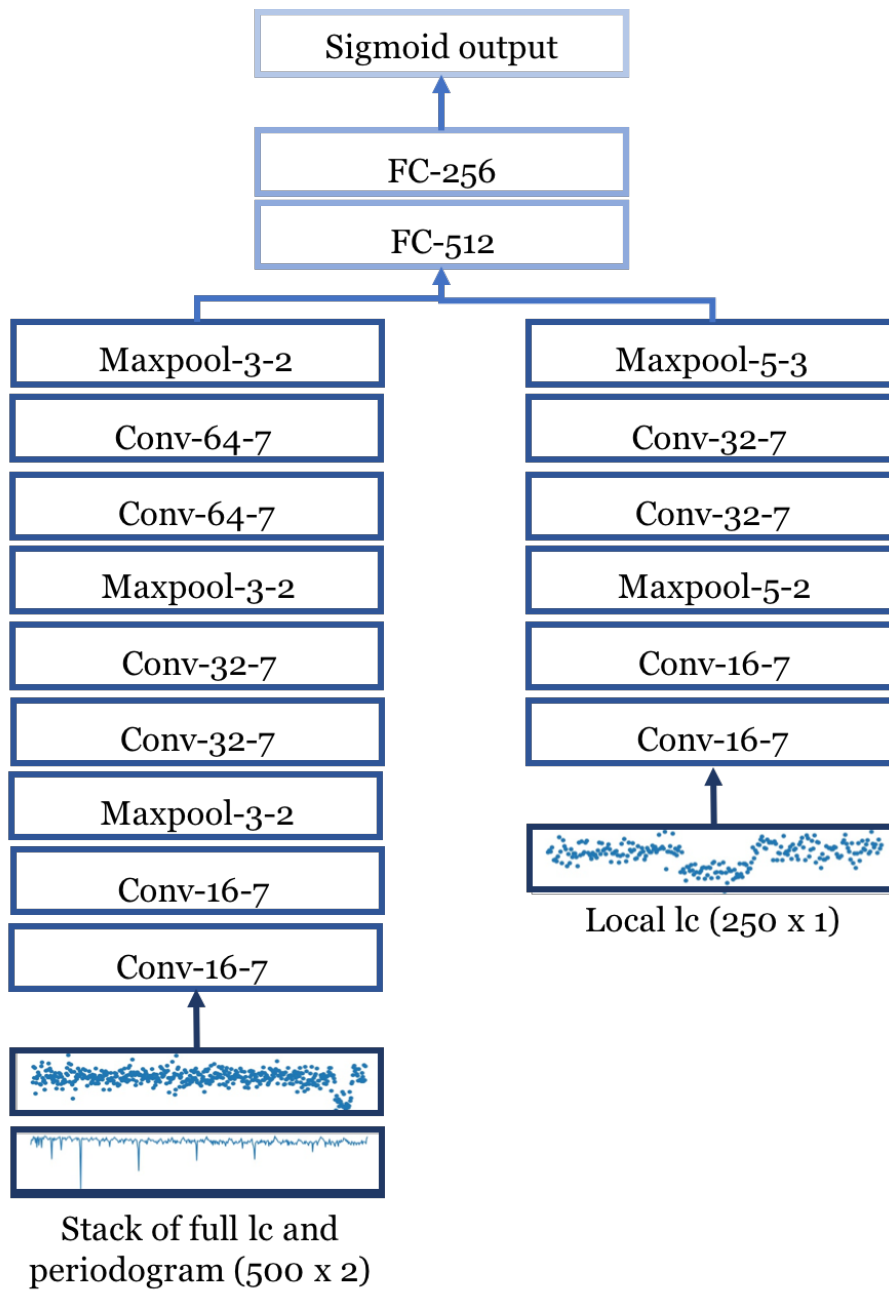
Figure 3.5: Structure of the CNN model used as one of the models incorporated into the final stacking model.

where one category is used as a baseline to which the log-odds of all other categories are found. The probability is given by:

$$p(y = c \mid x) = \frac{e^{w_c \cdot x + b_c}}{\sum\limits_{j=1}^{k} e^{w_j \cdot x + b_j}} \tag{3.2}$$

where c is the class, x represents the features, k is the number of classes, and w and b are the weights and biases that are learned via training. We use the cross entropy loss function, as described in 2.3.2 with the "lbfgs" solver, an approximation for the Broyden-Fletcher-Goldfarb-Shanno algorithm, for optimization. This solver is recommended by SCIKIT-LEARN for use with small datasets, and indeed when testing different solvers this one showed the best performance. To help reduce overfitting on the training data, we use the default L2 regularization, which adds the additional term

$$L2 : \frac{1}{2} \sum_{j=1}^{k} w_j^2 \tag{3.3}$$

to the cost function. In the SKLEARN implementation, the strength of the regularization is controlled by the parameter C, which we set to 100.

3.2.3 K-Nearest Neighbors

K-nearest neighbors (KNN; Altman, 1992) is a simple algorithm by which a classification is made based on what the majority of K-neighboring datapoints are labelled, with the K value being determined for the dataset at hand. This is a form of non-generalized learning, as it doesn't create a model from the data, but merely stores the training examples to use for classifying further data. We find that this method performs poorly regardless of the K-value, but in general performs less poorly for identifying planets with low values of K. We therefore implement a K value of 3. We also weight the contribution of each of the neighbors by the inverse of the distance to the testing datapoint so that close neighbors have a larger impact on the final classification.

3.2.4 Support Vector Classification

A Support Vector Classifier (SVC; Boser et al., 1992) looks for the best decision boundary to separate the different classes of data. This decision boundary in an N-dimensional space is a flat N-1 dimensional plane that divides the space into two parts, known as the hyperplane. For a multi-class problem, the SVC implemented in SCIKIT-LEARN uses a one-vs-one scheme, where a boundary is fit between each unique pair of classes, leading

to (k*(k-1))/2 total models. Each model will make a prediction on a new datapoint, with the classification given by the majority vote.

To find the best decision boundary, the SVC will try to maximize the margin between the data points nearest to the hyperplane for each class, called the support vectors, and the hyperplane itself. The data are converted to a higher dimension to better separate the categories. We use the polynomial kernel function with the default 3 degrees. The main tunable parameters for the SVC are the C value, which controls the amount of slack that the model gives to outliers in the data and the tolerance value which affects the stopping criteria for the optimization function.

3.2.5 Final Model

The training dataset is broken into 5 sections. Over 5 iterations, one of the sections is withheld and a RFC, CNN, LR, KNN, and SVC model are trained on the remaining 4 sections. In the case of the RFC, CNN, and LR model, the probabilities that a sample falls in the four classes is calculated and the prediction is made by taking the maximum value. In this case, the four probability values are stored for later use. The SVC and KNN models we implemented only return final classifications, and therefore only a single prediction value is stored.

The entire training data set is then used to train a series of new models with predictions and probabilities made on the test dataset saved as before.

The second-level training and test datasets are comprised of only the saved probability and prediction values, that is to say the final model does not have access to any of the features describing the physical system. Instead, the final logistic regression model uses the confidence of the models and the agreement or disagreement amongst models to make a final prediction. An overview of the process is shown in Figure 3.6.

3.3 Results

Several different metrics can be used when assessing the performance of a model and are derived from four different values that describe the results of prediction on the testing set. These values are the True Positives (TP), when the model correctly predicts the positive class; True Negatives (TN), when the model correctly predicts the negative class; False Positives (FP), when the model wrongly identifies the negative class as positive; and False Negatives (FN) when the model wrongly identifies the positive class as negative. For this work, we are most interested in finding planets and reducing the number of false positives. We can therefore simplify the problem by treating the results as a binary problem, with
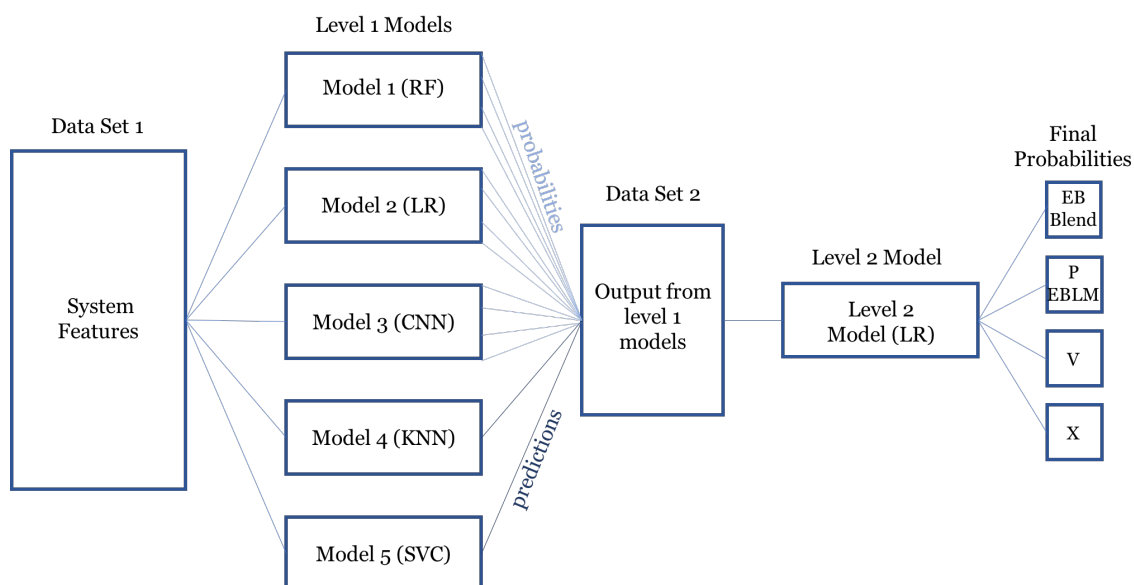
Figure 3.6: Final configuration of the stacked model used for planet candidate identification and ranking.

"P/EBLM" representing the positive class and all other categories making up the negative class. The TP, TN, FP, and FN values can then be determined from the appropriate values from the confusion matrix as shown in Fig. 3.7.

The four values can be interpreted in different ways depending on the desired objective of the predictor. The precision is a measure of how well the model does at only identifying members of the positive class as positive, and is defined as:

$$\text{precision} = \frac{TP}{TP + FP} \tag{3.4}$$

Recall, on the other hand, is a measure of the ability of the model to find all of the positive class in the dataset:

$$\text{recall} = \frac{TP}{TP + FN} \tag{3.5}$$

The recall is often also referred to as the true positive rate (TPR). Often it is desirable to strike a balance between the two, so that the model is able to find a lot of the positive class, but not at the expense of additional false positives. In this case, the F1 score is often used:

$$F1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{3.6}$$

This is more informative than the simple average of the precision and recall because it penalizes extreme values for either.

Finally, the False Positive Rate (FPR) is a measure of the number of negative cases

Figure 3.7: Interpretation of the confusion matrix when converting to a binary classification problem with P/EBLM as the positive class. All empty squares represent True Negatives.

that are identified as positive:

$$FPR = \frac{FP}{FP + TN} \tag{3.7}$$

The results of the constituent models and the final stacked model are shown in Table 3.3. While the RFC shows the overall best precision, it suffers greatly in recall indicating that it successfully rejects non planets at the expense of rejecting a lot of true signals. The CNN shows the opposite trend, where the precision is quite low but has the highest recall rate, missing only 8 planets and EBLMs in the test set. The final stacked model shows a more balanced approach, finding 64 of 73 P/EBLMs with only 152 false positives.

Of the 9 missed objects in the final model, 6 are EBLMs labeled as EB/Blends. The remaining 3 objects that were missed are planets. The first one (WASP-43; Hellier et al., 2011) was labeled as an EB/Blend, likely because this planet has an unusually deep and V-shaped transit with a very short period, making it an outlier in the planetary distribution. The remaining two (XO-1; McCullough et al., 2006) and (Kepler-16; Doyle et al., 2011) were rejected with an "X". XO-1b has a complex periodogram with several spurious signals disguising the harmonic pattern. Kepler-16, on the other hand, is on a 229-day orbit, which is not detectable by WASP at all. The signal found by WASP, and the one seen here, is a false detection at ~8 days caused by phasing, making the X prediction correct for the data at hand. These three planets as well as three misidentified EBLMs are shown in Figure 3.8.

The above analysis only takes into consideration the final classification and not the
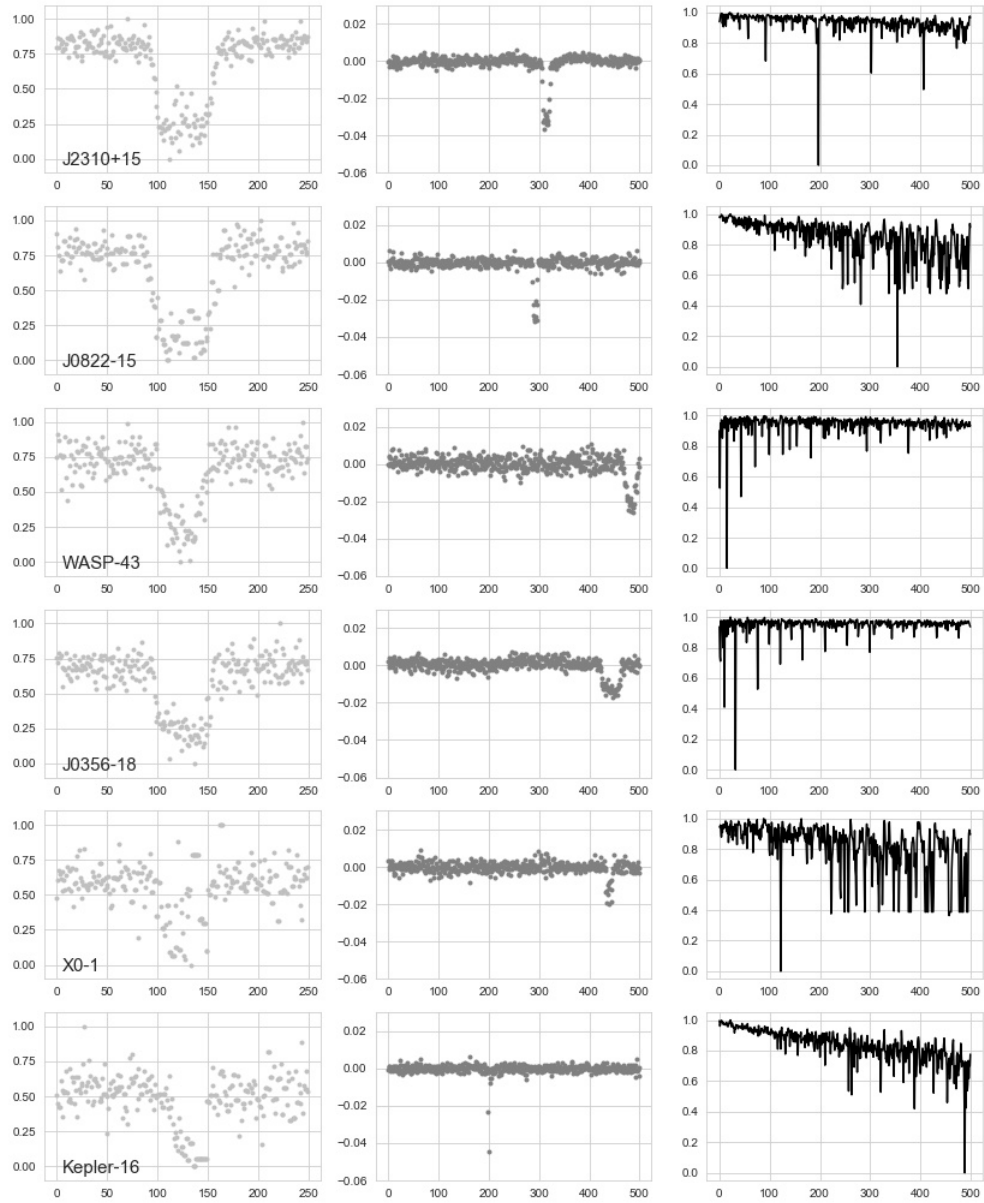
Figure 3.8: Example of the binned local lightcurve (left), full lightcurve (middle), and periodogram (right) for the three planets and three of nine EBLMs in the test data set missed by the stacked model.

Table 3.3: Results of the individual models and the final stacked model, assuming the P/EBLM class is positive and all other classes are negative.

Model	TP	FP	FN	Precision	Recall	F1	FPR
RFC	42	53	31	0.442	0.575	0.500	0.032
KNN	9	29	64	0.237	0.123	0.162	0.018
LR	60	221	13	0.214	0.822	0.339	0.135
SVC	51	137	22	0.271	0.699	0.391	0.084
CNN	65	209	8	0.237	0.890	0.375	0.128
Stacked LR	64	152	9	0.296	0.877	0.443	0.093

probability that a given sample falls into the category. As we also are interested in the relative strength of each candidate, the probabilities are an important measure. In particular, we are interested in knowing what an acceptable cutoff threshold is for planet candidates to maximize the number of true positives while limiting false positives. A commonly used tool to interpret the probability output of a model is the Receiver Operating Characteristic (ROC) curve.

An ROC curve is shown by plotting the false positive rate versus the true positive rate for a range of threshold values. A perfect classifier would follow the left border of the plot and then the top border forming an inverted L-shape. A very poor classifier would follow the diagonal from bottom left to upper right. A useful property of the ROC curve, then, is finding the area under the curve (AUC), where the perfect classifier would have an area of 1 and the poor classifier would have an area of .5. If the classifier performs below that threshold, it would mean that essentially the output labels are flipped, so by changing the labels the classifier would have an improved performance.

The ROC curve is inherently a binary measurement, so we calculate 4 different curves for each model, each assuming a different class as the positive class while the other 3 are lumped together as the negative class. As the curves measure the trade off between the TPR and FPR as a function of probability threshold, we only measure the curves for the LR, RFC, and CNN models. The final ROC curves are shown in Figure 3.9. The plots also indicate the average AUC weighted for the size of each class for the models. The initial LR model shows the lowest performance with an AUC of .868, followed by the CNN with a .886 and the RFC with an AUC of .920. The final stacked LR model has an AUC of .930. The ROC curves clearly show that all models show the lowest performance for the EB/Blend category, possibly reflecting the wide range of objects that make up this category.

As the ROC curve is calculated as a function of threshold value, we can use it to inform a cutoff value based on our follow-up potential. If resources and time are limited, we may restrict ourselves to the point on the P/EBLM curve before the curve makes a
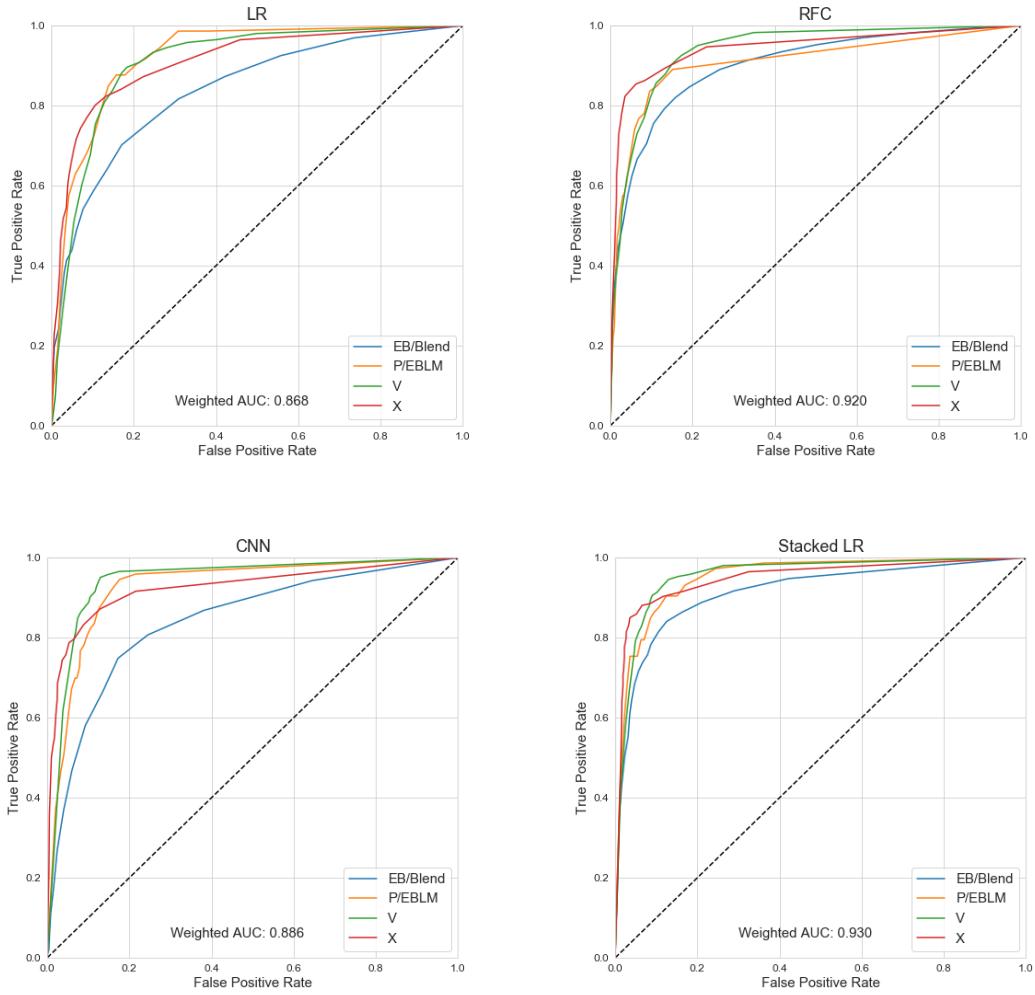
Figure 3.9: ROC curves for the LR, RFC, and CNN models as well as the final stacked LR model.

sudden small step to the right, with an FPR of 0.044 and a TPR of .753. This occurs when considering only objects where the final probability for the P/EBLM class is greater than or equal to 0.75. If the goal is instead to conduct a more thorough survey and we are interested in identifying smaller and harder to detect planets, we may choose the point in the curve before flattening horizontally. This lowers the threshold significantly, with stars with a P/EBLM probability greater than 0.2 included, giving a FPR of 0.170 and a TPR of 0.932.

3.4 Candidate Selection

After training and testing the model, we apply the final model to a set of 108,215 SuperWASP lightcurves for which no label is currently applied. Prior to running the RFC, SVC, LR, and KNN models the data set is scaled using the parameters set by the training

67

data.

In total, the highest probability value was for the planet or low-mass eclipsing binary class for 2,498 of these lightcurves. Using the conservative threshold value of a P/EBLM probability larger than 0.75, as discussed in section 3.3, there are 506 candidates, while a generous threshold of 0.2 leads to 6,704 candidates. These candidates are not intended to be prioritized sight-unseen, but rather compose a much smaller pool to be inspected by WASP team members.

The objects classified as P/EBLMs have stars ranging from a *V*-magnitude of 7.74 to 17.97 with a median value of 13.18. For practical observing, brighter targets are preferred, especially when taking radial velocity measurements. If we restrict the results to candidates with a *V*-magnitude less than 12, the pool of candidates shrinks dramatically (as would be expected as the sampling volume decreases by a factor of $10^{-0.6}$ for each magnitude), with only 325 P/EBLM predictions, 278 of which have magnitudes between 10 and 12. 34 and 1,094 candidates fall above the 0.7 and 0.2 thresholds respectively.

With the magnitude restriction in place, we identify the 12 highest-rated objects in the archive to discuss further. The CNN input, consisting of the local and global view of the transit as well as the periodogram for the highest-ranked candidates are shown in Figures 3.10 and 3.11. Note the y-axis of the local view and the periodogram are shifted as the data are adjusted to range between 0 and 1 before input into the CNN. A table containing all of the features for these candidates as well as the predictions and probabilities used in the stacked model are available in Appendix A.

By analyzing the targets with the highest probabilities in the P/EBLM class we can gain a better insight into the strengths and failure modes of the model. Here we discuss the results in the context of the WASP data available, including the lightcurve, periodogram, descriptive features, and data available on the WASP candidate webpage not directly available to the ML algorithms. We comment on the additional information gained from the TESS survey in Section 4.3.

Reassuringly, the local lightcurves in all cases show reasonable, flat-bottomed transit shapes, although the scatter present even in the binned data makes exact shape determination impossible. Note that J1454+14 to the observer may look like a poor fit as we the human observers expect the local view to be centered on the transit, but the CNN is only looking at small-scale features and is therefore robust to shifts in the data. The CNN therefore may highly rank candidates with spurious signals such as this.

The full lightcurves (middle column) in most cases appear plausible, but four of them do call attention to themselves and require extra checks. J0937+35, J0027+20, and J1454+14
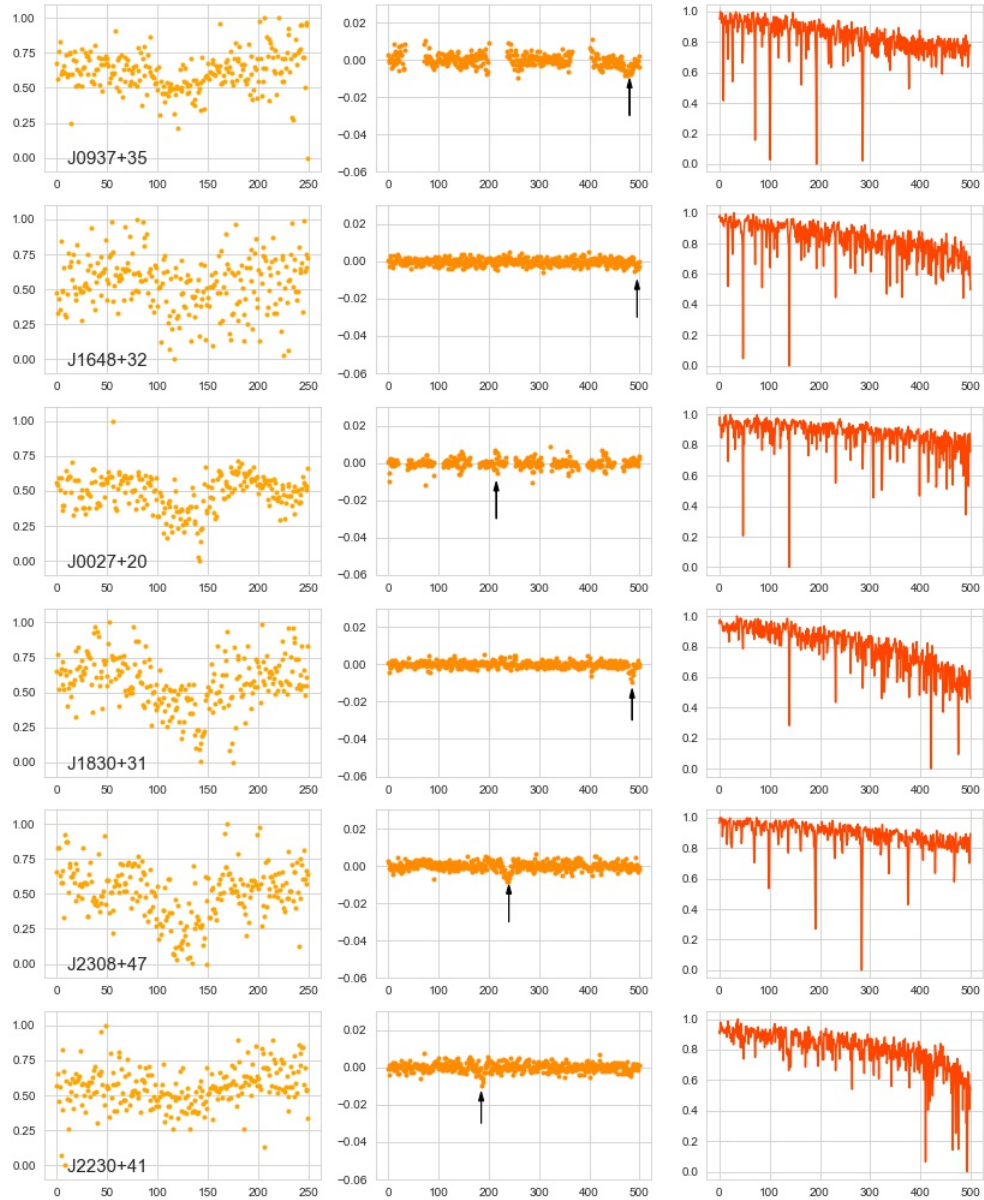
Figure 3.10: CNN input for the top 12 candidates found by the stacked model. The local view is on the left, full lightcurve in the middle with a black arrow denoting the detected transit event, and periodogram on the right, all plotted against bin number. The local view and periodogram are normalized between the values of 0 and 1.
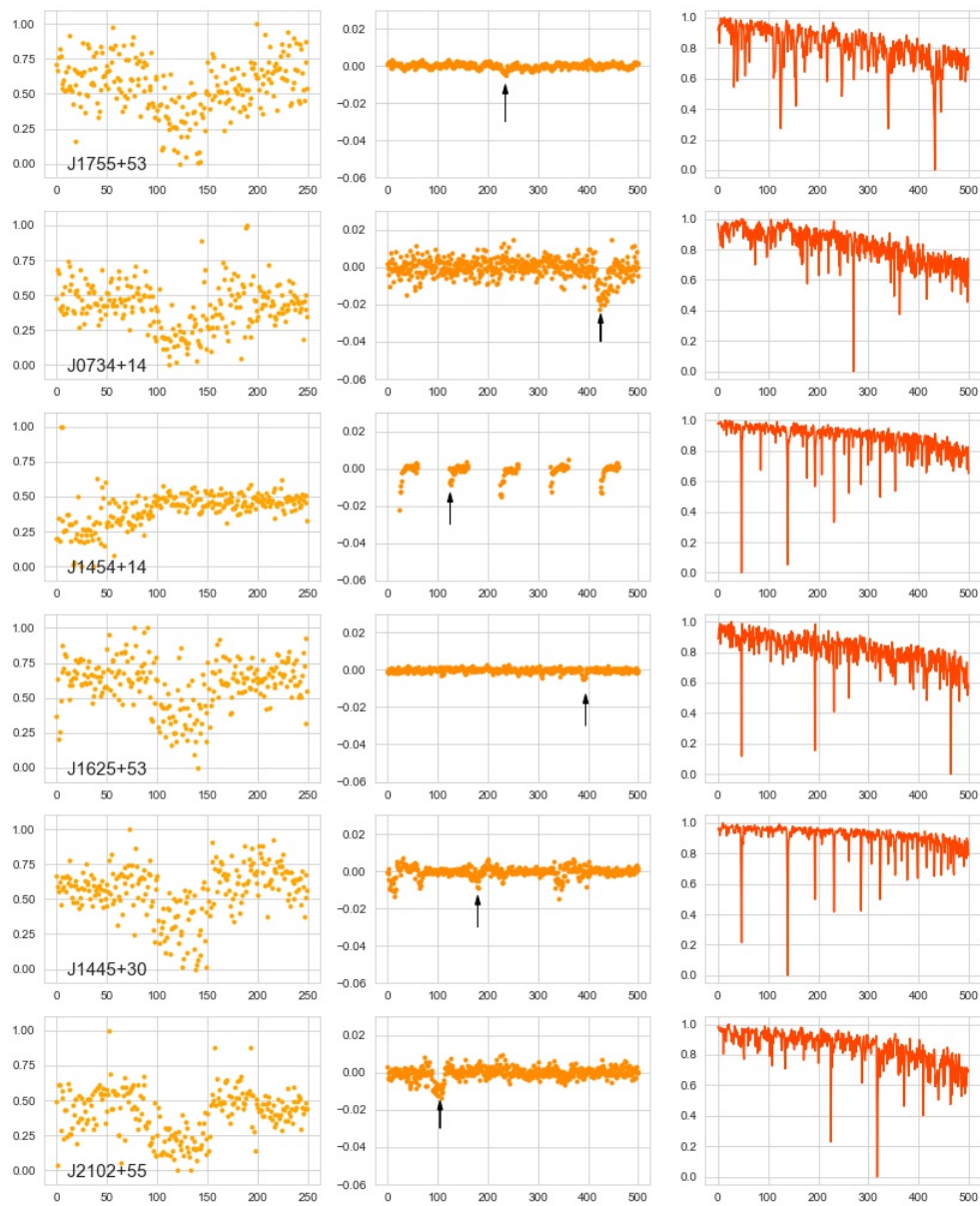
Figure 3.11: Top 12, cont.

all show gaps in the data due to folding on the periods of 1.496 ($\sim$1.5), 6.98 ($\sim$7), and 4.99 ($\sim$5) days. We check the individual transits that make up the folded lightcurve on the WASP webpage and find that transits of J0937+35 are not due to the day/night cycle but instead appear to be genuine in the WASP data as several individual transit detections are in the middle of a single observing night. J1454+14 fails the individual-transit test as all transit detections are at the beginning of an observing night and are most likely due to the changing conditions at the start of each night. J0027+20 appears to be folded at the wrong period, as the full lightcurve shows 7 small transit dips that were flattened out in the binning process. It is therefore most likely that this is a systematic issue, as the period is $\sim$7 days. The final questionable lightcurve is J1445+30, which appears to have a double-transit event in 3 intervals. As was the case for J1454+14, we conclude that, given the period of 1.497 days and the timing of the individual transits seen only at the beginning phase of individual observing nights, we expect that the signal is due to the influence of the day/night cycle.

The final input to the CNN is the periodogram. An observer would look for obvious dips in the $\Delta\chi^2$ value that are distinguishable from the baseline noise and show a series of harmonics at multiples and sub-multiples of the transit period. As noted, the periodograms are binned in log P space making the harmonic patterns for different periods related to each other by translation rather than scaling, with higher harmonics decreasing in depth.

As the CNN takes 1-dimensional data, the association between the period and the harmonics is lost which can lead to high, but incorrect, probabilites from the CNN. For example, J1625+53 shows what could have been interpreted as a harmonic pattern in the middle of the periodogram, but the transit period is 11.48 days corresponding to the dip on the far right. Looking at the full periodogram data, it is clear that no harmonics exist for that period. In addition, the full lightcurve shows a series of small pulses with many "mini transits", indicating the target is a false positive. While most methods identified this target as an "X", the CNN predicted with very high confidence (0.887) that the object is a planet, which was reflected in the final probability rating. J1755+53 also has an unclear harmonic pattern in the binned data. The full periodogram shows there is a harmonic at P/2, but nothing at 2P. In addition, this star has a faint (vmag=15) companion within the WASP aperture. We conclude that the transit is likely due to blending with the nearby star.

The final important check for all WASP objects is looking at the field surrounding the target star to check the potential for blending. This check quickly rules out J0734+14 as a visual binary with a star one magnitude fainter, indicating the transit depth is diluted by the other star. Even if the brighter star were the source of the transits, the transitor's

radius would be $\sim 2.2 R_J$, placing it above planetary limits. Four of the remaining stars, J1830+31, J2230+41, J2308+47, and J2102+55, are all in crowded fields with one or more faint stars within either the inner or middle WASP aperture. The source of the transit and subsequent transit depth could be determined with follow-up photometry. Finally, J1648+32 is alone in the field, but is at the faint limit with a V magnitude of 12, making RV follow-up difficult.

From this analysis, it is clear that while the stacked model is sensitive to shallow transit-like events, the performance of the final model when applied to new data does not obtain the same precision as the training and test set. This is in part due to the fact that the training and test data are made up of the "low hanging fruit" with the clearest cases already identified in the data. The remaining lightcurves are therefore inherently more difficult cases to classify. The unlabeled data also includes an overall fainter sample, reflected in the mean magnitude mismatch between the training/test sets and the unlabeled set of 11.46 to 12.95. More than 2000 of the P/EBLM predictions are on stars fainter than V-mag $= 12$. In WASP data, faint stars such as this do not have a high likelihood of genuine signals reaching above the white noise, while brighter targets are strongly affected by red noise (Pont et al., 2006; Günther et al., 2017).

However, the ability of the model to reject non-planet like shapes in the lightcurve is an immensely useful tool. The final probabilities reported in Appendix A can be used to take lightcurves out of the potential pool of candidates by eliminating stars that rate highly in any of the other three categories. The lightcurves showing the highest probabilities for the other three classes can be found in Appendix B. These lightcurves are very clearly members in their class. The same process as described for determining a planet candidate threshold can be used to find reasonable thresholds for each class where the periodicity found does not correspond to a planet. This can reduce the lightcurves waiting for human inspection and disposition from over 100,000 to only a few thousand.

4

Validation

There's still much to do, still so much to learn.

- Captain Picard, *Star Trek TNG: The Neutral Zone*

Once a candidate has been identified in the WASP data, the next step is to verify the source of the signal by taking additional observations of the object. Once sufficient follow-up data are obtained to confirm the presence of a planet, the focus shifts to using the observations to characterize the planet parameters. As there are many parameters needed to describe the exoplanet system that need to be fit simultaneously, the Markov-Chain Monte Carlo (MCMC) approach has been widely used (See: Holman et al., 2006; Collier Cameron et al., 2007; Barclay et al., 2012; Alsubai et al., 2013; Cáceres et al., 2014; Rowe et al., 2014; Kostov et al., 2020, among others). The MCMC method enables an efficient exploration of parameter space to both fit the model and provide meaningful posterior probabilities for the variables. Often the stellar parameters are determined separately from the transit and RV fit; here, we develop a new method to fit them simultaneously.

In the present chapter, we provide a brief overview of the equations used to describe a planet and its orbit. Next we describe the integrated MCMC approach and apply the method to the discovery of two new hot Jupiter planets, WASP-186b and WASP-187b.
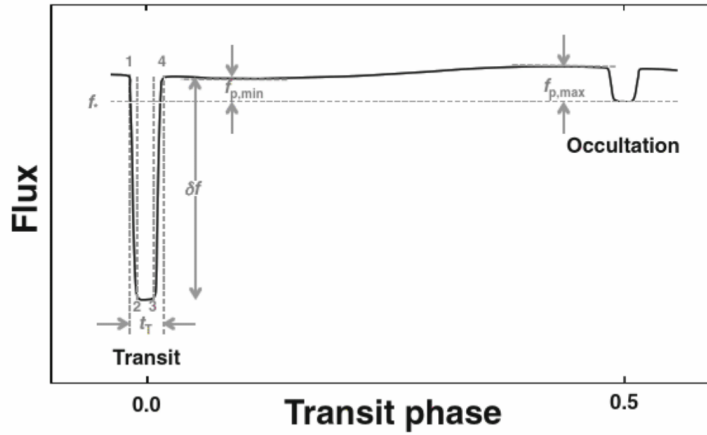
Figure 4.1: Example thermal-infrared lightcurve for a planet in a circular orbit. Adapted from Bozza et al. (2016) Figure 2.2.

Finally, we discuss how data from *Gaia* and *TESS* can be used to quicken the validation or rejection process for ground based surveys such as WASP, using the 12 candidates identified in Chapter 3 example cases.

4.1 Modelling an Exoplanet

4.1.1 Transits

From a transit lightcurve alone, there are four main observable quantities that can be used to characterize the system: the depth of the transit, the duration of the full transit event, the duration of ingress/egress, and the orbital period (when multiple transits are observed). Figure 4.1 shows a model of a planet transit. Unlike the BLS model used to find transits where the out-of-transit and in-transit fluxes are treated as constant values, a real transit shows substantially more variation. The out-of-transit points in Fig 4.1 are varying due to the starlight reflecting off of the planet, adding to the total flux at different levels based on the fraction of the planet's dayside that is visible from the perspective of the observer. The occultation occurs when the planet passes behind the star, and is therefore the only time during the orbit when all of the measured flux is coming from the star itself. Likewise, the transit event is not a perfect box shape, but rather shows four contact points (indicated by the numbers 1-4) of the primary transit. The first contact point shows the moment the planet reaches the stellar limb. The second contact point signifies the moment when the planet is fully in front of the star. Points three and four represent the same configuration but when the planet is moving past the star. Even while in full eclipse, the transit depth is not constant due to the effect of limb darkening. We review here how these measured features contribute to the understanding of the planet.

The depth of the transit indicates the change in flux resulting from the exoplanet

blocking the starlight, and is directly related to the stellar and planetary radius:

$$\frac{\Delta F}{F} \simeq \frac{R_p^2}{R_*^2} \tag{4.1}$$

The transit duration from mid-ingress to mid-egress T and the period P can be combined to get an estimate of the stellar density via:

$$\frac{T}{P} = \frac{1}{\pi} \sin^{-1} R_* \left(\frac{4\pi^2}{GM_*P^2} \right)^{1/3} \tag{4.2}$$

for a planet in a circular orbit with an inclination of 90°. In the event that the inclination is not 90°, this equation generalizes to:

$$\sin(\frac{t_{14}\pi}{P}) = \frac{R_*}{a} \left(\frac{(1 + (R_p/R_*))^2 - ((a/R_*)\cos i)^2}{1 - \cos^2 i} \right)^{1/2} \tag{4.3}$$

where t_{14} is the full transit duration from the first contact to fourth contact of the transit (Seager & Mallén-Ornelas, 2003). In the event that the orbit is not circular, an additional term is needed:

$$\frac{\sqrt{1 - e^2}}{1 + e\sin\omega} \tag{4.4}$$

where e is the eccentricity and ω is the argument of periastron, described more in section 4.1.2. In order to estimate the amount of flux that is blocked at any given moment of a transit, the model must take into account two key parameters: the projected separation of the center of the planet to the star and the ratio of the planetary to stellar radii. The separation of the planet to the star is dependent on the orbit configuration, which can be described using the epoch T_c, period P, inclination i, scaling parameter a/R_*, eccentricity e, and argument of periastron ω.

The transit is then modeled by integrating the surface brightness of the photosphere over the solid angle blocked by the planet, with the estimation dependent on the chosen limb darkening model. Several different formulations can be used to model limb darkening, ranging from a simple linear model to a quadratic approximation to more complicated 4-coefficient models. Morello et al. (2017) compared different limb-darkening models and concluded that the power-2 model (Hestroffer, 1997) showed the best performance, especially with regards to cool stars. The power-2 model is written as:

$$I_\lambda(\mu) = 1 - c(1 - \mu^\alpha) \tag{4.5}$$

where $I_\lambda(\mu)$ is the specific intensity at a given wavelength, μ is the cosine of the angle between the line of sight and the stellar surface normal, and c and α are coefficients to be fit for the star. We adopt this model, using coefficients calculated by Maxted (2018). The final transit model provides flux values calculated for the same times as the observations that are being fitted, and some form of χ^2 minimization is used to determine the best-fitting model for the observed flux.

4.1.2 Radial Velocities

The RV method has been an extremely useful tool for both the discovery and characterization of exoplanets. We briefly describe in this section the method for fitting radial velocity models to the data. For a much more complete discussion of the orbital elements, we refer the reader to Chapter 2 of Perryman (2018).

RV measurements describe the star's projected motion along the line-of-sight as it moves around the system barycenter as a result of interaction with another object. Fig. 4.2 shows this configuration, with M_p the mass of the planet and M_* the mass of the star. The two objects share the same orbital period and eccentricity, but the size of the ellipse is determined by the mass.

The motion of a single elliptical orbit is shown on the left in Figure 4.3. When the orbit is not observed face on, the two-dimensional view is not enough, and instead the three dimensional view shown on the right hand side of 4.3 is needed. The orientation of the orbit relative to the observer is described with the additional parameters of the inclination of the orbital plane i, longitude of the ascending node Ω, and the argument of periastron ω.

The measured radial velocity can be described by:

$$v_r = K[\cos(\omega + \nu) + e\cos\omega] \tag{4.6}$$

where K is the RV semi-amplitude, described by:

$$K \equiv \frac{2\pi}{P} \frac{a_* \sin i}{(1 - e^2)^{1/2}} \tag{4.7}$$

and ν is the true anomaly, which is a time-dependent angle that describes where the object is in the orbit. The RV curve built up from RV measurements over a time span
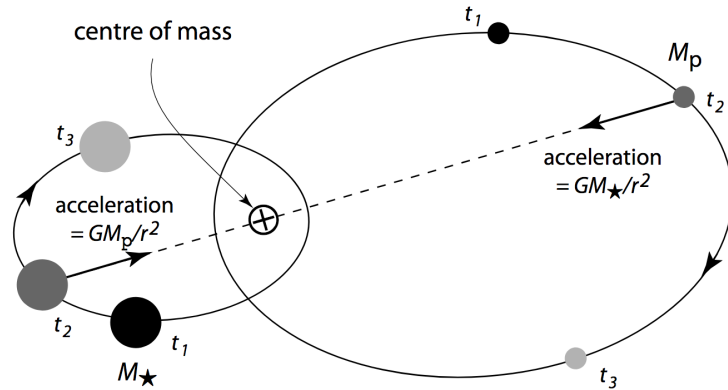
Figure 4.2: Motion of two bodies orbiting their common barycenter, taken from Perryman (2018) figure 2.3
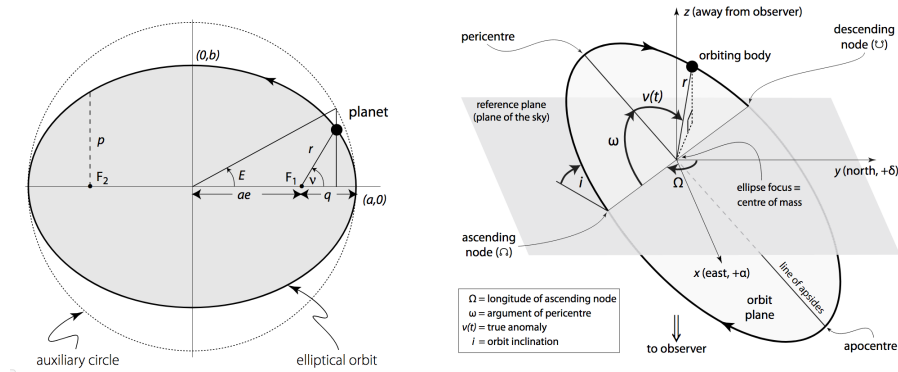
Figure 4.3: Geometry of an elliptical orbit projected on to two dimensions (left) and three dimensions (right), taken from Perryman (2018) figures 2.1 and 2.2

varies between the limits of $K(1 + e\cos\omega)$ and $K(-1 + e\cos\omega)$, with the shape of the curve determined by the e and ω values.

Another key insight from RV measurements comes from an alternative expression of K:

$$K = \left(\frac{2\pi G}{P}\right)^{1/3} \frac{M_p \sin i}{(M_* + M_p)^{2/3}} \frac{1}{(1 - e^2)^{1/2}} \tag{4.8}$$

As the equation suggest, if the mass of the star can be estimated, $M_p \sin i$ can be estimated from the RV measurements. Because the planetary mass is tied to $\sin i$, RV measurements alone do not allow the two values to be disentangled, meaning that RV observations alone can not distinguish a small planet sized object nearly edge on or a massive object that is observed nearly face on. However, when a transit is observed, the inclination is constrained and object mass measurements can be made to determine whether the motion is due to an object of planetary size.

As with the transit model, the RV model is estimated for the times of real data measurements, and the values of K, ω, and e are optimized by searching the parameter space and applying some type of χ^2 minimization.

4.2 WASP-186b and WASP-187b

We present in this work an overview of the discovery of the hot Jupiter planets WASP-186b and WASP-187b, published in Schanche et al. (2020). We refer the reader to that publication for more details, particularly on the observations taken by WASP, SOPHIE, and *TESS*.

4.2.1 Observations

Both WASP-186 and WASP-187 were originally flagged as planet candidates in the WASP survey data after an initial search of the data using the BLS implementation from Collier Cameron et al. (2006). WASP-186 was observed by SuperWASP beginning in 2006 and ending in 2014. The star was originally flagged for further review in July 2016 after inspection of the lightcurve. WASP-187 was observed from 2004 until 2011 and in 2014, the target was flagged for further observations. In addition, WASP-187 was identified by the RFC described in Schanche et al. (2019b) as a good planetary candidate, while WASP-186 was not included in the subset of candidates that were used in the study and therefore was not ranked.

The planetary nature of both WASP-186 and WASP-187 was established, and further

observations were then taken to determine their masses and orbital eccentricities, using the SOPHIE échelle spectrograph at the 1.93-m telescope of the Haute-Provence Observatory (Perruchot et al., 2008; Bouchy et al., 2009)). SOPHIE is dedicated to high-precision RV measurements, and widely used for SuperWASP follow-up (e.g. Hébrard et al. (2013); Schanche et al. (2019b)). Two observing modes are available from SOPHIE. The High-Resolution (HR) mode has a resolving power $R = 76\,000$, while the High-Efficiency (HE) mode has a resolving power $R = 40\,000$. WASP-187 was observed using both modes, so in the analysis HE and HR observations are considered independent datasets to account for different systematic radial velocity shifts. WASP-186 was only observed using HE mode, so all RV data are modelled together. Two low signal-to-noise observations of WASP-186 were not used.

The radial velocities have larger uncertainties than is typical for SOPHIE due to the rotational line broadening. The resulting CCFs have full width at half maximum of 22 and 21 km/s for WASP-186 and WASP-187, respectively. Regardless, they show significant variations in phase with the SuperWASP transit ephemeris, with semi-amplitudes corresponding to companions in the planetary-mass regime. The RVs measured using different stellar masks (F0, G2, K0, or K5) produce variations with similar amplitudes, so it is unlikely that these variations are produced by blend scenarios composed of stars of different spectral types. The line bisector was checked using the approach of (Boisse et al., 2010), and results indicate that the RV variations are not due to spectral-line profile changes attributable to blends or stellar activity, but rather to Doppler shifts due to planetary-mass companions.

The final dataset used comes from TESS, which in October and November 2019 observed both stars in Sector 17. WASP-186 and WASP-187 were given the designations as TESS Objects of Interest (TOI)-1494.01 and TOI-1493.01 with four transits observed for each target. However, a momentum dump (needed to reset the momentum wheel speed) at the beginning of the observation run coincided with the first transit of WASP-186, so we remove the first transit from further analysis. Although the event did not disrupt a transit for WASP-187, we remove the data for this time frame to remove any impact on the out-of-transit measurement. Additionally, the data surrounding spacecraft perigee was removed for both lightcurves.

4.2.2 Stellar Properties

Our knowledge of an exoplanet are directly tied to the knowledge of the host star. Before fitting the transit and RV models, we first conduct a preliminary characterization of the star. We approach the stellar analysis from three different angles.

Firstly, the spectrum was analyzed to estimate the temperature (T_{eff}), surface gravity (log g), and metallicity (Fe/H). For each star, the SOPHIE spectra not polluted by Moonlight were shifted to a common radial velocity and co-added. The spectral analyses were performed using the process outlined in Doyle et al. (2013), whereby T_{eff} was found using the $H\alpha$ line; log g was determined from the Na D and Mg b lines; and the Fe/H was estimated from the width of several Fe I lines; and the projected rotational velocity ($v \sin i$) was determined by convolving the data with the instrument resolution and then fitting the line profile of Fe I, in agreement with results obtained from the CCF using the calibration of Boisse et al. (2010).

Next, using the above derived stellar parameters, stellar masses and ages for WASP-186 and WASP-187 were determined by the isochrone placement functionality of the MCMCI tool (Bonfanti & Gillon, 2020). Following an MCMC approach, with five chains of 100 000 steps and a burn-in fraction of 20 per cent, the stellar masses were calculated by interpolating over grids of stellar isochrones and evolutionary tracks.

The final analysis uses the infrared flux method (IRFM), introduced by Blackwell & Shallis (1977). IRFM is a semi-direct way to measure stellar angular diameter (leading to stellar radius) and effective temperature by combining flux measurements at different wavelengths with stellar atmospheric models to determine the stellar properties. The IRFM method has been implemented by several groups, e.g. (Alonso et al., 1994; Ramírez & Meléndez, 2005; González Hernández & Bonifacio, 2009; Casagrande et al., 2010). For this application, we expanded the method by incorporating data from the *Gaia* spacecraft.

The magnitudes and corresponding uncertainties for WASP-186 and WASP-187 in the *Gaia* G, G_{BP} and G_{RP} bandpasses (Riello et al., 2018) were retrieved from the second data release (DR2) archive (Gaia Collaboration et al., 2018), along with data taken in the J, H, and K filters from the 2MASS survey (Skrutskie et al., 2006) and in the W1 and W2 (3.4 and 4.6 μm) bandpasses from the WISE survey (Wright et al., 2010). The magnitude information is compared to the stellar synthetic spectra atlas of Castelli & Kurucz (2003) to find initial estimates of stellar effective temperature and angular diameter (θ). The radius estimate can be determined by using the *Gaia* parallax measurement (ϖ):

$$R_*/a_\oplus = \theta/2\varpi \qquad (4.9)$$

where $a_\oplus = 1$ au, and θ and ϖ are both expressed in mas.

The three approaches give the stellar estimates provided in Table 4.1. WASP-186 appears to be a typical mid-F type star, while WASP-187's age and radius indicated

Table 4.1: Initial stellar parameters from the spectroscopic (T_{eff}, log g, Fe/H, and $v \sin i$), isochrone placement, (M_* and Age), and IRFM (ϖ, R_*) analysis of WASP-186 and WASP-187.

Parameter	WASP-186	WASP-187
T_{eff} (K)	6300 ± 100	6100 ± 100
log g	4.1 ± 0.2	3.8 ± 0.2
Fe/H	-0.08 ± 0.14	0.0 ± 0.11
$v \sin i$ (km s^{-1})	15.6 ± 0.9	15.3 ± 1.0
M_*	$1.21^{+0.07}_{-0.08}$	$1.53^{+0.07}_{-0.09}$
Age (Gyr)	$3.1^{+1.0}_{-0.8}$	$2.55^{+0.49}_{-0.25}$
ϖ	3.563 ± 0.042	2.667 ± 0.048
R_*	1.46 ± 0.02	2.87 ± 0.05

that the star is a slightly evolved early F-type star. The T_{eff}, log g, and M_* as well as the original ϖ measurement from *Gaia* DR2 are used to constrain the priors for the final MCMC fit. The IRFM radius and parallax measures are used as an independent confirmation for the final MCMC values.

4.2.3   MCMC analysis

The MCMC implementation used is modeled on that of Collier Cameron et al. (2007). The code, written in PYTHON, aims to fit the stellar parameters along with the transit and RV data simultaneously.

The MCMC uses the input parameters to derive models to which the observations are compared to obtain the log likelihood of the scenario for each data set independently. The log likelihoods are combined with the log likelihood that the parameters from the step in the chain match the prior stellar information. The input parameters are then randomly perturbed, new models created, and new likelihoods produced. If the fit to the data is better, the step is accepted. If the fit is worse, the step may be accepted based on a random threshold value, ensuring that the parameter space is well sampled and the MCMC chain does not fall into a local minimum.

To increase the efficiency of MCMC convergence, the jump parameters should be independent of each other (Ford, 2005); therefore the algorithm relies on the transformation of the MCMC jump parameters to the physical variables. The jump parameters used here are: the transit epoch (T_c), period (P), impact parameter (b), transit width (w), transit depth (($R_P/R_S)^2$), received flux (f), parallax (ϖ), stellar radius (R_*), extinction ($E(B-V)$), log of the system error which accounts for zero-point uncertainties in the definitions of the flux-to-magnitude conversions for the different bandpasses in the IRFM calculation (log σ_{sys}), RV amplitude (K), RV offset (γ), and RV jitter (σ_{jit}). As noted

above, the RV offset and jitter are treated separately for the High-Resolution and High-Efficiency modes for WASP-187. The eccentricity e and argument of periastron ω are parameterized as $\sqrt{e}\cos\omega$ and $\sqrt{e}\sin\omega$ (See, e.g. Anderson et al. (2011)); however the eccentricity for WASP-187 was found to be consistent with an eccentricity of 0. Therefore the final MCMC held these values constant. Finally, the transit depth of WASP-187 was underestimated when WASP data was included in the fit, likely due a slight dilution of the signal from the WASP data detrending. To model this star, we fit an additional term to account for the offset.

In order to convert the MCMC jump parameters to the necessary physical descriptions of the system, we perform the following transformations. Firstly, the eccentricity and argument of periastron are found directly from $\sqrt{e}\cos\omega$ and $\sqrt{e}\sin\omega$ using:

$$e = (\sqrt{e}\cos\omega)^2 + (\sqrt{e}\sin\omega)^2 \tag{4.10}$$

and

$$\omega = \tan^{-1}(\sqrt{e}\sin\omega, \sqrt{e}\cos\omega) \tag{4.11}$$

The transit duration is dependent on the impact parameter, so we correct the duration to the value that would be observed if the planet crossed directly past the center of the star:

$$w_0 = \frac{w(1+k)}{\sqrt{(1+k)^2 - b^2}} \tag{4.12}$$

where w is the proxy for the transit width and k is the ratio of the planetary to stellar radii, R_p/R_*. The ratio of the stellar radius to the semi-major axis is then given by

$$\frac{R_*}{a} = \frac{\pi\, w_0\, (1 + e\sin\omega)}{(1+k)P\sqrt{(1-e^2)}} \tag{4.13}$$

where P is the orbital period in days. As described by Winn (2009), $\cos i$ and $\sin i$ are then calculated as

$$\cos i = b\frac{R_*}{a}\frac{1 + e\sin\omega}{(1-e^2)} \quad \text{and} \quad \sin i = \sqrt{1 - \cos i^2} \tag{4.14}$$

The next step is to get the angular radius of the star. First the radius and parallax of the star are used to get the angular radius ($\theta \equiv \dfrac{R_*}{d}$):

$$\theta = \frac{R_* \, (\varpi + 0.082)}{1000 * 180 * 3600} \tag{4.15}$$

where ϖ is the parallax in milli-arcseconds, with the correction of Stassun & Torres (2018) applied to the parallax.

The orbital separation is then found by:

$$\frac{a}{a_\oplus} = \frac{\theta}{\varpi} \frac{a}{R_*} \tag{4.16}$$

with $a_\oplus$ being one astronomical unit and ϖ now expressed in radians.

The mass of the star in solar units can then be calculated using Kepler's third law, using the appropriate unit conversions:

$$\frac{M_*}{M_\odot} = \left(\frac{a}{a_\oplus}\right)^3 \left(\frac{P}{1\text{year}}\right)^{-2} \tag{4.17}$$

The angular radius is also utilized to find the stellar surface flux F:

$$F = f/\theta^2 \tag{4.18}$$

The stellar surface flux jump parameter can be used to derive the effective temperature (T_{eff}) via the Stefan-Boltzmann law. This initial value is adjusted by iteratively applying an extinction correction

$$F_{ext} = F \, 10^{0.4E(B-V)R_{\text{bol}}} \, . \tag{4.19}$$

Here R_{bol} is the bolometric extinction to reddening ratio, determined by optimizing the fit to stellar radii in the asteroseismic samples provided by Silva Aguirre et al. (2015, 2017):

$$R_{bol} = 2.31 + 0.000509(T_{\text{eff}} - 5894) \tag{4.20}$$

In this way, the T_{eff} and $E(B-V)$ values are decoupled from the stellar radius.

With the mass and radius of the star now known, we can calculate the $\log g$ value. Finally, the power-2 limb darkening parameters are interpolated from a grid using the calculated values of $\log g$ and T_{eff}.

The final physical variables that we retrieve from the jump parameters are $\sin i$, $\dfrac{R_*}{a}$, k, T_{eff}, $\log g$, θ, M_*, e, and ω.

4.2.4   MCMC initialization

The initial predictions for the depth, width, impact parameter, period, and epoch for the photometric datasets were found using the Transit Model in the `PyCHEOPS` v0.6.0 Python package[1], which uses a χ^2 optimization with respect to the model parameters using the Levenberg-Marquardt algorithm from the `lmfit` package for Python. The power-2 limb darkening coefficients are interpolated from tables (Maxted, 2018) for *TESS* and WASP separately for this initial fit, as well as at every step in the MCMC chain. The initial values for the RV semi-amplitude, offset, and jitter as well as the eccentricity and argument of periastron were obtained with the RADVEL package [2].

For each step in the MCMC, a log probability value is found by adding the log likelihoods of the transit and RV data with the prior contribution. A transit model is created, and the likelihood that the *TESS* and WASP data fit the model is evaluated. Note that the *TESS* FFI data is binned to a 30-minute cadence, so the model flux is averaged over that time interval as well. The model is created by using the T2Z and QPOWER2 functions of PYCHEOPS to calculate the star-planet separation and generate the lightcurve model using the power-2 limb darkening law, respectively. The RV model used to find the SOPHIE likelihood is computed using the VRAD function.

The system priors come from several different sources. Given T_{eff}, $\log g$, and Fe/H, the stellar surface flux spectrum is determined and attenuated by a galactic extinction law characterised by $E(B-V)$. The resulting reddened spectrum is folded through the photon-weighted filter transmission curves for Gaia, 2MASS, and WISE and scaled by the zero-points and θ^2 to obtain synthetic apparent magnitudes. The residuals of the observed magnitudes minus the synthetic values in the eight bandpasses and the measurement uncertainties are incorporated into the likelihood at each jump in the MCMC chain. The remaining priors come from the parallax measurement in *Gaia* DR2, T_{eff} and $\log g$ from the spectroscopic analysis, and M_* from the isochrone placement.

The MCMC will converge to near-optimal values before settling in to explore the optimal parameter space; this burn-in phase is removed so that the final MCMC chain is only sampling around the solution. We implement 3 burn-in phases of 6,000, 2,000, and 2,000 steps, with updates of the jump lengths after each phase. The final MCMC was run for 50,000 steps.

---

[1] https://github.com/pmaxted/pycheops
[2] https://github.com/California-Planet-Search/radvel
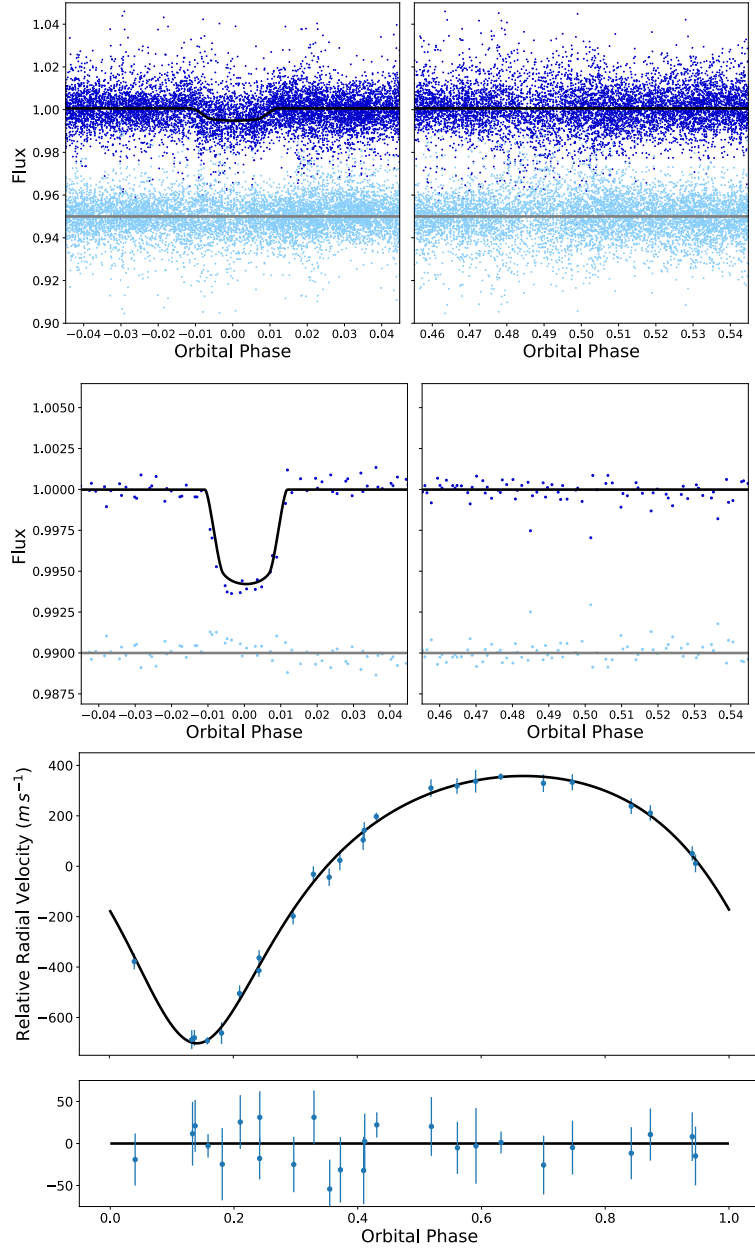
Figure 4.4: WASP (top), *TESS* (middle), and SOPHIE (bottom) data for WASP-186b phase folded on the best-fitting period. Residuals to the fit are shown below the data.

4.2.5 WASP-186b and WASP-187b Parameters

The final transit and RV models after the MCMC fit can be seen overlayed on the data in Figures 4.4 and 4.5, with the final stellar and planetary values reported in Table 4.2.

WASP-186 is a mid-F type star with an effective temperature of 6361^{+105}_{-82} K, agreeing with the spectral determination within errors. The parallax estimate corresponds to a distance of $280.71^{+13.05}_{-11.96}$ pc (Bailer-Jones et al., 2018). The star is rotating with a $v \sin i$ of $15.6 \pm 0.9 \, \text{km s}^{-1}$. As can be seen in Figure 4.6, the planet has a radius typical for a hot
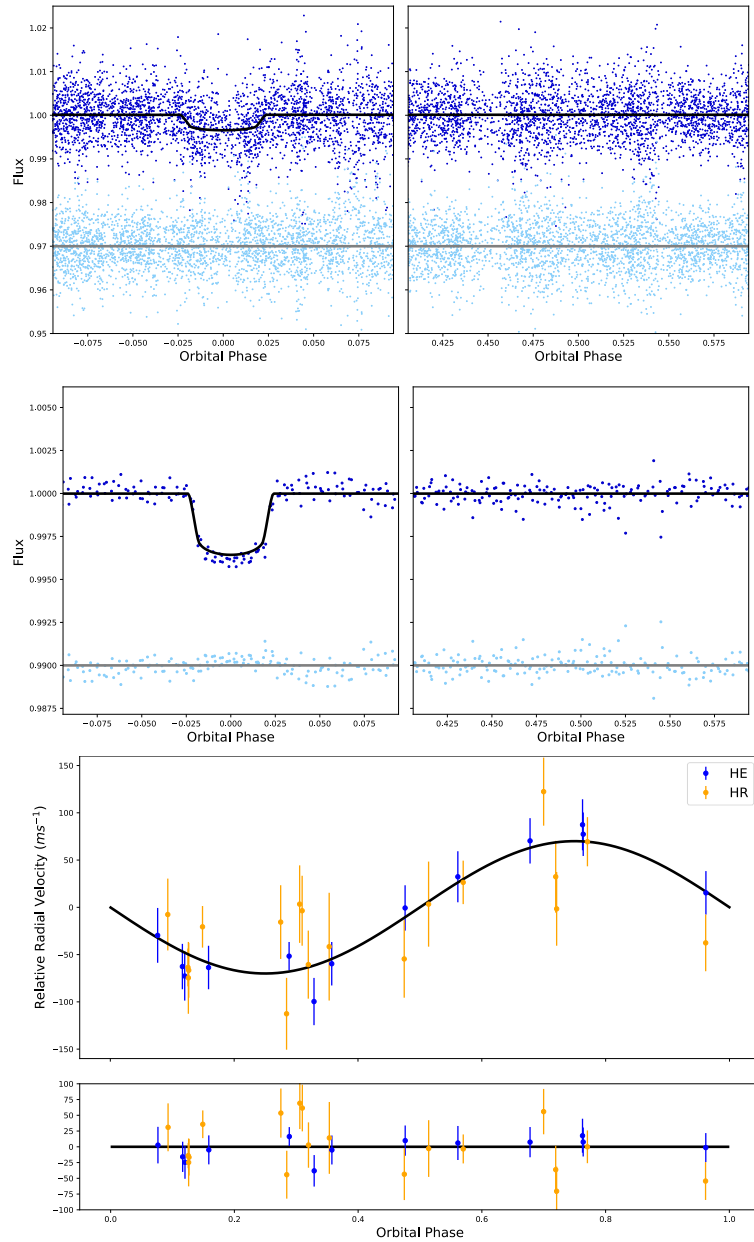
Figure 4.5: WASP (top), *TESS* (middle), and SOPHIE (bottom) data for WASP-187b phase folded on the best-fitting period. Residuals to the fit are shown below the data.

Table 4.2: System parameters for WASP-186 and WASP-187

Parameters	Symbol (Unit)	WASP-186	WASP-187
Stellar Parameters			
WASP ID		J011558.85+213700.9	J010953.96+254054.0
TESS ID		TOI-1494.01	TOI-1493.01
		TIC-411608801	TIC-15692883
Gaia ID		2790691147020786816	306410392895767680
Right Ascension	RA (hh:mm:ss)	01:15:58.85	01:09:53.96
Declination	Dec (dd:mm:ss)	+21:37:00.9	+25:40:54.0
Visual Magnitude	Vmag (mag)	10.82	10.30
TESS Magnitude	Tmag (mag)	10.30	9.71
Gaia Magnitude	Gmag (mag)	10.65	10.13
Stellar Mass	M_* $(M_\odot)$	$1.22^{+0.07}_{-0.08}$	1.54 ± 0.09
Stellar Radius	R_* $(R_\odot)$	1.47 ± 0.02	2.83 ± 0.05
Effective Temperature	$T_{\rm eff}$ (K)	6361^{+105}_{-82}	6150^{+92}_{-85}
Parallax	ϖ (mas)	$3.571^{+0.044}_{-0.042}$	$2.663^{+0.046}_{-0.043}$
Stellar Density	ρ_s $(\rho_\odot)$	$0.387^{+0.028}_{-0.027}$	0.068 ± 0.005
Surface Gravity	log g (cgs)	$4.193^{+0.028}_{-0.029}$	3.722 ± 0.029
Received Flux	f*1e-9 (cgs)	$1.266^{+0.007}_{-0.006}$	2.055 ± 0.008
Extinction	$E(B{-}V)$ (mag)	0.03 ± 0.02	0.09 ± 0.02
Planet Parameters			
Period	P (d)	$5.026799^{+.000012}_{-.000014}$	$5.147878^{+.000005}_{-.000009}$
Transit Epoch	Tc-2450000	6237.1195 ± 0.0009	$5197.3529^{+0.002}_{-0.0022}$
Transit Width	w (hr)	$2.704^{+0.048}_{-0.051}$	$5.82^{+0.095}_{-0.091}$
Transit Depth	$(R_P/R_S)^2$	0.0061 ± 0.0003	0.0035 ± 0.0002
Planet Mass	M_p (M_{Jup})	4.22 ± 0.18	0.8 ± 0.09
Planet Radius	R_p (R_{Jup})	1.11 ± 0.03	1.64 ± 0.05
Semi-major Axis	a (au)	$0.06^{+0.0012}_{-0.0013}$	0.0653 ± 0.0013
Impact Parameter	b	$0.84^{+0.01}_{-0.02}$	0.76 ± 0.02
Orbital Eccentricity	e	0.33 ± 0.01	0 (Fixed)
Argument of Periastron	ω	$3.02^{+0.05}_{-0.06}$	0 (Fixed)
Planet Density	ρ_p (ρ_{Jup})	$2.881^{+0.3}_{-0.279}$	$0.169^{+0.026}_{-0.023}$
Surface Gravity	log g (cgs)	3.93 ± 0.03	$2.87^{+0.05}_{-0.06}$
RV Semi-amplitude	K $(m\ s^{-1})$	$530.29^{+8.43}_{-8.59}$	$70.0^{+6.84}_{-7.75}$
0-point Uncertainty	log σ_{sys}	$-2.19^{+0.41}_{-0.34}$	$-2.16^{+0.32}_{-0.25}$
RV offset	γ $(m\ s^{-1})$	$-5866.2^{+5.7}_{-5.6}$	$-20275.3^{+6.9}_{-6.8}$ (HE)
			$-20220.4^{+9.2}_{-9.3}$ (HR)
RV Jitter	$\sigma_{\rm jit}$	$0.8^{+7.1}_{-7.3}$	$3.6^{+4.8}_{-2.8}$ (HE)
			$17.0^{+12.7}_{-8.7}$ (HR)
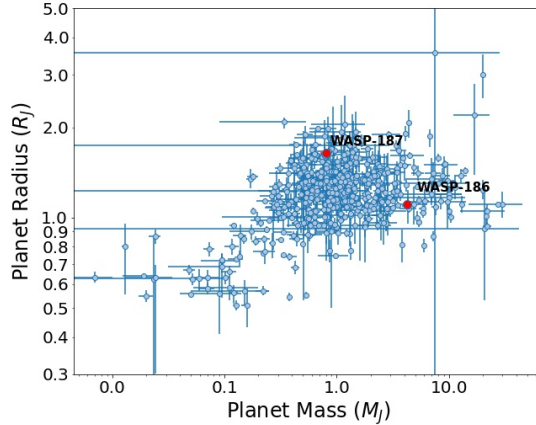
Figure 4.6: Planet mass versus planet radius for all Jupiter-sized planets ($R_P > 0.5R_J$) with mass and radius measurements. Data for this and subsequent plots was obtained from the NASA Exoplanet Archive http://exoplanetarchive.ipac.caltech.edu.

Jupiter ($1.11 \pm 0.03 \, R_{\mathrm{Jup}}$), but is quite massive at $4.22 \pm 0.18 \, M_{\mathrm{Jup}}$. WASP-186b therefore fits among the most massive and dense hot Jupiters known (See Fig. 4.7). WASP-186b is also notable as the orbit has an eccentricity of 0.33 ± 0.01, pointing to late-time high-eccentricity migration, rather than disc migration (Rasio & Ford, 1996; Ford & Rasio, 2008). Using equation 1 of Dobbs-Dixon et al. (2004) and an estimate for Q'_p of 10^6 in line with the estimation from Yoder & Peale (1981), the time scale of eccentricity damping via tidal disturbance is on the order of 15.7 Gyr, well above the estimated age of $3.1^{+1.0}_{-0.8}$ Gyr, suggesting that the system is still undergoing circularization. Finally, the planet has an equilibrium temperature (T_{eq}) of 1348^{+23}_{-22} K, assuming zero albedo and isotropic re-radiation.

The host star for WASP-187b has begun to evolve away from the main sequence (see Fig. 4.8), indicated by the effective temperature of 6150^{+92}_{-85} K, mass of $1.54 \pm 0.09 \, M_{\odot}$, and radius of $2.83 \pm 0.05 \, R_{\odot}$. The star is $375.52 \pm .45$ pc away (Bailer-Jones et al., 2018) and has a projected rotation is similar to that of WASP-186 at 15.3 $\pm 1.0 \, \mathrm{km \, s^{-1}}$, indicating the rotation has slowed since leaving the main sequence (Wolff & Simon, 1997). WASP-187b is hotter ($T_{eq} = 1726^{+31}_{-29}$ K) and significantly less dense than WASP-186b with a mass of $0.80 \pm 0.09 \, M_{\mathrm{Jup}}$ and a radius of $1.64 \pm 0.05 \, R_{\mathrm{Jup}}$, suggesting that this planet could be undergoing re-inflation (Hartman et al., 2016). Both planets have a period of around 5 days, but their radii lie near the low and high radius boundaries for known hot Jupiters of that period (see Fig. 4.9).
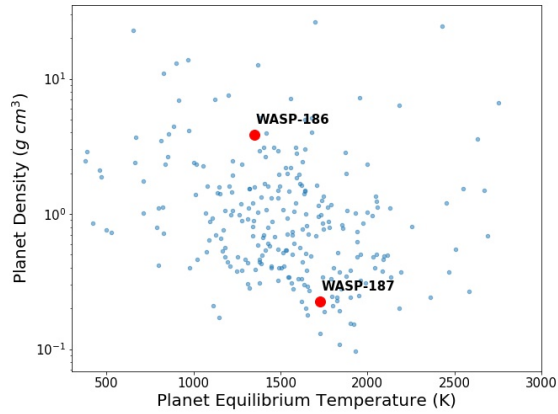
Figure 4.7: Estimated planet equilibrium temperature assuming zero albedo and isotropic re-radiation vs planet density. Note that KELT-9 is not shown because of the high equilibrium temperature (4,050 K).
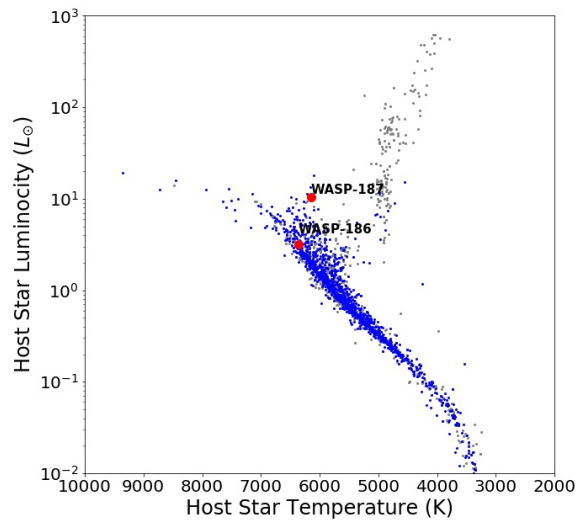
Figure 4.8: H-R diagram showing temperature versus stellar luminosity for all stars known to host exoplanets. Stars hosting planets with periods less than 10 days are shown in blue, while stars with planets with longer periods are in gray.
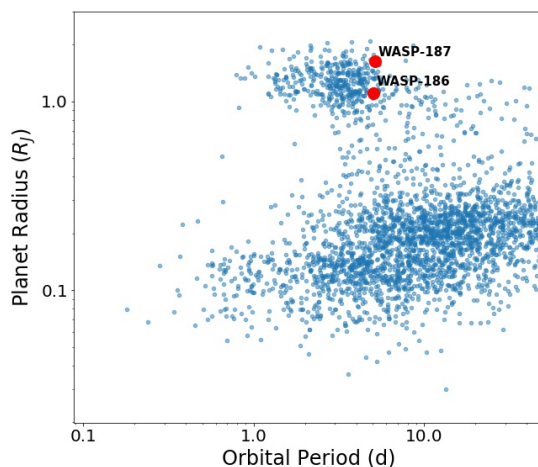
Figure 4.9: Orbital period versus planet radius for all known exoplanets with a period less than 50 days.

4.3 Future Candidate Validation

The model for WASP and other ground-based observatories until recently has been to identify candidates and pass them on to partner observatories to schedule observations when observing time allowed. However, the recent launches of *Gaia* and *TESS* have changed the validation process substantially.

Launched in 2013, *Gaia*'s mission is to produce the largest and most accurate map of the galaxy by monitoring the motion of around one billion stars (Gaia Collaboration et al., 2016). The usefulness of the resulting stellar catalog extends to all branches of astronomy. For exoplanet validation specifically, the parallax measurements of the stars can be used to find stellar radii estimates that greatly expand the catalog to new stars and improve upon estimates made with previous surveys. Historically around 6 per cent of candidates followed-up upon by WASP (43 of the 568 candidates dispositioned by SOPHIE and 60 of 1234 candidates dispositioned by CORALIE, Schanche et al. (2019b)) are discovered to have host stars in the giant regime making the transitors' radii super-planetary. Access to the *Gaia* data nearly eliminates this form of false positive. In the SuperWASP sample presented in Chapter 3, updating the WASP dataset with radii estimates from *Gaia* when available and recalculating the candidate planet's radius assuming this new value pushed 27 359 objects above a threshold of $2R_J$, quickly eliminating them from the candidate pool without need for additional observations.

The launch of *TESS* in 2018 is having an even more profound effect on the future of candidate vetting. During the initial 2-year mission, *TESS* is surveying the majority of the sky in strips, starting in the southern hemisphere for the first year followed by

the northern hemisphere for the second year (Ricker et al., 2015b). Each strip of sky is observed for 27 days, with stars near the ecliptic poles observed in multiple sectors extending the baseline coverage in some regions to over 300 days. However, mid-way through each sector's observations when the spacecraft is at perigee, there is a small gap in observations reducing the total time coverage slightly. *TESS* has a Candidate Target List (CTL) of 200,000 objects for which data is available at a 2-minute cadence. However, the entire full frame image (FFI) containing many more stars is available at a 30-minute cadence. Access to the data is open source, which provides a rich source of information for candidate vetting.

The vast majority of WASP candidates will be observed in at least one *TESS* sector. Once the observations for the sector containing the target are complete, the stack of FFIs can be downloaded and the region surrounding the star of interest extracted. As WASP candidates have a maximum period of 16 days due to the BLS search space, most astrophysical transits should be seen at least twice for each sector in which it is visible. This means that data for a large number of candidates can be quickly downloaded and lightcurves extracted to be checked for confirmation of transits. If no transit is seen at all, the candidate can be rejected as a likely systematic error in WASP. In many cases, there is a transit but the higher resolution data shown that the depth, shape, or deep secondary is not consistent with a planetary source. Similar to WASP, *TESS* has a large pixel size of 21", so there is a potential for blends. However, it is possible to check for any centroid shifts to indicate the source of the transit. This information can be combined with *Gaia* measurements of nearby stars to see whether the blending scenario is likely.

The wealth of new data has a huge impact, as now dozens of candidates can be checked for transits in a matter of hours without the need of any additional follow-up on the part of the vetter. Once the initial triage is complete, the remaining candidates that pass *TESS* inspection can be prioritized for RV follow up and, if necessary, further photometric observation.

To demonstrate this point, we use *TESS* data to check the validity of the top 12 candidates found in Chap 3. Of the 12, two candidates (J1648+32 and J1830+31) have not yet been observed but are in upcoming sectors and one candidate (J1454+14) is not currently in any scheduled sector. In addition, one candidate (J0937+35) was located on a strip of bad pixels and therefore no useful data could be extracted. Of the remaining 8 candidates, shown in Figures 4.10 and 4.11, three (J2308+47, J0734+14, and J2102+55) were clearly eclipsing binaries blended with nearby stars, as evidenced by their deep transits and pronounced secondaries. J2230+41, J1755+53, J1625+53, and J1445+30 show no strong signal in the target or surrounding stars, and are therefore likely to be due to

systematic issues in the WASP data. Only J0027+20 shows a possible small transit-like dip, although the depth is below the detection limit for WASP and the period does not match that found by WASP.

While many of the highest-ranked WASP candidates are found to be false positives, the process highlights the important roll that *TESS* now plays in validating targets identified by ground surveys. The threshold value used to prioritize candidates from ML ranking can be lowered as riskier targets (those with smaller transit depths, fainter host stars, or suspicious periods) can be checked without risking the limited observing time on other instruments.

Of course, *TESS* is also conducting searches through its archive to identify new planet candidates (See Crossfield et al. (2018) for full details on the vetting process). Similar to previous surveys, the data is automatically processed and further analysis and vetting plots are created for all Threshold Crossing Events (TCEs). The TCEs are all reviewed by a group of individuals who can label the object as a planet candidate or a selection of other false positives. When multiple human vetters flag the object as a planet candidate, it enters a group vetting round where the object is discussed in more detail. Objects that are still likely candidates at this stage are disseminated to the community as TOIs.

Many objects that are found by searching the WASP archive inevitably will have already been identified as TOIs, as evidenced by WASP-186 and WASP-187. However, this is not a deterrent to continuing to use archival WASP data. As was shown by Hellier et al. (2019), WASP and *TESS* have a symbiotic relationship, where the long baseline coverage achieved with the inclusion of WASP data can substantially improve the accuracy of the transit ephemeris from that found by *TESS* alone. This information can then be passed to the *TESS* team to improve the transit timing forecasts for detailed observations with other instruments such as the *CHEOPS* satellite (Broeg et al., 2013) and the upcoming James Webb Space Telescope, for which precise timings are essential for optimized scheduling.
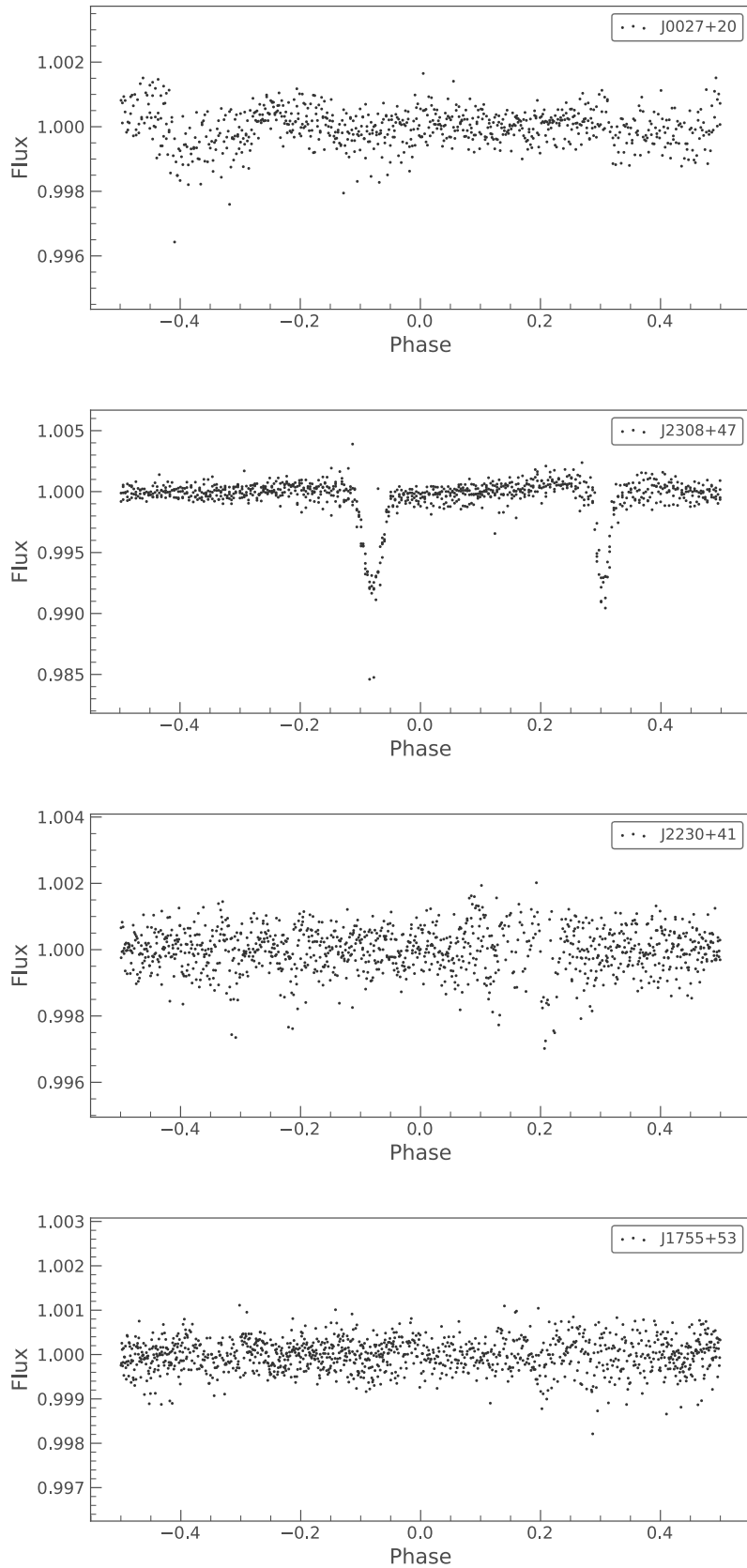
Figure 4.10: The *TESS* lightcurves top candidates found by the stacked ML model. The data were downloaded and the lightcurves were extracted and long term trends were removed using the functions provided in the LIGHTKURVE package, and then folded on the best-fitting period by Astropy's BOXLEASTSQUARES routine.
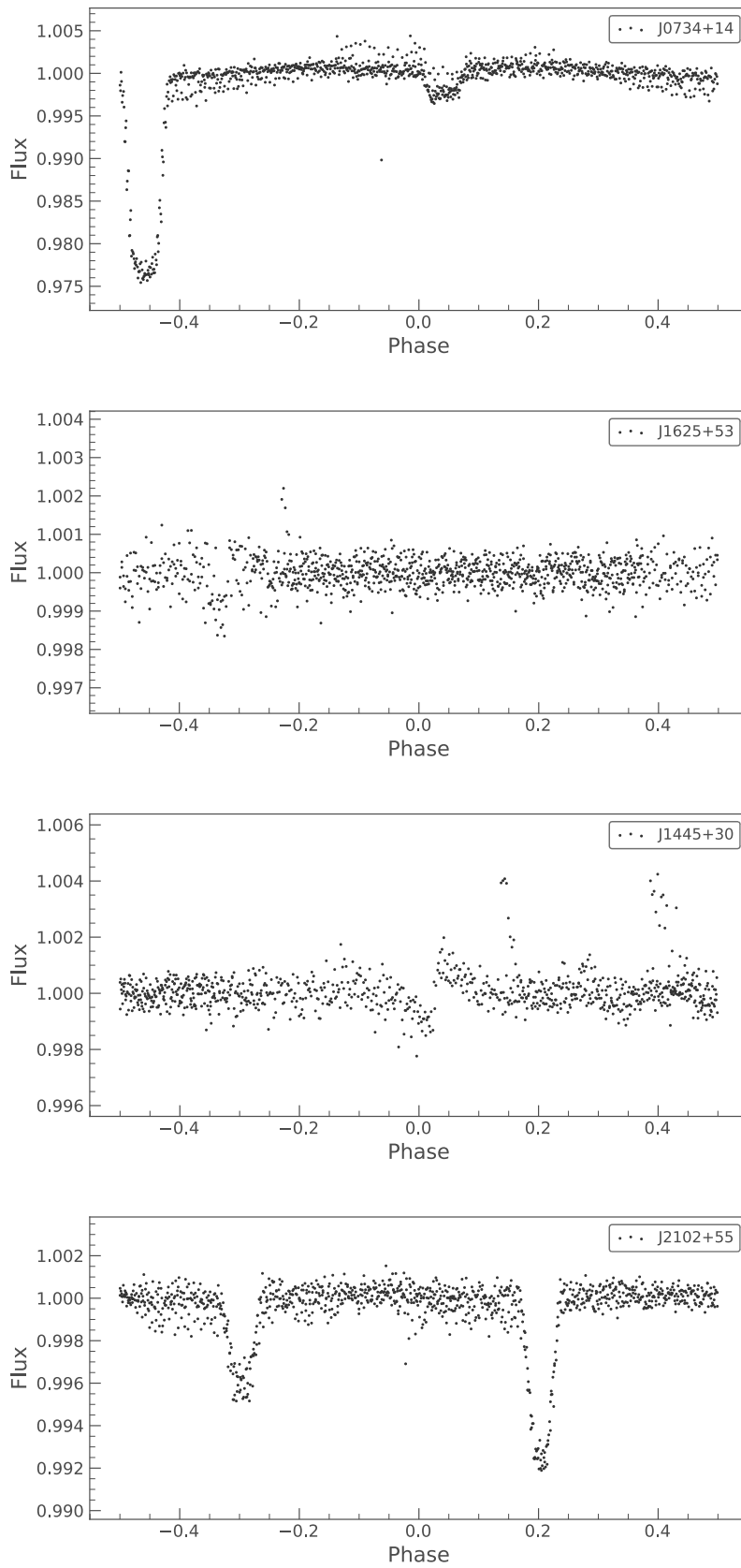
Figure 4.11: 4.10, cont.

94

5

Conclusions

And now the conclusion.

- Majel Barrett, *Every two part Star Trek episode*

5.1 Summary

This thesis outlined a model of exoplanet candidate selection beginning with target identi-fication and ranking using a machine learning approach to validation and characterization of planets using a novel MCMC approach, as applied to the discovery of the new hot Jupiters WASP-186b and WASP-187b. While the method used in this thesis was devel-oped specifically for WASP data, the principles are applicable to any transit search, either ground or space based with modifications to account for the data and errors specific to each survey.

In Chapter 2 we discussed the WASP project and the dataset that is available to train the machine learning models. As part of the development of the training data, we standardized the labelling of all of the SuperWASP false positives that were discovered after follow-up observations were taken. This catalog is now publicly available on CDS[1] so that other planet searches can avoid spending follow-up time on the same objects. The

---

[1]https://cds.u-strasbg.fr/

information can also be used by other surveys to add new pre-vetted datapoints to their own ML datasets.

As a preliminary application of ML, we developed two supervised models - a Random Forest Classifier and a Convolutional Neural Network. The RFC uses 34 features that describe the target star and the candidate system. The CNN takes as input binned lightcurve data, with two different variants: one only containing the full lightcurve, while the second also contains the "local" view to better capture the transit shape. In both cases, the input data are reliant on the results of the BLS algorithm implemented as part of the WASP pipeline.

The results of the two methods are discussed in Chapter 3. In particular, while in the process of manually inspecting the candidates, certain patterns began to emerge where the RFC and CNNs performed very differently. In many of these cases, the difference in the ratings was indicative of an underlying trend in the data. This realization sparked the development of a stacked ML model, in which the predictions and classifications of an ensemble of ML models are used as input to a second-level model which produces the final candidate probabilities.

In addition to applying three new ML models (Support Vector Machine, K-Nearest Neighbor, and Logistic Regression), we also made changes to the training data for the RFC and CNN. The training data for the initial implementation used only known planets for the positive "P" class. The relative low number of planets resulted in a poor performance, so artificial data points were required either through the SMOTE technique (RFC) or through artificial injection of transit signals into real WASP data (CNN). While this improved the performance of both methods, it is not clear how the artificial data may have affected the algorithm's performance when applied to new objects. Instead of relying on artificial data, we nearly doubled the members of the planet class for the stacked models by adding the known low-mass stars and brown dwarfs to the "P" class as their lightcurves are nearly identical to those of planets. We also modified the CNN to take the 1-dimensional BLS periodogram, binned in log space, into account in addition to the local and full lightcurves.

The output probabilities (CNN, RFC, and LR) and predictions (SVC, KNN) were used as the input data for the second-level logistic regression model. As we demonstrated with the resulting ROC curve and AUC values, the stacked model outperforms each constituent component. As a specific example, when applied to the testing set, if we require that 90% of planets are recovered, ~ 5 times the number of planets would be false positives. In other words, an observing program that aimed to have a 90% recovery rate would be expected to have an overall success rate of $\sim 20\%$ for follow-up observations. We applied the stacked model to over 100,000 unlabelled SuperWASP lightcurves to identify new candidates. As

a demonstration, we manually inspected the 12 highest-ranked candidates. While all of the candidates have reasonable transit-like shapes, many could be rejected through other means. Nevertheless, this demonstrates the utility of the method in reducing the total number of lightcurves for which labour-intensive manual checks must be done.

Finally, we discussed the steps that are taken once a planet is confirmed. We developed an MCMC method that fits photometric and radial velocity data simultaneously, while also incorporating prior information from various sources. This includes the use of optical/IR spectrophotmetry to get the angular diameter and the parallax measurements from *Gaia*, to constrain the stellar radius.

We applied this model to two new hot Jupiter systems. The eccentric (e∼0.33) orbit of WASP-186b and circular orbit of WASP-187b both have periods just above 5 days, but the planets' structural compositions are quite different. WASP-186b is a very dense object with a mass 4.2 times that of Jupiter and radius only 1.1 times larger. WASP-187b in contrast is very puffy, with a radius of 1.6 Jupiter radii and a mass only 0.8 times the mass of Jupiter. WASP-187b notably orbits around a star currently evolving off of the main sequence, possibly influencing its puffed-up nature.

Finally, we discussed how the current model of follow-up observations of ground-based targets is undergoing rapid changes. Open-access data from new space-based surveys can be used to quickly confirm or reject large numbers of candidates. This is reducing the amount of telescope time needed to follow up on the remaining targets, particularly with photometry. While ground surveys like WASP soon will no longer play a primary role in planet discovery, the archival data are extremely useful in establishing the epoch and periods of new planets.

5.2 Additional Follow-up Observations

One direct outcome of this work is the identification of a number of new planet candidates. The initial RFC/CNN model produced 72 candidates, a subset of which were pursued using photometric observations of Las Cumbres Observatory (LCO).

Three planets have already been validated, two by other surveys (HATNet and HAT-South) and the third is WASP-187b, presented in Chapter 4. One further object was found to have a plausible radius in follow-up and is currently waiting for RV confirmation of the mass. Of the remaining targets for which LCO data was obtained, 12 turned out to be caused by blending, 4 were eclipsing binaries, 4 were found to be low mass eclipsing binaries (3 from masses found by RV follow-up and one by the refined radius measure of $3R_J$) and for 6 no transit was visible at all in the LCO data meaning either the detection

by WASP was false or the transit ephemeris estimate has drifted and the observation scheduled missed the transit event.

Since *TESS* began observing the northern hemisphere, targets can now be assessed using lightcurves derived from *TESS* FFIs. Of the remaining ML candidates for which *TESS* data was available, 3 targets had the WASP period and depth confirmed by *TESS* and are awaiting RV observations, 4 were found to be blends, 7 showed transit depths and/or deep secondaries inconsistent with a planet signal and are therefore classed as EBs, and 10 were not detected by *TESS* and were therefore likely caused by systematics in the WASP data.

If we make the optimistic assumption that all of the objects that were supported photometrically will have masses that also put them in the planetary regime, the follow-up success rate of the initial ML candidate pool is just short of 13%. While this does not reach the success rates predicted by the RFC and CNN, it does exceed the overall success rate of 5% shown by Schanche et al. (2019a) for the SuperWASP project. We suspect that the drop in performance from that projected by the ML methods is reflective of the fact that the training and test datasets are made up of the lightcurves that have already been identified in the WASP archive, which tend to be the better candidates with stronger transit signals. The unclassified objects that make up the new candidates are less obvious and more difficult to find, either because they have shallow transits, fainter host stars, or red noise that was not removed sufficiently during detrending.

5.3 Machine Learning Improvements

While it is not possible to eliminate all false positives, it is still possible to further improve the performance of the classifier. The first most obvious way to improve a ML model is to have better and more informative data. The feature-based data used by the RFC, LR, KNN, and SVC in particular could be improved by adding new information that aids classification.

One possibility would be to incorporate the impact parameter, which for WASP is found by the MCMC fit. If a transit is found to be very short for the determined stellar density, the impact parameter needs to be high to account for the duration. For a planet, the probability of a transit is $\frac{R_*}{a}$, while the probability for a grazing transit is $\frac{2R_p}{a}$. It is much less likely for a planet to be observed at a grazing angle, so objects with a high impact parameter would be suggestive of an alternative model. This could be a useful feature for space based searches looking for small planets, where the chances of observing a grazing transit are much smaller. Another possibility would be to update the features using data from the *TESS* FFIs when available. This would help to eliminate errors due

to WASP systematics, as *TESS* would not observe any periodic dips at the period found by WASP.

While it is currently uncertain what the overall performance of the stacked model will be as the resulting targets have not been followed up upon, based on the ROC curves we expect it to improve upon the initial RFC and CNN methods. We can neverthe-less speculate on several ways by which the stacked model could be further improved. First, more individual models could be incorporated into the first level of the architecture. We implemented 5 models in this work, but there is no programmatic reason why more models couldn't be created. As stacking is more effective when the constituent models have more variation, it would be desirable to implement new methods that process the data in different ways but have been shown to be effective in transit analysis, such as Locality Preserving Projections (Thompson et al., 2015) or the unsupervised method of Self-Organizing Maps (Armstrong et al., 2017, 2018).

In addition, it would be interesting to explore not just different machine learning models, but also different transit-finding methods. Several studies show that different detection algorithms can have different performances in transit retrievals (e.g. Tingley, 2003; Moutou et al., 2013; Enoch et al., 2012). The first level models could be developed to run using the results of different transit finding models as input, and incorporate the predictions based on these differing results to make a final prediction.

5.4 Conclusions

From the work done in association with this thesis, two broad concluding points can be drawn that address the use of machine learning in the search for new planets going forward.

First, while stacking makes marked improvements over the individual models, there is a limit to the performance. The lightcurves of WASP and other ground-based surveys contain variable gaps in data and have errors significantly larger than the transit itself. Subsequent transit detection relies on the application of some other algorithm to find the signal, making the ML performance limited by the chosen transit detection algorithm's performance. In turn, the transit detection performance is limited by the ability to prop-erly de-trend the data, especially in terms of removing red noise. For these reasons, we conclude that while modest improvements could be made to the final classifier, they are unlikely to reveal a substantial number of new, high-quality WASP candidates. However, several candidates have been identified using ML in this study that are awaiting further observations.

Second, we recommend stacking as a technique going forward into new surveys. Large

space-based missions like TESS and *Plato* could use their large community of scientists to create the first level models. As long as all participants are given the same training, validation, and test objects and standardized output classes, different individuals or groups could create their own ML model of choice and submit the final architecture for use in a master second level model to make final predictions that are distributed to the community. This framework also has strong potential in the context of "Hack-Weeks" which have become a popular way to get participants together to focus on solving an existing computing problem in astronomy.

A

Appendix A

Table A.1: Features of the top 12 candidates found in SuperWASP data. A description of the features used in training can be found in Table 3.1. The results of the first level RF, LR, and CNN contain four columns, one giving the probability the object is in each class, whereas the KNN and SVC only return a single prediction value. The probabilities marked with an * show the final probabilites from the stacked model.

SWASP ID	Period	ntrans	$\Delta\chi^2$	sn_red	sn_ellipse	depth
J0937+35	1.496	21	-211.0	-10.7	1.5	-0.006
J1648+32	5.385	32	-133.0	-10.5	0.1	-0.004
J0027+20	6.980	14	-353.2	-18.3	3.4	-0.008
J1830+31	8.298	16	-340.7	-7.5	0.5	-0.006
J2308+47	2.946	12	-213.1	0.0	3.1	-0.006
J2230+41	7.639	12	-87.9	-9.1	1.1	-0.005
J1755+53	9.035	21	-334.4	-8.0	1.8	-0.004
J0734+14	2.676	14	-531.2	-11.7	1.1	-0.016
J1454+14	4.986	30	-263.1	-9.6	2.4	-0.008
J1625+53	11.476	11	-160.2	-8.3	3.3	-0.003
J1445+30	1.497	18	-176.2	-13.2	0.7	-0.007
J2102+55	3.807	9	-459.5	0.0	5.2	-0.009

SWASP ID	width	vmag	teff_jh	npts_intrans	clump_idx	sde
J0937+35	0.050	11.5	6078	298.0	0.2	18.64
J1648+32	0.014	12.0	5890	511.0	0.2	24.81
J0027+20	0.006	10.5	5838	269.0	0.4	27.15
J1830+31	0.017	10.5	5632	199.0	0.1	17.27
J2308+47	0.039	10.5	6707	120.0	0.2	18.07
J2230+41	0.015	11.8	6113	157.0	0.2	18.19
J1755+53	0.027	10.7	6078	736.0	0.2	16.26
J0734+14	0.056	11.8	6218	169.0	0.2	12.57
J1454+14	0.017	10.8	6078	323.0	0.2	22.84
J1625+53	0.017	10.6	6069	365.0	0.2	23.17
J1445+30	0.034	11.1	5986	208.0	1.0	27.89
J2102+55	0.050	10.8	6043	159.0	0.2	13.32

SWASP ID	R_*	R_p	trans_ratio	near_int	skew	kurtosis
J0937+35	1.17	0.87	0.911	0.496	-0.17	1.540
J1648+32	1.10	0.69	0.491	0.385	-0.21	0.122
J0027+20	0.67	0.60	1.389	0.020	-0.56	1.405
J1830+31	1.13	0.87	0.581	0.298	-0.34	0.382
J2308+47	1.46	1.13	0.651	0.054	-0.29	0.220
J2230+41	1.79	1.30	0.564	0.361	-0.12	1.306
J1755+53	1.85	1.15	0.472	0.035	-0.38	0.127
J0734+14	1.52	1.91	0.356	0.324	0.11	0.564
J1454+14	1.14	1.00	1.147	0.014	-0.11	4.098
J1625+53	1.27	0.73	0.319	0.476	-0.60	0.327
J1445+30	1.09	0.89	0.381	0.497	-0.80	0.752
J2102+55	1.35	1.29	0.482	0.193	-0.13	0.081

SWASP ID	dilution_v	RF_ebprob	RF_plprob	RF_vprob	RF_xprob	KNN_pred
J0937+35	2.00	0.316	0.627	0.016	0.04	X
J1648+32	0.00	0.203	0.757	0.001	0.04	EB/Blend
J0027+20	0.00	0.554	0.249	0.020	0.18	EB/Blend
J1830+31	0.00	0.174	0.731	0.018	0.08	X
J2308+47	0.00	0.520	0.361	0.049	0.07	EB/Blend
J2230+41	8.00	0.153	0.550	0.005	0.29	P/EBLM
J1755+53	1.00	0.248	0.574	0.042	0.14	X
J0734+14	1.00	0.278	0.646	0.017	0.06	EB/Blend
J1454+14	4.00	0.293	0.522	0.024	0.16	EB/Blend
J1625+53	0.00	0.350	0.317	0.005	0.33	X
J1445+30	0.00	0.422	0.206	0.041	0.33	X
J2102+55	4.00	0.487	0.383	0.062	0.07	P/EBLM

SWASP ID	LR_ebprob	LR_plprob	LR_vprob	LR_xprob	SVC_pred	CNN_ebprob
J0937+35	0.065	0.833	0.095	0.007	P/EBLM	0.105
J1648+32	0.099	0.762	0.008	0.131	P/EBLM	0.167
J0027+20	0.026	0.966	0.002	0.006	P/EBLM	0.076
J1830+31	0.075	0.647	0.002	0.277	P/EBLM	0.097
J2308+47	0.087	0.836	0.026	0.052	P/EBLM	0.102
J2230+41	0.104	0.716	0.003	0.177	P/EBLM	0.136
J1755+53	0.059	0.197	0.001	0.743	X	0.042
J0734+14	0.308	0.389	0.091	0.212	P/EBLM	0.060
J1454+14	0.104	0.842	0.019	0.035	EB/Blend	0.102
J1625+53	0.070	0.467	0.000	0.463	X	0.016
J1445+30	0.123	0.780	0.062	0.035	P/EBLM	0.024
J2102+55	0.137	0.652	0.023	0.188	P/EBLM	0.053

Table 1, Cont.

SWASP ID	CNN_plprob	CNN_vprob	CNN_xprob	EB_prob*	Pl_prob*	V_prob*
J0937+35	0.848	0.003	0.011	0.040	0.959	0.000
J1648+32	0.511	0.000	0.007	0.056	0.942	0.000
J0027+20	0.883	0.000	0.002	0.064	0.930	0.000
J1830+31	0.350	0.001	0.084	0.078	0.916	0.001
J2308+47	0.916	0.000	0.001	0.082	0.914	0.000
J2230+41	0.546	0.000	0.019	0.076	0.911	0.001
J1755+53	0.949	0.000	0.001	0.085	0.906	0.001
J0734+14	0.919	0.000	0.001	0.093	0.905	0.000
J1454+14	0.454	0.000	0.002	0.096	0.895	0.001
J1625+53	0.887	0.000	0.004	0.075	0.894	0.001
J1445+30	0.965	0.000	0.000	0.085	0.893	0.001
J2102+55	0.814	0.000	0.003	0.123	0.869	0.001

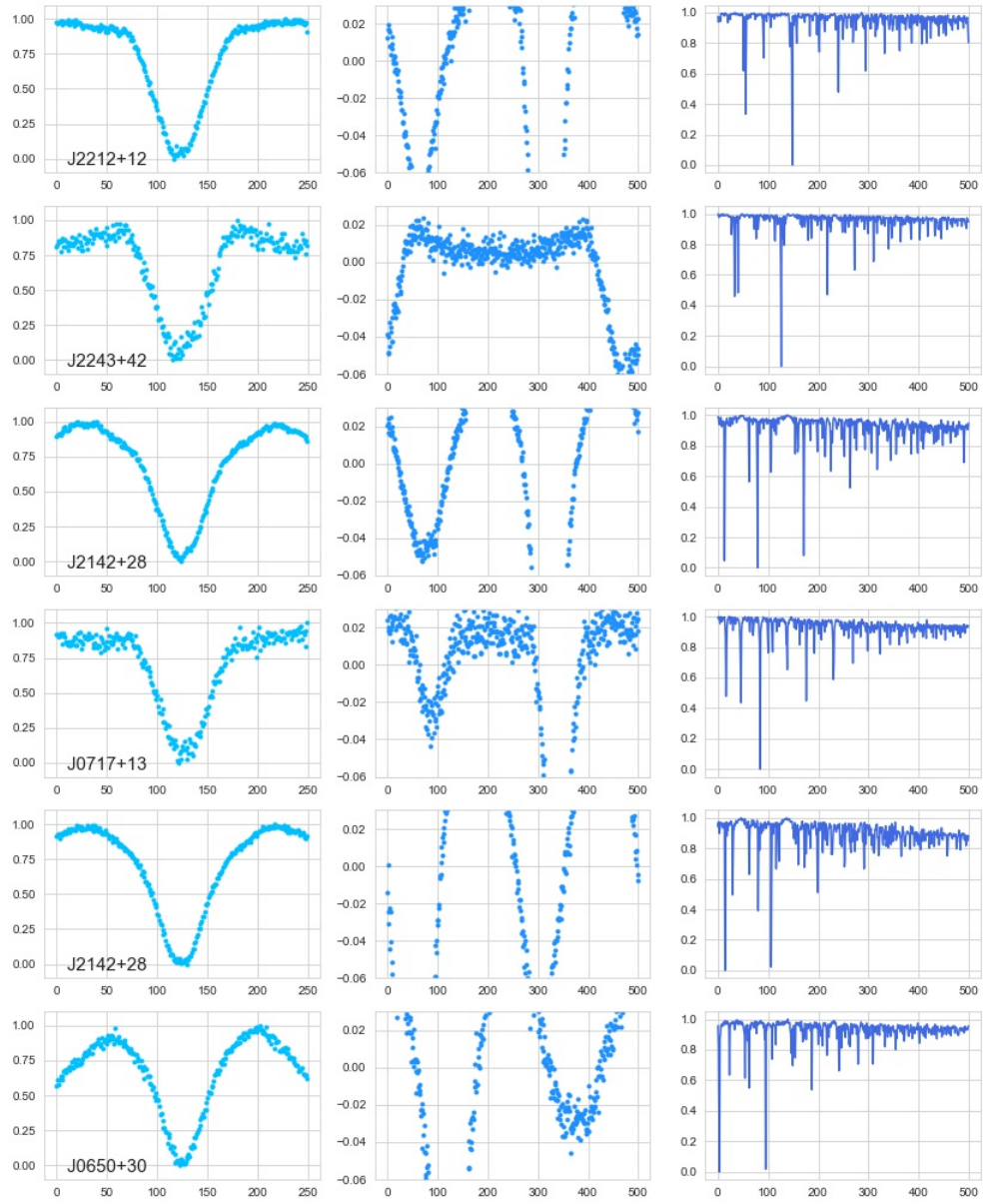| SWASP ID	X_prob*	Final Pred
J0937+35	0.001	P/EBLM
J1648+32	0.001	P/EBLM
J0027+20	0.006	P/EBLM
J1830+31	0.005	P/EBLM
J2308+47	0.003	P/EBLM
J2230+41	0.012	P/EBLM
J1755+53	0.009	P/EBLM
J0734+14	0.002	P/EBLM
J1454+14	0.009	P/EBLM
J1625+53	0.030	P/EBLM
J1445+30	0.021	P/EBLM
J2102+55	0.007	P/EBLM

B

Appendix B

Figure B.1: CNN input for the 12 lightcurves with the highest probabilities in the class containing eclipsing binaries and blended stars.
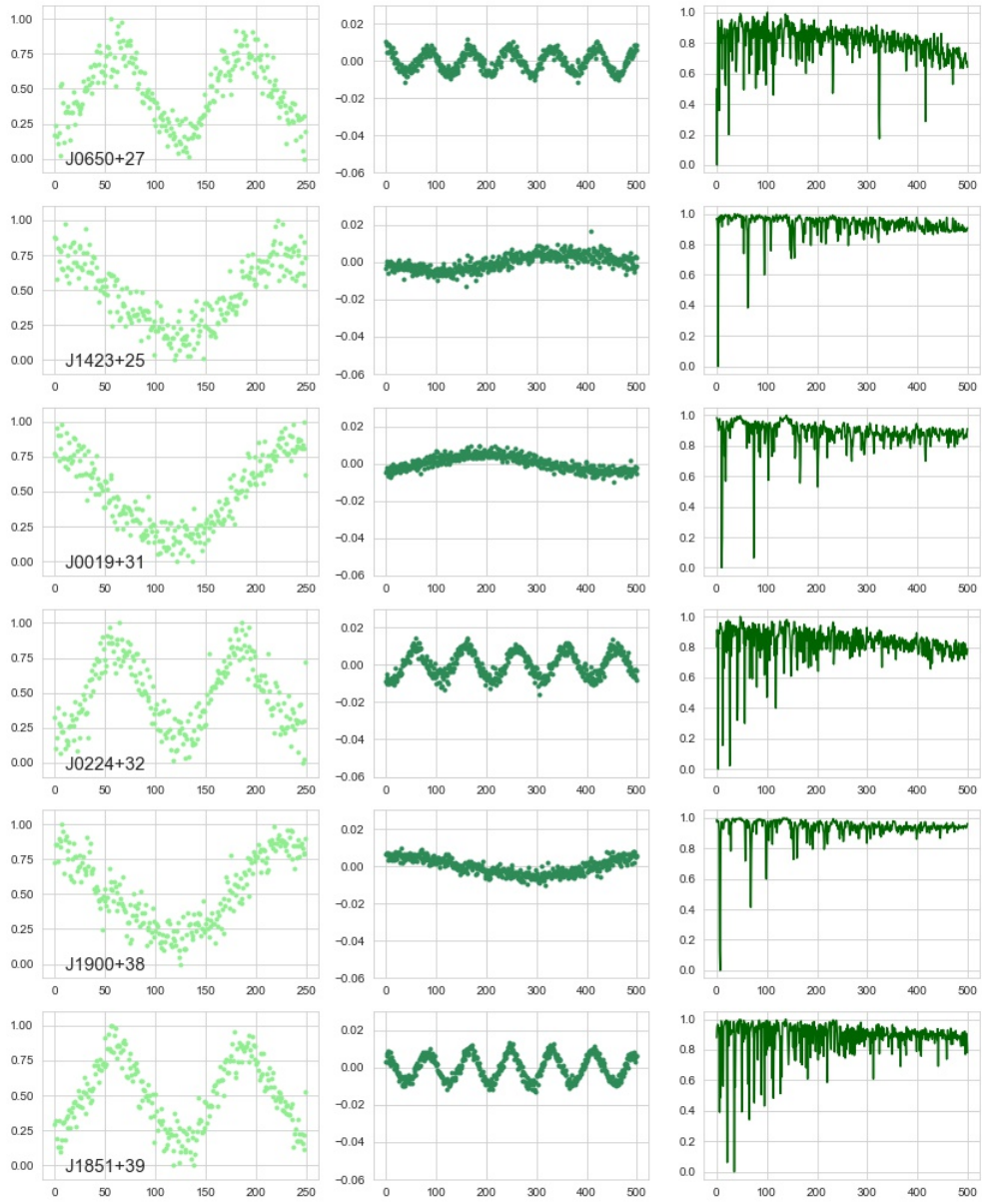
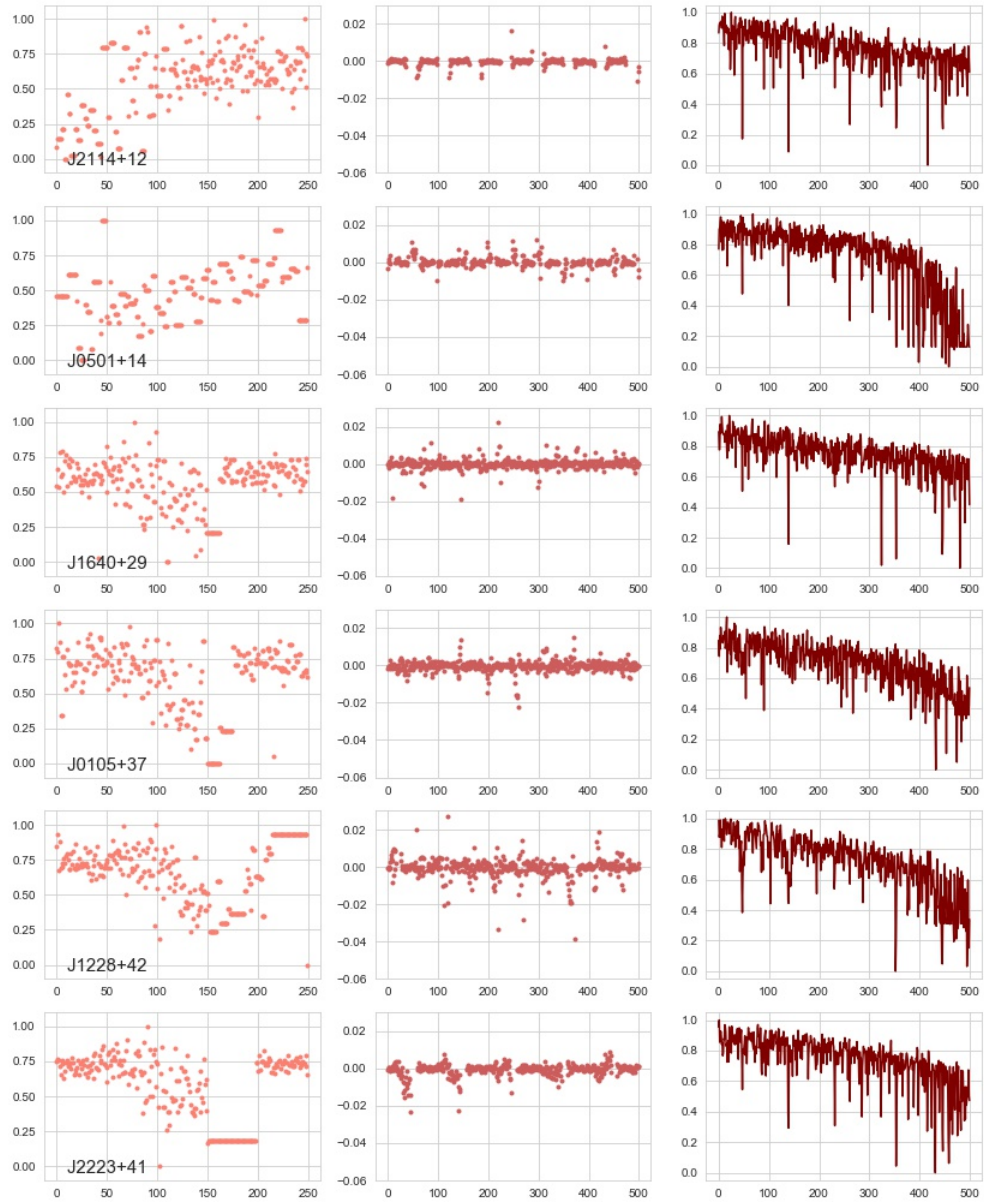**Figure B.2:** CNN input for the 12 lightcurves with the highest probabilities in the variable star class.

Figure B.3: CNN input for the 12 lightcurves with the highest probabilities in the X class.

Bibliography

Aggarwal, A., Schanche, N., Reeves, K. K., Kempton, D., & Angryk, R. 2018, The Astrophysical Journal Supplement Series, 236, 15

Aigrain, S., & Favata, F. 2002, Astronomy and Astrophysics, 395, 625, astro-ph/0209179

Alibert, Y., & Venturini, J. 2019, Astronomy and Astrophysics, 626, A21, 1903.00320

Alonso, A., Arribas, S., & Martinez-Roger, C. 1994, Astronomy and Astrophysics, 282, 684

Alsubai, K. A. et al. 2013, Acta Astronomica, 63, 465, 1401.1984

Altman, N. S. 1992, The American Statistician, 46, 175

Anderson, D. R. et al. 2011, The Astrophysical Journal Letters, 726, L19, 1010.3006

Andrae, R. et al. 2018, Astronomy and Astrophysics, 616, A8, 1804.09374

Ansdell, M. et al. 2018, The Astrophysical Journal Letters, 869, L7, 1810.13434

Armstrong, D. J. et al. 2018, MNRAS, 478, 4225, 1805.07089

Armstrong, D. J., Pollacco, D., & Santerne, A. 2017, MNRAS, 465, 2634, 1611.01968

Auvergne, M. et al. 2009, Astronomy and Astrophysics, 506, 411, 0901.2206

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, The Astronomical Journal, 156, 58, 1804.10121

Bakos, G. Á. et al. 2013, Publications of the Astronomical Society of the Pacific, 125, 154, 1206.1391

Bakos, G. Á. et al. 2011, The Astrophysical Journal, 742, 116, 1008.3388

Baraffe, I., Chabrier, G., Barman, T. S., Allard, F., & Hauschildt, P. H. 2003, Astronomy and Astrophysics, 402, 701, astro-ph/0302293

Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015, Astronomy and Astrophysics, 577, A42, 1503.04107

Barclay, T. et al. 2012, The Astrophysical Journal, 761, 53, 1210.4592

Barnes, G., Schanche, N., Leka, K. D., Aggarwal, A., & Reeves, K. 2017, in IAU Symposium, Vol. 325, Astroinformatics, ed. M. Brescia, S. G. Djorgovski, E. D. Feigelson, G. Longo, & S. Cavuoti, 201–204

Bayliss, D. et al. 2018a, MNRAS, 475, 4467, 1710.11099

Bayliss, D. et al. 2018b, The Astronomical Journal, 155, 119

Bayliss, D. D. R., & Sackett, P. D. 2011, The Astrophysical Journal, 743, 103, 1112.0359

Bentley, S. 2009, PhD thesis, Keele University

Blackwell, D. E., & Shallis, M. J. 1977, MNRAS, 180, 177

Boisse, I. et al. 2010, Astronomy and Astrophysics, 523, A88, 1006.4984

Bonfanti, A., & Gillon, M. 2020, Astronomy and Astrophysics, 635, A6, 1912.12632

Bordé, P., Fressin, F., Ollivier, M., Léger, A., & Rouan, D. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 366, Transiting Extrapolar Planets Workshop, ed. C. Afonso, D. Weldrake, & T. Henning, 145, 0709.3727

Borucki, W. J. et al. 2010, Science, 327, 977

Boser, B. E., Guyon, I. M., & Vapnik, V. N. 1992, in Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT 92 (New York, NY, USA: Association for Computing Machinery), 144152

Bouchy, F. et al. 2009, Astronomy and Astrophysics, 505, 853

Bouma, L. G., Masuda, K., & Winn, J. N. 2018, The Astronomical Journal, 155, 244, 1804.07764

Bozza, V., Mancini, L., & Sozzetti, A. 2016, Methods of Detecting Exoplanets : 1st Advanced School on Exoplanetary Science (Springer)

Breiman, L. 1996, Mach. Learn., 24, 49

Breiman, L. 2001, Mach. Learn., 45, 5

Broeg, C. et al. 2013, in European Physical Journal Web of Conferences, Vol. 47, European Physical Journal Web of Conferences, 03005, 1305.2270

Cabrera, J., Csizmadia, S., Erikson, A., Rauer, H., & Kirste, S. 2012, Astronomy and Astrophysics, 548, A44, 1211.6550

Cabrera-Vives, G., Reyes, I., Förster, F., Estévez, P. A., & Maureira, J.-C. 2017, The Astrophysical Journal, 836, 97, 1701.00458

Cáceres, C. et al. 2014, Astronomy and Astrophysics, 565, A7, 1403.2723

Caceres, G. A., Feigelson, E. D., Jogesh Babu, G., Bahamonde, N., Christen, A., Bertin, K., Meza, C., & Curé, M. 2019a, The Astronomical Journal, 158, 58, 1905.09852

Caceres, G. A., Feigelson, E. D., Jogesh Babu, G., Bahamonde, N., Christen, A., Bertin, K., Meza, C., & Curé, M. 2019b, The Astronomical Journal, 158, 57, 1901.05116

Campbell, B., Walker, G. A. H., & Yang, S. 1988, The Astrophysical Journal, 331, 902

Carrasco, D. et al. 2015, Astronomy and Astrophysics, 584, A44, 1405.5298

Casagrande, L., Ramírez, I., Meléndez, J., Bessell, M., & Asplund, M. 2010, Astronomy and Astrophysics, 512, A54, 1001.3142

Castelli, F., & Kurucz, R. L. 2003, in IAU Symposium, Vol. 210, Modelling of Stellar Atmospheres, ed. N. Piskunov, W. W. Weiss, & D. F. Gray, A20, astro-ph/0405087

Cecil, D., & Campbell-Brown, M. 2020, Planetary and Space Science, 186, 104920

Chabrier, G., & Baraffe, I. 2000, Annual Review of Astronomy and Astrophysics, 38, 337, astro-ph/0006383

Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, The Astrophysical Journal Letters, 529, L45, astro-ph/9911436

Chaushev, A. et al. 2019, MNRAS, 488, 5232, 1907.11109

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2002, J. Artif. Int. Res., 16, 321

Chintarungruangchai, P., & Jiang, I.-G. 2019, Publications of the Astronomical Society of the Pacific, 131, 064502, 1904.12419

Christian, D. J. et al. 2006, MNRAS, 372, 1117

Collier Cameron, A. 2012, Nature, 492, 48

Collier Cameron, A. et al. 2010, MNRAS, 407, 507, 1004.4551

Collier Cameron, A. et al. 2006, MNRAS, 373, 799, astro-ph/0609418

Collier Cameron, A. et al. 2007, MNRAS, 380, 1230, 0707.0417

Collier Cameron, A. C., & Jardine, M. 2018, MNRAS

Collins, K. A. et al. 2018, ArXiv e-prints, 1803.01869

Crossfield, I. J. M. et al. 2018, The Astrophysical Journal Supplement Series, 239, 5, 1806.03127

Cumming, A., Butler, R. P., Marcy, G. W., Vogt, S. S., Wright, J. T., & Fischer, D. A. 2008, Publications of the Astronomical Society of the Pacific, 120, 531, 0803.3357

David, T. J. et al. 2018, The Astronomical Journal, 155, 222, 1803.05056

Defaÿ, C., Deleuil, M., & Barge, P. 2001, Astronomy and Astrophysics, 365, 330

Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441, 1503.07077

Bibliography

D'Isanto, A., Cavuoti, S., Brescia, M., Donalek, C., Longo, G., Riccio, G., & Djorgovski, S. G. 2016, MNRAS, 457, 3119, 1601.03931

Dittmann, J. A. et al. 2017, Nature, 544, 333, 1704.05556

Dobbs-Dixon, I., Lin, D. N. C., & Mardling, R. A. 2004, The Astrophysical Journal, 610, 464, astro-ph/0408191

Domijan, K., Bloomfield, D. S., & Pitié, F. 2019, Solar Physics, 294, 6

Doyle, A. P. et al. 2013, MNRAS, 428, 3164, 1210.5931

Doyle, L. R. et al. 2011, Science, 333, 1602, 1109.3432

Doyle, L. R. et al. 2000, The Astrophysical Journal, 535, 338, astro-ph/0001177

du Buisson, L., Sivanandam, N., Bassett, B. A., & Smith, M. 2015, MNRAS, 454, 2026, 1407.4118

Dubath, P. et al. 2011, MNRAS, 414, 2602, 1101.2406

Enoch, B., Haswell, C. A., Norton, A. J., Collier-Cameron, A., West, R. G., Smith, A. M. S., & Parley, N. R. 2012, Astronomy and Astrophysics, 548, A48

Evans, T. M., & Sackett, P. D. 2010, The Astrophysical Journal, 712, 38, 1002.0886

Fisher, C., Hoeijmakers, H. J., Kitzmann, D., Márquez-Neila, P., Grimm, S. L., Sznitman, R., & Heng, K. 2020, The Astronomical Journal, 159, 192, 1910.11627

Florios, K., Kontogiannis, I., Park, S.-H., Guerra, J. A., Benvenuto, F., Bloomfield, D. S., & Georgoulis, M. K. 2018, Solar Physics, 293, 28, 1801.05744

Ford, E. B. 2005, The Astronomical Journal, 129, 1706, astro-ph/0305441

Ford, E. B., & Rasio, F. A. 2008, The Astrophysical Journal, 686, 621, astro-ph/0703163

Fortney, J. J., Marley, M. S., & Barnes, J. W. 2007, The Astrophysical Journal, 659, 1661, astro-ph/0612671

Fukushima, K. 1980, Biol. Cybernetics, 36, 193

Gaia Collaboration, Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Babusiaux, C., & Bailer-Jones, C. A. L. 2018, ArXiv e-prints, 1804.09365

Gaia Collaboration et al. 2016, Astronomy and Astrophysics, 595, A1, 1609.04153

Gaudi, B. S. et al. 2017, Nature, 546, 514, 1706.06723

Gill, S. et al. 2019, Astronomy and Astrophysics, 626, A119, 1904.12695

Gillon, M. et al. 2017, Nature, 542, 456, 1703.01424

Glorot, X., Bordes, A., & Bengio, Y. 2010, in

Gómez Maqueo Chew, Y. et al. 2014, Astronomy and Astrophysics, 572, A50, 1408.6900

González Hernández, J. I., & Bonifacio, P. 2009, Astronomy and Astrophysics, 497, 497, 0901.3034

Gray, D. F. 1992

Günther, M. N. et al. 2017, Monthly Notices of the Royal Astronomical Society, 472, 295, https://academic.oup.com/mnras/article-pdf/472/1/295/19687274/stx1920.pdf

Gural, P. S. 2019, MNRAS, 489, 5109

Hartman, J. D., Bakos, G., Stanek, K. Z., & Noyes, R. W. 2004, The Astronomical Journal, 128, 1761, astro-ph/0405597

Hartman, J. D. et al. 2016, The Astronomical Journal, 152, 182, 1609.02767

He, K., Zhang, X., Ren, S., & Sun, J. 2015, in 2015 IEEE International Conference on Computer Vision (ICCV), 1026–1034

Hebb, L. et al. 2010, The Astrophysical Journal, 708, 224, 1001.0403

Hébrard, G. et al. 2013, Astronomy and Astrophysics, 549, A134, 1211.0810

Hellier, C. et al. 2011, Astronomy and Astrophysics, 535, L7, 1104.2823

Hellier, C., Anderson, D. R., Gill, S., & West, R. G. 2019, Research Notes of the American Astronomical Society, 3, 156, 1910.07276

Henry, G. W., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000, The Astrophysical Journal Letters, 529, L41

Hestroffer, D. 1997, Astronomy and Astrophysics, 327, 199

Hippke, M., & Heller, R. 2019, Astronomy and Astrophysics, 623, A39, 1901.02015

Hodžić, V. et al. 2018, MNRAS, 481, 5091, 1807.07557

Høg, E. et al. 2000, Astronomy and Astrophysics, 355, L27

Holman, M. J. et al. 2006, The Astrophysical Journal, 652, 1715, astro-ph/0607571

Horne, K. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 294, Scientific Frontiers in Research on Extrasolar Planets, ed. D. Deming & S. Seager, 361–370, astro-ph/0301249

Howard, A. W. et al. 2012, The Astrophysical Journal Supplement Series, 201, 15, 1103.2541

Howell, S. B. et al. 2014, Publications of the Astronomical Society of the Pacific, 126, 398, 1402.5163

Bibliography

Jenkins, J. M., Caldwell, D. A., & Borucki, W. J. 2002, The Astrophysical Journal, 564, 495

Jenkins, J. M. et al. 2010, The Astrophysical Journal Letters, 713, L87, 1001.0258

Jenkins, J. M., Doyle, L. R., & Cullers, D. K. 1996, Icarus, 119, 244

Jurkevich, I. 1971, Astrophysics and Space Science, 13, 154

Kim, E. J., & Brunner, R. J. 2017, MNRAS, 464, 4463, 1608.04369

Kingma, D. P., & Ba, J. 2014, ArXiv e-prints, 1412.6980

Kodikara, G. R. L., & McHenry, L. J. 2020, Icarus, 345, 113719

Konacki, M., Torres, G., Jha, S., & Sasselov, D. D. 2003, Nature, 421, 507

Kostov, V. B. et al. 2020, The Astronomical Journal, 159, 253, 2004.07783

Kovács, G., Zucker, S., & Mazeh, T. 2002, Astronomy and Astrophysics, 391, 369, astro-ph/0206099

Kreidberg, L. 2015, Publications of the Astronomical Society of the Pacific, 127, 1161, 1507.08285

Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, in Advances in Neural Information Processing Systems 25, ed. F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Curran Associates, Inc.), 1097–1105

Lafler, J., & Kinman, T. D. 1965, The Astrophysical Journal Supplement Series, 11, 216

Lam, K. W. F. et al. 2017, Astronomy and Astrophysics, 599, A3, 1607.07859

Latham, D. W., Mazeh, T., Stefanik, R. P., Mayor, M., & Burki, G. 1989, Nature, 339, 38

LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., & Jackel, L. D. 1990, in Advances in Neural Information Processing Systems 2, ed. D. S. Touretzky (Morgan-Kaufmann), 396–404

LeCun, Y. A., Bottou, L., Orr, G. B., & Müller, K.-R. 1998, Efficient BackProp, ed. G. Montavon, G. B. Orr, & K.-R. Müller (Berlin, Heidelberg: Springer Berlin Heidelberg)

Liu, C., Deng, N., Wang, J. T. L., & Wang, H. 2017, The Astrophysical Journal, 843, 104, 1706.02422

Lo, K. K., Farrell, S., Murphy, T., & Gaensler, B. M. 2014, The Astrophysical Journal, 786, 20, 1403.0188

Marcy, G., Butler, R. P., Fischer, D., Vogt, S., Wright, J. T., Tinney, C. G., & Jones, H. R. A. 2005, Progress of Theoretical Physics Supplement, 158, 24, https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.158.24/5251052/158-24.pdf

Marcy, G. W., & Benitz, K. J. 1989, The Astrophysical Journal, 344, 441

Márquez-Neila, P., Fisher, C., Sznitman, R., & Heng, K. 2018, Nature Astronomy, 2, 719, 1806.03944

Marton, G. et al. 2019, MNRAS, 487, 2522, 1905.03063

Masci, F. J., Hoffman, D. I., Grillmair, C. J., & Cutri, R. M. 2014, The Astronomical Journal, 148, 21, 1402.0125

Maxted, P. F. L. 2018, Astronomy and Astrophysics, 616, A39, 1804.07943

Mayor, M., & Queloz, D. 1995, Nature, 378, 355

McCauliff, S. D. et al. 2015, The Astrophysical Journal, 806, 6, 1408.1496

McCullough, P. R. et al. 2006, The Astrophysical Journal, 648, 1228, astro-ph/0605414

Ment, K. et al. 2019, The Astronomical Journal, 157, 32, 1808.00485

Mislis, D., Bachelet, E., Alsubai, K. A., Bramich, D. M., & Parley, N. 2016, MNRAS, 455, 626, 1511.03456

Monet, D. G. et al. 2003, The Astronomical Journal, 125, 984, astro-ph/0210694

Morello, G., Tsiaras, A., Howarth, I. D., & Homeier, D. 2017, The Astronomical Journal, 154, 111, 1704.08232

Morii, M. et al. 2016, Publications of the Astronomical Society of Japan, 1609.03249

Moutou, C. et al. 2013, Icarus, 226, 1625, 1306.0578

Nair, V., & Hinton, G. E. 2010, in Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML10 (Madison, WI, USA: Omnipress), 807814

O'Donovan, F. T. et al. 2006, The Astrophysical Journal, 644, 1237, astro-ph/0603005

Ollivier, M. et al. 2012, Astronomy and Astrophysics, 541, A149

Osborn, H. P. et al. 2020, Astronomy and Astrophysics, 633, A53, 1902.08544

Pearson, K. A., Palafox, L., & Griffith, C. A. 2018, MNRAS, 474, 478, 1706.04319

Pepper, J. et al. 2007, Publications of the Astronomical Society of the Pacific, 119, 923, 0704.0460

Perruchot, S. et al. 2011, in , Vol. 8151, Techniques and Instrumentation for Detection of Exoplanets V, 815115, 1110.2256

Perruchot, S. et al. 2008, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7014, The SOPHIE spectrograph: design and technical keypoints for high throughput and high stability, 70140J

Perryman, M. 2018, The Exoplanet Handbook, 2nd edn. (Cambridge University Press)

Petigura, E. A. et al. 2018, The Astronomical Journal, 155, 89, 1712.04042

Pollacco, D. L. et al. 2006, Publications of the Astronomical Society of the Pacific, 118, 1407, astro-ph/0608454

Pont, F., Zucker, S., & Queloz, D. 2006, MNRAS, 373, 231, astro-ph/0608597

Queloz, D. et al. 2000, Astronomy and Astrophysics, 354, 99

Ramírez, I., & Meléndez, J. 2005, The Astrophysical Journal, 626, 446, astro-ph/0503108

Rappaport, S. et al. 2012, The Astrophysical Journal, 752, 1, 1201.2662

Rasio, F. A., & Ford, E. B. 1996, Science, 274, 954

Régulo, C., Almenara, J. M., Alonso, R., Deeg, H., & Roca Cortés, T. 2007, Astronomy and Astrophysics, 467, 1345, 0705.4557

Reis, I., Baron, D., & Shahaf, S. 2019, The Astronomical Journal, 157, 16, 1811.05994

Revsbech, E. A., Trotta, R., & van Dyk, D. A. 2018, MNRAS, 473, 3969, 1706.03811

Ricker, G. R. et al. 2015a, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

Ricker, G. R. et al. 2015b, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

Riello, M. et al. 2018, Astronomy and Astrophysics, 616, A3, 1804.09367

Rowe, J. F. et al. 2014, The Astrophysical Journal, 784, 45, 1402.6534

Schaefer, C., Geiger, M., Kuntzer, T., & Kneib, J. P. 2018, Astronomy and Astrophysics, 611, A2, 1705.07132

Schanche, N. et al. 2019a, MNRAS, 488, 4905

Schanche, N. et al. 2019b, MNRAS, 483, 5534, 1811.07754

Schanche, N. et al. 2020, MNRAS

Schwarzenberg-Czerny, A., & Beaulieu, J. P. 2006, MNRAS, 365, 165, astro-ph/0509833

Seager, S., & Mallén-Ornelas, G. 2003, The Astrophysical Journal, 585, 1038, astro-ph/0206228

Shallue, C. J., & Vanderburg, A. 2018, The Astronomical Journal, 155, 94, 1712.05044

Shporer, A. et al. 2017, The Astrophysical Journal Letters, 847, L18, 1708.08455

Silburt, A., Ali-Dib, M., Zhu, C., Jackson, A., Valencia, D., Kissin, Y., Tamayo, D., & Menou, K. 2019, Icarus, 317, 27, 1803.02192

Silva Aguirre, V. et al. 2015, MNRAS, 452, 2127, 1504.07992

Silva Aguirre, V. et al. 2017, The Astrophysical Journal, 835, 173, 1611.08776

Skrutskie, M. F. et al. 2006, The Astronomical Journal, 131, 1163

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, Journal of Machine Learning Research, 15, 1929

Stassun, K. G., & Torres, G. 2018, The Astrophysical Journal, 862, 61, 1805.03526

Stellingwerf, R. F. 1978, The Astrophysical Journal, 224, 953

Struve, O. 1952, The Observatory, 72, 199

Sullivan, P. W. et al. 2015, The Astrophysical Journal, 809, 77, 1506.03845

Tamuz, O., Mazeh, T., & Zucker, S. 2005, MNRAS, 356, 1466, astro-ph/0502056

Temple, L. Y. et al. 2018, ArXiv e-prints, 1802.00766

Thompson, S. E., Mullally, F., Coughlin, J., Christiansen, J. L., Henze, C. E., Haas, M. R., & Burke, C. J. 2015, The Astrophysical Journal, 812, 46, 1509.00041

Tingley, B. 2003, Astronomy and Astrophysics, 403, 329, astro-ph/0303200

Toscher, A., & Jahrer, M. 2009

Triaud, A. H. M. J. et al. 2013, Astronomy and Astrophysics, 549, A18, 1208.4940

Triaud, A. H. M. J. et al. 2017, Astronomy and Astrophysics, 608, A129, 1707.07521

Ulmer-Moll, S., Santos, N. C., Figueira, P., Brinchmann, J., & Faria, J. P. 2019, Astronomy and Astrophysics, 630, A135, 1909.07392

von Boetticher, A. et al. 2017, Astronomy and Astrophysics, 604, L6, 1706.08781

von Boetticher, A. et al. 2019, Astronomy and Astrophysics, 625, A150, 1903.10808

Warner, B., & Robinson, E. L. 1972, MNRAS, 159, 101

Wheatley, P. J. et al. 2018, MNRAS, 475, 4476, 1710.11100

Winn, J. N. 2009, in IAU Symposium, Vol. 253, Transiting Planets, ed. F. Pont, D. Sasselov, & M. J. Holman, 99–109, 0807.4929

Wolff, S., & Simon, T. 1997, Publications of the Astronomical Society of the Pacific, 109, 759

Wolpert, D. 1992, Neural Networks, 5, 241

Wolszczan, A., & Frail, D. A. 1992, Nature, 355, 145

Wright, E. L. et al. 2010, The Astronomical Journal, 140, 1868, 1008.0031

Wright, J. T., Marcy, G. W., Howard, A. W., Johnson, J. A., Morton, T. D., & Fischer, D. A. 2012, The Astrophysical Journal, 753, 160, 1205.2273

Yoder, C. F., & Peale, S. J. 1981, Icarus, 47, 1

Yu, L. et al. 2019, The Astronomical Journal, 158, 25

Zeng, L., Sasselov, D. D., & Jacobsen, S. B. 2016, The Astrophysical Journal, 819, 127, 1512.08827

Zheng, S., Zeng, X., Lin, G., Zhao, C., Feng, Y., Tao, J., Zhu, D., & Xiong, L. 2016, New Astronomy, 45, 54

Zhu, W. W. et al. 2014, The Astrophysical Journal, 781, 117, 1309.0776

Zitlau, R., Hoyle, B., Paech, K., Weller, J., Rau, M. M., & Seitz, S. 2016, MNRAS, 460, 3152, 1602.06294

Zucker, S., & Giryes, R. 2018, The Astronomical Journal, 155, 147, 1711.03163