Navigating the Paper: Identifying Exoplanet Candidates Using WaveCeptionNet

Sixuan Han

April 3, 2025

1 Background of the Study

In recent years, the search for exoplanets has witnessed remarkable progress, fueled by advancements in observational techniques and data analysis methodologies. Among these techniques, transit photometry stands out as a powerful method for detecting exoplanets by measuring the periodic dimming of a star's light as a planet passes in front of it.[4] However, the sheer volume of data generated by transit photometry missions, such as the Transiting Exoplanet Survey Satellite (TESS), presents a significant challenge for efficient analysis and identification of exoplanet candidates. In response to this challenge, the paper titled "Identifying Exoplanet Candidates Using WaveCeptionNet" proposes an innovative approach that combines wavelet-transform-based preprocessing with a convolutional neural network (CNN) model based on Inception-v3 for fast and accurate classification of light curves [2]. In this review, we will summarize the key findings and contributions of the paper, highlighting its significance in the field of exoplanet research and its implications for future studies.

1.1 Object of the Study

An exoplanet, short for "extrasolar planet," refers to any planet located outside of our solar system that orbits a star other than the Sun. These planets vary greatly in size, composition, and orbit, and their discovery has revolutionized our understanding of planetary systems and the universe. Exoplanets are detected through various methods, including the transit method (observing the slight dimming of a star as a planet passes in front of it), the radial velocity method (measuring the gravitational influence of a planet on its parent star), and direct imaging (capturing the light emitted or reflected by the planet itself). Studying exoplanets provides valuable insights into planetary formation, evolution, and the potential for habitability beyond our own solar system.

The paper attempted to classify different types of exoplanet candidates based on light curve data obtained from the Transiting Exoplanet Survey Satellite (TESS) using the transit method.[2] Specifically, the exoplanet candidates they aimed to classify include:

- Confirmed planets (CPs)
- Known planets (KPs)
- Planet candidates (PCs)

These categories were based on the TESS objects of interest (TOI) catalog, which assigned targets to various categories including confirmed planets, known planets, and planet candidates. The paper grouped CP, KP, and PC targets together as planet candidates (PC).

In addition to exoplanet candidates, the paper also aimed to classify other types of celestial objects and phenomena present in the light curve data, such as eclipsing binaries (EBs), stellar variability (V), and instrument noise/systematic (IS).

1.2 Authors of the Study

The authors are Huiping Liao, Guangyue Ren, Xinghao Chen, and Yuxiang Li, affiliated to the Key Laboratory of In-Fiber Integrated Optics of Ministry of Education College of Physics and Optoelectronic Engineering, Harbin Engineering University, as well as Guangwei Li from the Key laboratory of Space Astronomy and Technology National Astronomical Observatories, Chinese Academy of Sciences. Guangwei Li's research focus includes galactic substructure, massive stars, flare stars, and astronomical Image Processing. Some of his recent papers include "Discovery of Two Different Full Disk Evolutionary Patterns of M-type T Tauri Stars with LAMOST DR8", "A Meteor Detection Algorithm for GWAC System", and "Magnetic Activity and Parameters of 43 Flare Stars in the GWAC Archive." (https://orcid.org/0000-0001-7515-6307)

The paper was received on May 21, 2023, revised on February 12, 2024 and accepted the following day. It was published on March 27, 2024.

1.3 Goal of the Study

The subject of this study is the classification of light curves obtained through transit photometry, particularly focusing on identifying exoplanet candidates. Transit photometry involves observing the periodic dimming of a star's light as a planet passes in front of it, providing valuable information about the size, orbit, and characteristics of exoplanetary systems. The study utilizes data from the Transiting Exoplanet Survey Satellite (TESS), which collects vast amounts of light curve data in its search for exoplanets. The light curves obtained from TESS observations are subject to various sources of noise, including instrument noise, background light, and interference from celestial objects such as eclipsing binaries and variable stars. Efficiently and accurately classifying these light curves is crucial for identifying potential exoplanet candidates and planning follow-up observations.

1.4 Wavelet Transform

The paper employs wavelet transform as a mathematical method for preprocessing light curve data to enhance the performance of their machine learning model, WaveCeptionNet, in classifying potential exoplanet candidates. Wavelet transform is a multiresolution analysis method in both the time and frequency domains, decomposing signals into different frequency components while extracting periodicities and separating noise. Specifically, the authors use the discrete wavelet transform (DWT) with a chosen basis function and threshold selection method to decompose the original light curve data into low-frequency and high-frequency components. They then employ wavelet packet decomposition tree to further decompose the signal recursively into constituent parts at different levels of resolution. The lengths of the resulting low-frequency (CA6) and high-frequency (CD6) components after six levels of decomposition are expressed mathematically in terms of the length of the original signal and the length of the wavelet used (Figure 1). After preprocessing, the authors conduct Min-Max normalization on the CA6 and CD6 components and resample them using cubic spline interpolation method to achieve consistent dimensions for input data required by neural networks. The high- dimensional data from multiple observations of exoplanet research,

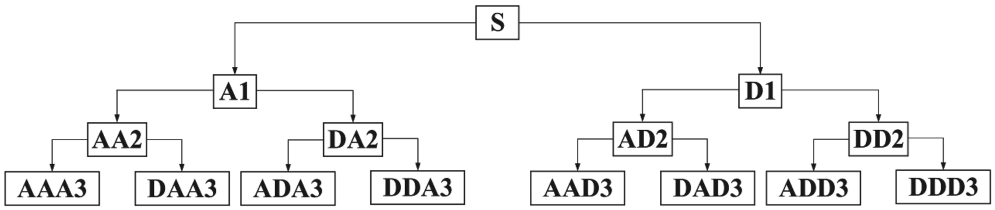

Figure 1: Wavelet packet decomposition tree [2]

after being processed by their method, has been reduced in length by a factor of 32 compared to the original data. This preprocessing pipeline aims to reduce noise, enhance signal resolution, and ensure consistent input data dimensions for training the machine learning model.

The wavelet transform preprocessing method that does not require folding the light curves or period parameters, which is the traditional approach of using machine learning to classify exoplanets. Applying wavelet transform to the data significantly reduces the noise and anomalies present in the raw light curve data, thereby enhancing the neural network's efficiency and dependability.

1.5 Neural Network Architecture

1.5.1 CNN

Inception-v3 is a deep learning model architecture primarily used for image classification tasks. It was developed by Google researchers as part of the Inception project, which aimed to explore and improve the efficiency and performance of convolutional neural networks (CNNs) [5]. The Inception-v3 model is characterized by its deep architecture with multiple layers of convolutional, pooling, and fully connected layers. It incorporates several key design features to enhance its effectiveness.

1.5.2 spatial dropout layer

Spatial dropout is applied during training by randomly setting some neuron outputs to zero, which helps the network learn more robust features, reduces overfitting, and improves generalization. Spacial dropout sets some channels to zero for each time step, not only considering individual time steps but also taking into account the correlation between time steps.

1.5.3 optimization

The paper uses the softmax function for nonlinearity and the cross entropy loss function with L1 regularization minimized by the ADAM optimizer.

1.5.4 evaluation

The paper evaluates the performance of their machine learning model, WaveCeptionNet, for classifying light curves of potential exoplanet candidates (PCs) along with other celestial phenomena such as eclipsing binaries (EBs), variable stars (V), and instrument noise/systematic (IS). They employ standard evaluation practices by randomly dividing the labeled data into training, validation, and test sets, achieving an overall accuracy rate of 95.0368% across the four classes. Performance metrics such as accuracy, precision, recall, and F1-score are computed for each class, along with macroaverage and weighted average metrics to account for class imbalance. The model exhibits high precision rates exceeding 92% for all classes and recall rates exceeding 95% for PCs, EBs, and V. However, a relatively lower recall rate of 83.58% is observed for IS, indicating challenges in accurately classifying noise. The paper also assesses the model's sensitivity to various parameters such as signal-to-noise ratio (S/N), orbital period, transit duration, stellar radius, planet radius, and transit depth, demonstrating the model's effectiveness in identifying PCs across different system configurations. Additionally, comparisons with previous studies suggest that WaveCeptionNet outperforms other methods in terms of recall for PC classification. Overall, the evaluation showcases the robustness and effectiveness of WaveCeptionNet in automatically classifying light curves to identify potential exoplanet candidates amidst noisy astronomical data.

2 Results of the Study

2.1 Physical Principles

2.1.1 Transit Photometry

Understanding transit photometry is crucial for comprehending the significance of the proposed method for identifying exoplanet candidates using WaveCeptionNet. Transit photometry is a technique used to detect exoplanets by observing the periodic dimming of a star's light as a planet passes in front of it. This dimming, or transit, occurs when the planet crosses between the observer and the star, causing a temporary decrease in the star's brightness (Figure 2). By carefully monitoring these dips in brightness over time, astronomers can infer the presence of an exoplanet and gather valuable information about its size, orbit, and characteristics.[4]

Transit photometry also provides valuable data about exoplanetary systems, including the size and orbital period of the planets. The depth and duration of the transit signal can reveal information about the size of the planet relative to its host star and the inclination of its orbit. By analyzing these characteristics, astronomers can infer the nature of the exoplanet, such as whether it is a gas giant, a rocky planet, or a potentially habitable world.

2.1.2 Confounding Targets

When classifying the targets, the authors identified four categories of interest: Planet Candidates, Eclipsing Binaries, Stellar Variability, and Instrument Noise and Systematic Effects (IS). The objective is to correctly distinguish the Planet Candidates, but it is important to understand the other three categories and why they might produce results similar to those of transiting planets. We have covered Eclipsing Binaries in class, so I will focus on what Stellar Variability and IS are.

Stellar variability refers to the changes in brightness or other observable properties of a star over time. These variations can occur on various timescales, ranging from milliseconds to years, and can be caused by a variety of physical processes happening within the star itself. Understanding stellar variability is crucial in astronomy because it can provide valuable insights into the internal structure, evolution, and behavior of stars.[3]

Here are some of the main types of stellar variability and the underlying mechanisms responsible for them:

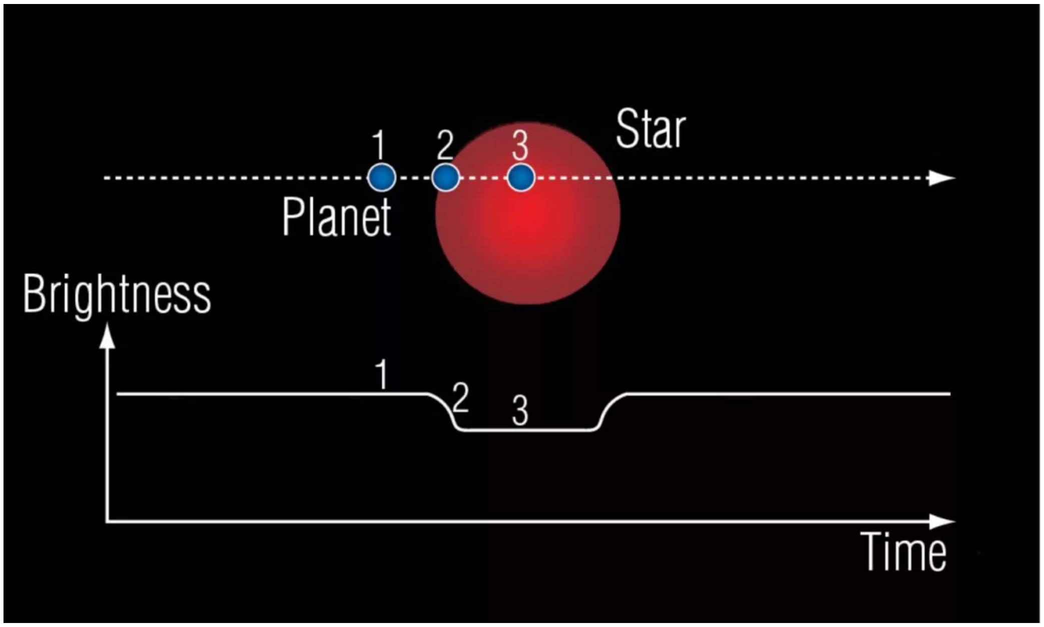

Figure 2: Transit Explanation. Source: Adapted from [4].

- Pulsations: Some stars undergo periodic expansions and contractions, causing changes in their brightness over time.

- Activity Cycles: Stars with magnetic fields, like the Sun, can exhibit cyclic variations in activity levels, including the number and size of sunspots, as well as the occurrence of solar flares (Figure 3). These activity cycles typically last for several years and are driven by the dynamo processes operating within the star's convective zone.

- Stellar Spots: Similar to sunspots on the Sun, starspots are regions of cooler temperature on the surface of a star caused by magnetic activity. As a star rotates, these spots can come into and out of view, leading to periodic variations in brightness. This type of variability is particularly common in young, rapidly rotating stars.

- Flares: Stellar flares are sudden, transient increases in brightness caused by the release of magnetic energy in a star's atmosphere. Flares can occur in stars with strong magnetic fields, such as young, active stars or certain types of binary systems. The energy released during a flare

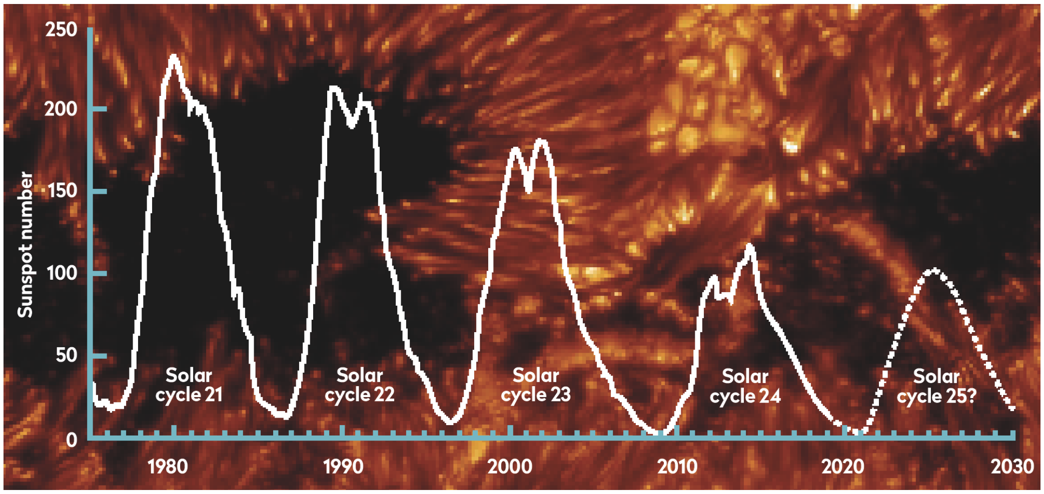

Figure 3: Solar Cycle. Source: Adapted from [1].

can be orders of magnitude higher than the total energy output of the star under normal conditions.

Instrument noise and systematic effects (IS) refer to unwanted variations or biases introduced into observational data by the instruments or methods used to collect the data. These effects are not astrophysical in nature but rather arise from limitations or imperfections in the instrumentation, data acquisition process, or data analysis techniques.

2.2 Significance

The study demonstrates the effectiveness of the wavelet processing method in reducing the dimensionality of the data by approximately 32-fold while largely removing noise. This reduction in dimensionality allows for more efficient and streamlined data analysis, enabling astronomers to sift through massive datasets from instruments like the Transiting Exoplanet Survey Satellite (TESS) more effectively. The demonstrated effectiveness of the proposed method in achieving about a 32-fold dimension reduction in light curve data is particularly noteworthy.

The integration of a convolutional neural network (CNN) model based on Inception-v3 represents a significant integration of deep learning techniques into the field of exoplanet research. Deep learning has emerged as a transformative technology in various fields, including astronomy, due to its ability to learn intricate patterns and representations from large datasets. By adapting a state-of-the-art CNN architecture to process wavelet-transformed light curve data, the study demonstrates the potential of deep learning for automated classification and detection of exoplanets. The CNN model achieves impressive accuracy and precision rates for classifying different types of light curves, including eclipsing binaries, planet candidates, variable stars, and instrument noise. With F1-scores ranging from 89.60% to 95.93% and precision rates exceeding 96%, the method demonstrates robust performance in identifying exoplanet candidates and minimizing false positives. the recall for (Yu et al. 2019) and Rao et al. (2021) are 61% and 74.3%, respectively, comparing to WaveCeptionNet's 95.38% in the PC, which also demonstrates the paper's significance.

The proposed method has practical applications for exoplanet research and the broader field of astrophysics. By automating the process of identifying exoplanet candidates and screening massive datasets, the method enables astronomers to efficiently analyze transit photometry data and prioritize targets for follow-up observations. This capability is crucial for accelerating the discovery and characterization of exoplanetary systems and advancing our understanding of planetary formation and evolution.

References

[1] Lucie Green. What causes peaks in activity on the sun? scientists are still trying to fully understand the 11-year solar cycle. Available online: `https://www.skyatnightmagazine.com/space-science/understanding-sun-science-solar-cycle`, 2024.

[2] Huiping Liao, Guangyue Ren, Xinghao Chen, Yuxiang Li, and Guangwei Li. Identifying exoplanet candidates using waveceptionnet. *The Astronomical Journal*, 167(4):180, mar 2024. doi: 10.3847/1538-3881/ad298f. URL `https://dx.doi.org/10.3847/1538-3881/ad298f`.

[3] A. McQuillan, S. Aigrain, and S. Roberts. Statistics of stellar variability from kepler. i. revisiting quarter 1 with an astrophysically robust systematics correction. *Astronomy &Astrophysics*, 539:A137, 2012. doi: 10.1051/0004-6361/201016148. URL `https://doi.org/10.1051/0004-6361/201016148`. Published online: 07 March 2012.

[4] The Planetary Society. Down in front!: The transit photometry method. Available online: https://www.planetary.org/articles/down-in-front-the-transit-photometry-method, 2020.

[5] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015. URL http://arxiv.org/abs/1512.00567.