IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ꁘ Peer-reviewed & Refereed journal ꁘ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11752

DEEP LEARNING EXOPLANETS DETECTION BY COMBINING REAL AND SYNTHETIC DATA

DILEEP M[1], USHA M[2]

Student, MCA, Bangalore Institute of Technology, Bengaluru, India[1]

Professor, MCA, Bangalore Institute of Technology, Bengaluru, India[2]

Abstract: Our work combines real observation data with synthetic data in exoplanet detection by deep learning. We collate datasets from space-based telescopes, like Kepler and TESS, from ground-based observatories for light curves, and spectral data. To further enrich the dataset, we produce synthetic data simulating a collection of astrophysical scenarios. Convolutional and recurrent neural networks enable model robustness and generalization. Accuracy and reliability of exoplanet detection will increase with training using the total dataset. Such integration will not only extend the scope of the training dataset to probe a far greater variety of astrophysical conditions but also speed up the discovery and characterization of exoplanets.

I. INTRODUCTION

Interest in space exploration has been the case with humankind since time immemorial. Out of many exciting new developments experienced in astrophysics over the last few decades, one is the discovery of exoplanets, which constitute those planets orbiting other stars other than the Sun the ones heliocentrically bound to our solar system. This rapidly growing field has benefited more from technological development, especially through the deployment of space-based telescopes like Kepler and TESS, while on the other hand, it also got a boost from progress in data analysis techniques for instance, deep learning.

Contrasted with these advances, the detection of exoplanets remains challenging. The transit signals of planets are typically faint, periodic, and buried between various sources of noise and artifacts in large stellar datasets. Traditional methods of analysis, although partly helpful in this regard, are not usually positioned to find these faint signals against the complexities of astrophysical phenomena and instrumental limitations.

The faster the pace of exoplanet exploration, the faster technology and the methods used in analyzing results must be developing. Among these crucial instruments, deep learning was inspired by the human brain itself. Algorithms from this area, such as CNNs and RNNs, have proven to be exceptionally good in recognizing exoplanetary signals from more complex data, like light curves and spectra. While it is the real observational data obtained from the telescopes that gives the essential insights, synthetic data from simulations broadens the dataset and can cover an extent of possible cases that are hardly reproducible through real data, which increases the effectiveness of deep learning models for discovering and understanding exoplanets.

This will enhance the strength and ability of deep learning models to generalize exoplanet detection. This development is critical in the reliable detection of exoplanetary signals against noisy data, as it will expand our understanding of planetary systems beyond our solar system. Deep learning techniques are therefore going to play a central role in uncovering new insights into the emerging discipline of exoplanetary science.

In our project, we focus on bridging the gap between real and synthetic data for exoplanet detection using deep learning methods. We draw on space-based and ground-based telescopes to leverage strengths from these approaches toward improving performance and reliability for deep neural networks. Accompanied by rigorous data preprocessing, model training, and validation, we aim to come up with algorithms that turn out to be more accurate, dependable, and adaptable in detecting and characterizing exoplanets.

Such a combined approach allows for the training dataset to reach further, and a much broader set of astrophysical conditions can be probed. In the long term, it holds the promise to expedite discovery and deepen our understanding of exoplanetary systems, bringing humankind closer to uncovering the hidden secrets of the universe.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11752

II. LITRATURE SURVEY

[1] This survey provides a comprehensive overview of deep learning techniques for exoplanet detection, covering methods such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs). It details their applications to light curves and spectra from space-based telescopes like Kepler and TESS, demonstrating how these techniques have improved the accuracy and efficiency of exoplanet discovery.

[2] This review discusses the application of deep learning to exoplanet detection, covering transit detection methods and signal processing techniques. It addresses the challenges of astronomical data analysis and outlines future research directions, highlighting the transformative potential of deep learning in enhancing data interpretation and analysis for exoplanet detection.

[3] This review explores the use of deep learning across various areas of modern astronomy, including exoplanet detection, astronomical object classification, and large survey data analysis. It evaluates the capabilities and limitations of deep learning approaches and their potential to significantly advance astronomical research through improved data processing and interpretation methods.

[4] This paper reviews the use of machine learning and data mining techniques in exoplanet research, with a focus on transit detection algorithms and feature engineering. It highlights the application of deep learning to spacebased telescopes and discusses the challenges and prospects of using these advanced techniques for detailed exoplanet detection and characterization.

[5] This review emphasizes the growing role of data-driven approaches in exoplanet astronomy, focusing on the challenges posed by noisy observational data and recent advances in using data-driven techniques for exoplanet detection. It discusses future research directions, advocating for the enhanced use of advanced data analytics to improve exoplanet discovery and understanding.

III. PROPOSED SYSTEM

In the proposed system of exoplanet detection, most of the inefficient traditional methods will be replaced by state-of-the-art computational techniques of machine learning and deep learning. The system components are explained below:

Data Acquisition and Preprocessing: The light curves and their spectra are collected by space- and ground-based telescopes. The acquired raw data will undergo stringent preprocessing to enhance quality through noise reduction, normalization, and instrumental artifact removal.

The two major features that can be extracted from this data are transit depth and transit duration. Then, more advanced feature selection techniques enable the identification of only the most important features for exoplanet detection, hence reducing model complexity and improving efficiency by dimensionality reduction.

It applies different high-end models like Convolutional Neural Networks, Recurrent Neural Networks, and Ensemble Methods. These models are then efficient at detecting complex trends within the data, very important in the identification of exoplanetary transits.

Model training and optimization will be conducted based on datasets with transit signals of confirmed exoplanets and transit-negative signals. This paper has employed several different training techniques, including crossvalidation and hyperparameter tuning, for model optimization in order to prove the developed models' robustness when applied to new data.

Statistical methods and anomaly detection for reliability are used to detect transit candidates and subsequently validate them to help get rid of false positives due to noise or other artifacts, refining the list of possible exoplanet candidates from the raw list.

Scalability and Efficiency: Tailored to manage huge datasets efficiently by parallel processing and cloud computing, this framework, therefore, can process huge volumes of data from missions like Kepler and TESS.

Real-Time Monitoring with Alerts: The system provides real-time monitoring, issuing alerts on their possible exoplanet candidates, so that follow-up observations can be done promptly for candidate verification and further study.

ISSN (O) 2393-8021, ISSN (P) 2394-1588
IARJSET
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ≍ Peer-reviewed & Refereed journal ≍ Vol. 11, Issue 7, July 2024
DOI: 10.17148/IARJSET.2024.11752

Continuous Learning and Adaptation: The continuous learning in the system keeps updating and refining its models at periodic intervals with the availability of new data. Feedback loops and retraining mechanisms in place will keep models effective against changing astronomy.

IV. METHODOLOGY

A. Data Collection :

Gather astronomical data from diverse sources, including space-based telescopes like Kepler and TESS, along with ground-based observatories. This data encompasses light curves, spectral data, and other pertinent information related to variations in star brightness and radial velocity measurements.

B. Data Preprocessing:

Clean and preprocess the collected data to eliminate noise, artifacts, and instrumental biases. This involves correcting for instrumental effects, removing outliers, and normalizing the data to prepare it for analysis.

C. Feature Engineering:

Extract meaningful features from the preprocessed data that indicate potential exoplanetary signals. These features may include transit depth, transit duration, orbital period, radial velocity variations, and spectral characteristics of the host star.

D. Model Selection:

Select suitable machine learning algorithms for tasks such as exoplanet detection and classification, tailored to the characteristics of the data and the project objectives. Commonly used algorithms include support vector machines (SVM), decision trees, random forests, as well as deep learning architectures like convolutional neural networks (CNN) and recurrent neural networks (RNN).

E. Training:

Train the chosen machine learning models using labeled datasets that contain confirmed exoplanet detections and non-exoplanetary signals. This step involves feeding the engineered features into the machine learning algorithms and adjusting model parameters iteratively to minimize prediction errors.

F. Validation:

Validate the trained machine learning models using separate validation datasets to evaluate their performance and ability to generalize. This ensures that the models can accurately identify exoplanet candidates in new, unseen data without overfitting to the training dataset.

G. Hyperparameter Tuning:

Fine-tune the hyperparameters of the machine learning models to enhance their performance and robustness. Techniques such as grid searches or randomized searches across defined parameter spaces are employed to identify optimal configurations for each model.

H. Evaluation:

Assess the performance of the trained machine learning models using appropriate metrics such as accuracy, precision, recall, and F1-score. This evaluation provides insights into the models' capability to correctly distinguish exoplanet candidates from false positives.

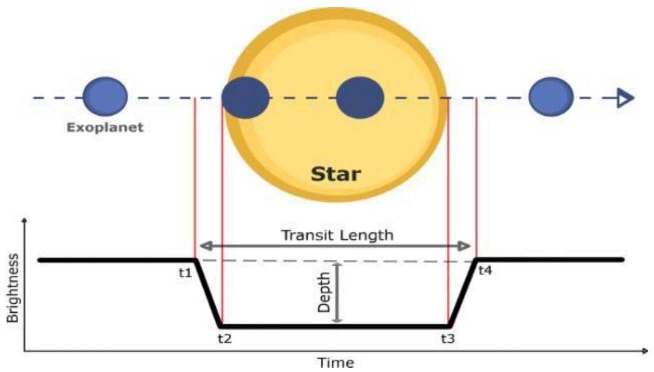


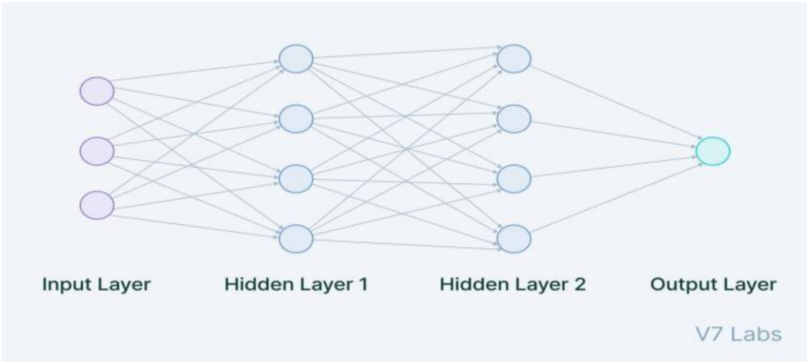
Fig.1 Transmit method

IARJSET
ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 7, July 2024
DOI: 10.17148/IARJSET.2024.11752

This approach involves observing stars and looking for periodic dips in their brightness, which occur when planets pass between the star and Earth, blocking a fraction of the star's light. By measuring the depth and duration of these dips, astronomers can infer properties of the orbiting planets, such as their size and orbital period. The transit method is particularly effective for detecting smaller planets and has contributed significantly to our understanding of the diversity of planetary systems. In addition to the transit method in Fig .1

V. ALGORITHM

A neural network model is a computational framework that learns structures and features of data, named after the human brain's architecture and functioning principle. It consists of a number of layers that are interconnected with nodes or neurons, which process and transform the data. Below is an outline of the neural network model:



Overview of the Neural Network Model

A neural network model is built to recognize patterns and make predictions out of those learned data through a process known as training. It generally contains three types of layers:

Input Layer: This is the first layer of the neural network. In this type of layer, raw data is entered into the model. Each neuron in this layer may represent a feature or even an attribute of the input data.

Purpose: To receive and pass the data to the layers further.

Hidden Layers: These are the intermediary layers which exist between the layers of input and output. Each hidden layer contains several neurons in them which take in the inputs, apply the weights and receive biases; then, they are passed through an activation function.

Advantages: The input could go through a series of weighted sums and successive activation functions toward complex patterns and representations found among the data.

Output Layer: It is the last layer of the network, which provides the predictions or the classification model, depending upon the input processed by the hidden layers.

Purpose: Final output from the network; can be anything like class label or continuous value for the task. Training Process: Training a neural network refers to the process of adjusting the weights and biases of its connecting links between neurons so that the difference between the predicted outputs of the network and the true target values is minimized. It is generally performed through the specific steps:

Forward Propagation: The data gets passed through the network from an input layer, through a number of hidden layers, to an output layer. Each neuron applies a weighted sum of its inputs and an activation function to produce its output. This function will calculate the network's predictions for a given set of inputs.

Calculating the Loss Function: The loss function measures the difference between the predicted output and the actual target values of the model (regression tasks can use mean squared error or cross-entropy loss for classification).

Purpose: The function to quantify how well the network is performing.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET
ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 7, July 2024
DOI: 10.17148/IARJSET.2024.11752

VI. CONCLUSION

In essence, the proposed exoplanet detection system integrated with machine and deep learning technologies extends far beyond the typical methods in place. The system is developed in computational algorithms and approaches for efficient automations, surpassing limiting restrictions of currently existing techniques for exoplanet identification from

observational data with higher sensitivity, precision, and productivity that can be produced through space and ground telescopes. It incorporates functions for data acquisition, data preprocessing, feature extraction, and model training and validation. Additionally, it is made up of monitoring and adaptation mechanisms associated with the relevant models. Excluding that, the system, coupled with various mechanisms of adaptation and feedback loops that foster continuous learning, is designed in such a manner that the system remains very effective and relevant enough in changing scenarios in the field of astronomy. The system, through its enormous data management, will provide real-time monitoring and alerts, allowing astronomers and researchers to make new exoplanetary system discoveries with unprecedented efficiency and precision. And, finally, seamlessly integrate with existing astronomical tools.

VII. ENHANCEMENT

Some of the most promising directions include:
Advanced Machine Learning Techniques: Implement state-of-the-art machine learning methodologies, such as reinforcement learning, transfer learning, and Bayesian optimization, to improve on the accuracy and efficiency of exoplanet detections.

Better Feature Engineering: Innovative techniques for feature extraction, such as time frequent analysis, wavelet transforms, or multi-scale feature integration, can reveal the detailed patterns and the characteristics of exoplanetary transits to improve the detection performance.

Multi-modal data fusion: A spectrum of data, including light curves, spectral data, and imaging, can offer a more holistic view of exoplanetary systems and therefore increase the strength of detection algorithms through the blending of diverse sources of information. the classification and characterization of exoplanets into detected classes and detailed descriptions, such as the physical characteristic of their orbits and atmospheres, will better describe the nature and possible support for life within these exoplanets.

Interpretability and Explainability: Techniques in interpretability and explainability in machine learning models are set to become increasingly important in rendering the decision processes of these models understandable by astronomers so that the inferences of the system are trusted even more.

Collaborative Platforms and Citizen Science: The availability of citizen science platforms in the domain of amateur astronomy greatly improves possibilities for analysis and validation through sharing knowledge and passion with much wider communities.

Integration with Next-Generation Telescopes: This system has to be compatible with the upcoming space-based observatories, among them the James Webb Space Telescope and the Nancy Grace Roman Space Telescope, in order to tap into their advanced functionality and further expand the exoplanetary research horizon.These avenues for enhancement are expected to lead the exoplanet detection system continuously forward into the advancement of the field of astronomy, bringing about new findings and excitement

REFERENCES

[1]. Khodayari, K., & Kouhestani, H. (2020). "A Survey on Deep Learning Techniques for Exoplanet Detection." IEEE Access.
[2]. Riberio, F., et al. 2018. "Exoplanet Detection Using Deep Learning: A Review." IEEE Signal Processing Magazine.
[3]. Zhang, Y., et al. (2018). "Deep Learning for Astronomy: A Review." Publications of the Astronomical Society of the Pacific.
[4]. Shallue, C. J., & Vanderburg, A. (2018). "Machine Learning and Data Mining for Exoplanet Research." Annual Review of Astronomy and Astrophysics.
[5]. Hsu, D.C., et al. (2019). "Data-driven Discovery in Exoplanet Astronomy." Annual Review of Astronomy and Astrophysics.