Deep Learning Algorithm for Exoplanet Classification

Brian Johnson
College of Arts and Sciences
Georgia State University
Atlanta, Georgia

ABSTRACT

With the discoveries of exoplanets increasing, creating techniques used to classify them are essential for efficiency. While there are ways to classify exoplanets, they often suffer from limitations. Therefore, this paper reflects a deep learning-based algorithm trained on a dataset consisting of information and observations from known exoplanetary systems for identifying planets outside of our solar system while also distinguishing these exoplanets from fake positives. The methodology involved training a machine learning model by using linear regression with various data about exoplanetary systems such as light curves. Although expectations were high, results demonstrate the lack of effectiveness of this approach, having an accuracy of less than 20%.

Keywords: Exoplanets, Linear Regression, Light Curves, Machine Learning

I. INTRODUCTION

Deep learning is a fairly new, well-researched and peer reviewed branch of artificial intelligence (AI) that was introduced to computer scientists in the mid 1980s. Later, it was introduced to artificial neural networks in the year 2000. It teaches a computer to interpret and process data in a human-like manner, allowing deep learning models to recognize significant amounts of data to accurately produce predictions and insights.

Due to deep learning models' ability to process extensive amounts of data, professionals in a plethora of fields use these models for numerous reasons. One being classification which is commonly done through Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs). CNNs are typically used for image-processing, computer vision, and pattern recognition [1] while LSTMs are typically used for its ability to learn and analyze from sequential data [2]. Both combined permits a computer to examine spatial features of images while extracting raw data to improve the accuracy of a system [3].

II. RELATED PUBLICATIONS

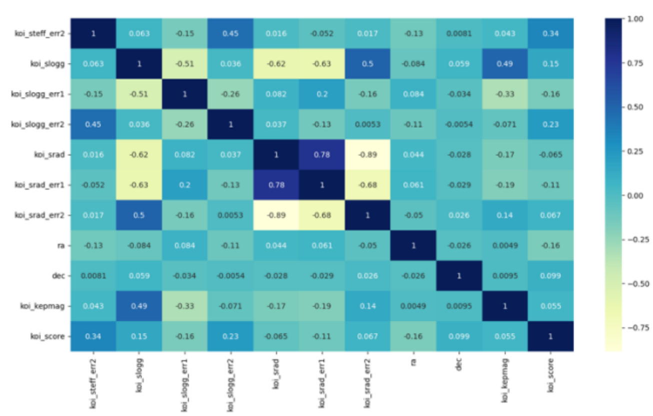
Deep learning models are heavily used by astronomers for classifying a planet outside of our solar system (exoplanet). The classification models typically used are SVM, ANN, and Naïve Bayes Classifiers [4]. Astronomers often utilize light curve data to detect an exoplanet, however, the data can be complex.

III. NEW METHOD AND IMPLEMENTATION

In my proposed idea, I am aiming to redefine the existing deep learning models used by astronomers to create a new deep learning algorithm for classifying exoplanets. I believe that there are more efficient and accurate ways to detect an exoplanet with the substantial amount of astronomical data available. I will be testing this using Python's various deep learning frameworks to implement my algorithm and the software will analyze and learn from a preprocessed dataset.

Data used by the algorithm consisted of disposition scores, orbital days, transit information, temperature, and more all within a dataset provided by the NASA Exoplanet Archive. The goal is to use the various data to predict disposition scores, which is a score used to determine the confidence that an object is either a candidate or false positive. After collecting data, it was then preprocessed, normalized, and cleaned for analysis, removing irrelevant data such as planets names from the raw data.

The data was then split for training and testing with X being every variable except for the target variable and y being the target variable. 80% of the data was then used for training. Using Python's pandas and matplotlib library, the data could finally by analyzed. With "koi_score", the disposition scores, being the target variable, I decided to look at its correlation between it and the other variables using a heatmap.



The architecture choice was linear regression due to its simplicity and ability to evaluate data and establishing relationships between multiple variables.

IV. Results and performance analysis

Using y_lr_train_pred = lr.preditct(X_train) and y_lr_test_pred = lr.predict(X_test), the results consisted of koi_scores array([0.6164439 , 0.31566165, 0.51102145, ..., 0.71523412, 0.72344471,0.04991117]) and array([0.54252793, 0.58221617, 0.24777767, ..., 0.37474478, 0.68083614, 0.311387]) respectively. For koi_scores, if the number is higher the more confidence it has. After getting results, I tested the performance of the model using Mean Squared Error and Root Mean Squared Error:
lr_train_mse = mean_squared_error(y_train, y_lr_train_pred) lr_train_r2 = r2_score(y_train,y_lr_train_pred)
lr_test_mse = mean_squared_error(y_test, y_lr_test_pred)
lr_test_r2 = r2_score(y_test,y_lr_test_pred).

After evaluating the performance of the model, it was shown that the model had less than a 20% accuracy for both the training and test models for Mean Squared Error and Root Mean Squared Error:
lr_train_mse = 0.19106838231928056
lr_train_r2 = 0.1584517887073953
lr_test_mse = 0.19430873366418716
lr_test_r2 = 0.15142411935812605.

According to the results, the model is not effective at determining whether a planet is outside of the solar system or not.

V. Conclusion

Exoplanets are being discovered more as of recent, making it a necessity to automate the process of discovering one. Although, there are many deep learning algorithms that are used to classify them, I was curious to know potential techniques to compete with them. Regardless of the model not being accurate, future plans entails using different datasets and models such as a Random Forest Model or a type of neural network.

VI. References

[1] D. Arora, M. Garg and M. Gupta, "Diving deep in Deep Convolutional Neural Network," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020, pp. 749-751

[2] M. Bkassiny, "A Deep Learning-based Signal Classification Approach for Spectrum Sensing using Long Short-Term Memory (LSTM) Networks," 2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2022, pp. 667-672

[3] M. Abdallah, N. A. L. Khac, H. Jahromi and A. D. Jurcut, “A Hybrid CNN-LSTM Based Approach for Anomaly Detection Systems in SDNs,” 2021 In Proceedings of the 16th International Conference on Availability, Reliability and Security (ARES '21), 2021, pp. 1–7

[4] R. Jagtap, U. Inamdar, S. Dere, M. Fatima and N. B. Shardoor, "Habitability of Exoplanets using Deep Learning," 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2021, pp. 1-6