Grenze International Journal of Engineering and Technology, June Issue

Extrasolar Planet Detection using Machine Learning

Maitreya Kanitkar[1], Prof. Minal Apsangi[2], Aanchal Agarwal[3], Prasad Deshmukh[4] and Mayuresh Jadhav[5]
[1-5]Pune Vidyarthi Griha's College of Engineering & G. K. Pate (Wani) Institute of Management, Pune, India
Email: maitreya.kanitkar@gmail.com, mra_it@pvgcoet.ac.in, {aanchalpvg, prasaddeshmukh156, mj.ethical}gmail.com

***Abstract*—Planets outside the solar system were mysterious to astronomers for a long time; astronomers in the past observed these changes in the night sky and noted them. Since then astrophysics has become a data-driven endeavour. The enormous streaming of data collected by scientific projects like the Kepler mission in the process of detecting Extrasolar planets(Any planet that lies beyond our solar system) by telescopes, satellites, and spaceships is becoming too large, which currently relies on human involvement during the selection and classification processes. This makes the analysis and classification of potential Extrasolar Planets a time-consuming matter. Therefore, there is a need for an automated and unbiased way to identify the Extrasolar Planets. The method used for identifying Extrasolar planets consists of a Convolutional Neural Network model. The model classifies objects denoted as 'Candidates' into either 'confirmed planets' or 'false-positives'. Extrasolar Planets Detection proceeds by analyzing the variations in the brightness of a remote star.**

***Index Terms*— CNN (Convolutional Neural Network), Extra-solar planets, Confirmed, False-positive, Candidate, Deep learning, Dummies.**

I. INTRODUCTION

This space we declare to be infinite... In it is an infinity of worlds of the same kind as our own."
– Giordano Bruno (1584)
Until around 30 years ago, we only knew of the eight planets in our solar system. However, it was long thought that different stars would jointly contain their planetary systems. Such planets would be known as Extrasolar Planets. There are several methods currently used to detect Extra-Solar Planets. If we were able to directly observe a planet, we can directly have a spectrum of its atmosphere using a conventional spectrometer or an integral field spectrometer. Hence, gathering information like physical and chemical characteristics of a planetary atmosphere is possible, despite the that direct imaging of exoplanets remains challenging. The reason why it is difficult to detect these objects is the fact that planets are dark (don't have the light of their own) and are very close to an extremely bright source of light, their host star. A planet is a billion times fainter than its host star in visible light. Therefore, we make use of various indirect methods to detect them. The advent of large-scale transit surveys revolutionized our understanding of Extra-Solar Planets. Both Ground-Based and Space-Based telescopes have provided us with an unprecedented volume and rate of discoveries. Perhaps the foremost notable of these surveys is the Kepler Space Telescope of NASA. The task of creating an automated system for identifying Planet Candidates in Kepler's data using a Deep Learning Algorithm, an Advanced Machine Learning Technique is to be done. Recently, different machine learning algorithms have begun to be used for this and other connected functions. An original machine learning algorithm for classifying planet candidates and false positives in Kepler data was the Auto-Vetter, a random forest classifier that made decisions based on

Grenze ID: 01.GIJET.8.2.15_1

metrics generated by the Kepler pipeline. More recently a logistic regression algorithm was used to attempt to detect non-transiting hot Jupiter, self-organizing maps and random forests were used to classify variability, eclipsing binaries, and transiting planets in both Kepler data and ground-based surveys data. Neural networks, a type of Deep Learning Model, have been used successfully for detecting transits in simulated Kepler-like data and for classifying planet candidates and false positives in Kepler data. The Machine learning method is expanded by using a supervised Convolutional neural network architecture but makes several key modifications to enhance the network's ability to classify signals in the qualitatively different datasets.

II. BACKGROUND

There is completely different analysis work done for detecting extrasolar planets, at first exoplanets were detected using scientific methods like research by W.J.Borucki & J.D.Scargle that's Detection of Extrasolar planetary by Transit[4]. This analysis proves that the photometric technique of finding out planets around stars depends on observing the decrease in stellar light. Similarly another analysis by Y. Tsapras that's Microlensing Searches for Exoplanets[5]. This covers basic theoretical ideas in microlensing. Some research involves machine learning methods to find exoplanets. The famous research by C.J.Shallue & A.Vanderburg that's Identifying exoplanets with Deep Learning, A five-planet Resonant Chain around Kepler-80 and an eight planet around Kepler-90 Present a technique for classifying potential planets using deep learning[2]. There is also completely different machine learning research like Improving Convolutional Neural Networks for Exoplanet Detection by C.Greeven, this work shows how extending the number of light curves views truly benefits CNN-based extrasolar planet detection. and comparing convolutional neural networks and recurrent neural networks for exoplanet detection by S.Koning, They compared CNN and RNN for exoplanet detection[1]. DATA AND PRE-PROCESSING The data required for the project has been taken from NASA Exoplanet archives[3]. The data is called a cumulative dataset and contains 9564 records and around 50 features. Most features have been retained with others either being dropped altogether or used as targets. The dataset primarily contains 3 classes i.e. CONFIRMED, FALSE POSITIVE, and CANDIDATE. The data cannot be used as it requires more refining to suit the model. The data has been collected by various telescopes that are currently in orbit. To reduce the time and complexity of the model we initially split the dataset into two i.e first dataset which contained the classes CONFIRMED and FALSE POSITIVE, and the second dataset containing the class CANDIDATE. Some of the columns which contained unique values and some columns which were empty reduced the size of our data. Dummies were created to get the features and targets. The former dataset was split into the training set and testing set and the latter was used to come up with predictions.

III. DATA AND PRE-PROCESSING

The data required for the project has been taken from NASA Exoplanet archives[3]. The data is called a cumulative dataset and contains 9564 records and around 50 features. Most features have been retained with others either being dropped altogether or used as targets. The dataset primarily contains 3 classes i.e. CONFIRMED, FALSE POSITIVE, and CANDIDATE. The data cannot be used as it requires more refining to suit the model. The data has been collected by various telescopes that are currently in orbit. To reduce the time and complexity of the model we initially split the dataset into two i.e first dataset which contained the classes CONFIRMED and FALSE POSITIVE, and the second dataset containing the class CANDIDATE. Some of the columns which contained unique values and some columns which were empty reduced the size of our data. Dummies were created to get the features and targets. The former dataset was split into the training set and testing set and the latter was used to come up with predictions.

IV. MODEL

The convolutional neural network model used in this project utilizes "supervised learning," which means the model is provided with a labelled set of data, from which it can learn. A CNN is used since it provides much better results than an MLP classifier. The CNN used in the model uses two convolutional 1D layers followed by a maxpool layer. The convolution part is followed by a 4 layer dense neural network(Figure.1). The convolutional part is used for feature learning and the dense network which follows is used for the final classification(Figure.2). The model runs for 50 epochs with a batch size of 5. An increase in the number of epochs sees no drastic increase in accuracy but takes a lot of time and may also lead to overfitting.

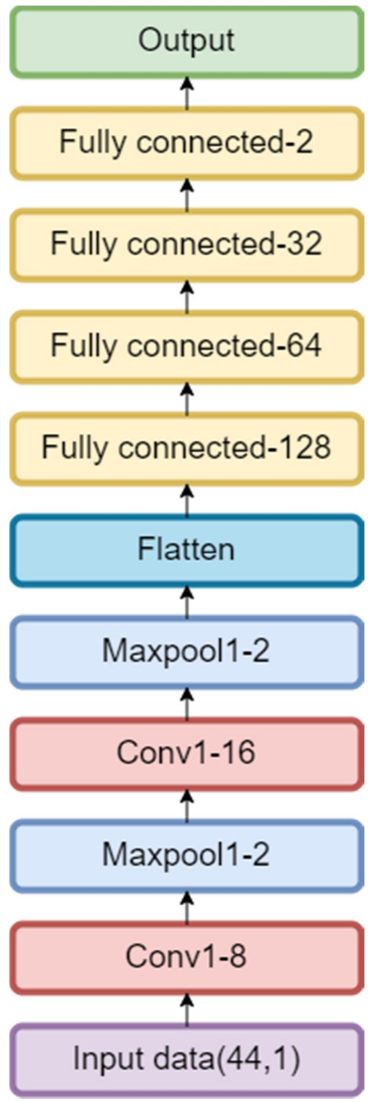

Figure.1. (Convolutional Neural Network Model

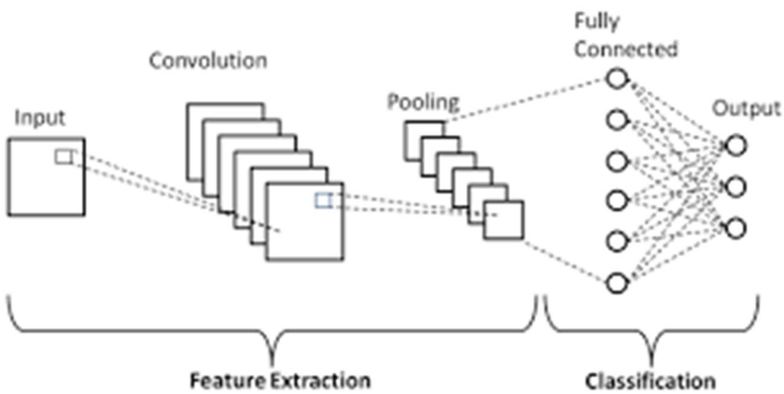

Figure.2. (Convolutional Neural Network illustration)

The training accuracy increases over time as the number of epochs increases whilst the validation accuracy does not see a great change in accuracy(Figure.3). As the accuracy increase, the loss decreases with an increase in the number of epochs(Figure.4).

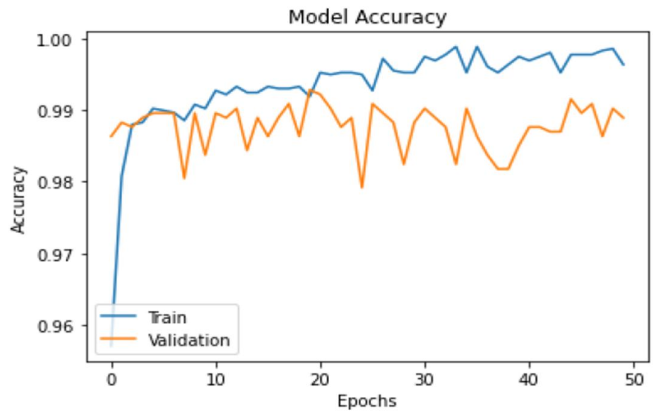

Figure.3. (Model Accuracy)

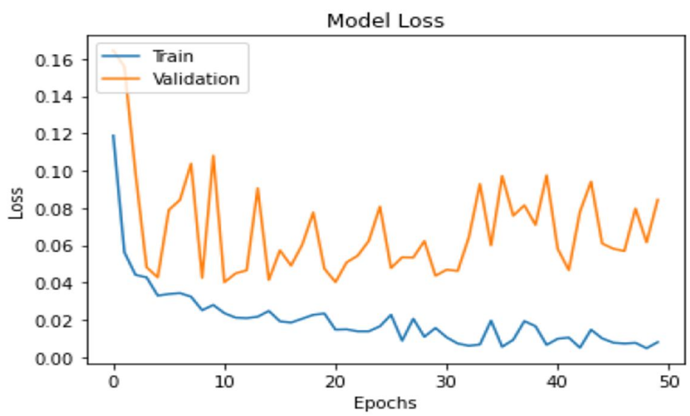

Figure.4. (Model loss)

Comparison of output between the CNN model and MLP Classifier
CNN model

	precision	recall	f1-score	support
0	0.98	0.98	0.98	729
1	0.99	0.99	0.99	1466
micro avg	0.99	0.99	0.99	2195
macro avg	0.99	0.99	0.99	2195
weighted avg	0.99	0.99	0.99	2195
samples avg	0.99	0.99	0.99	2195

MLP Classifier

	precision	recall	f1-score	support
0	0.79	0.84	0.81	689
1	0.97	0.97	0.97	1502
micro avg	0.91	0.93	0.92	2191
macro avg	0.88	0.91	0.89	2191
weighted avg	0.91	0.93	0.92	2191
samples avg	0.71	0.71	0.71	2191

0 refers to the CONFIRMED class and 1 refers to the FALSE POSITIVE class.

V. ALGORITHM

1. Preprocess the first dataset(containing classes CONFIRMED and CANDIDATE).
2. Split the first dataset into training data and test data.
3. Create the Convolutional Neural Network model.
4. Train the model on the training data for 50 epochs with a batch size of 5.
5. Make the predictions to test model accuracy.
6. Calculate the model accuracy by comparing predictions with actual data.
7. Now preprocess the second dataset(containing class CANDIDATE).
8. Make the future predictions for the CANDIDATE class by running the second dataset on the trained model.
9. Display the future predictions.

VI. APPLICATIONS

The ultimate goal of NASA's Exoplanet Exploration Program is to identify candidates that could harbour life. Exoplanets that could hold signs of life are waiting to be revealed by detailed analysis of the atmospheres of planets. When we analyze light shot by a star through the atmosphere of a distant planet, with the help of a technique known as the transit method, the effect looks like a shadow. The ingredients that are present in the alien atmosphere are known from the slices missing from the light spectrum. One of the patterns of black gaps might indicate methane or another oxygen. A strong argument for the presence of life could be seeing those ingredients together.

1. To discover planetary systems and Earth-like planets around nearby stars.
2. Building an essential step, toward the goal of discovering habitable planets for future generations.
3. To identify candidates for the next massive step for the human race.

VII. ADVANTAGES & DISADVANTAGES

Advantages

1. CNNs are much better at feature mapping.
2. More accurate than a multi-layer perceptron.

Disadvantages

1. Still requires lots of accurate human contribution.
2. A negative transit requires excessive processing.
3. Time-consuming as the data provided by Kepler may contain too many false positives.
4. The absolute value of error may increase with the increase in records.

VIII. CONCLUSION

Extra-solar Planets are those planets that orbit around a star other than our sun. Detection of Extra-solar Planets has always been a topic of interest. Kepler's satellite produces a large number of false positives in case of noisy

data, which have to be reviewed manually. To minimize manual work, to produce more accurate results, and to minimize the time required to predict the Extra-Solar Planet, automation in detecting
the Extra-Solar Candidates is required. We trained a model consisting of a Convolutional Neural Network to identify extrasolar planets in Kepler Cumulative data. Delivering more accurate and true candidates of Extrasolar planets, with less human involvement reduces the period of the whole detection process.
Our approach consists of three main steps.

- Firstly, split the first dataset into the training set and test set. Then train the model on the training set.
- Secondly, predict output and compare it with the testing set and generate accuracy for the model.
- Lastly, use the second data i.e the dataset which contains all the candidates to make a prediction using the previously trained model. Based on the accuracy our model has provided we can back our new predictions.

Even if a machine learning model works well in the development environment, it is prone to make mistakes on unseen data. Hence, such methods should be used alongside human supervision. Nonetheless, at the current stage, these models can provide a very reliable system to rule out a large number of false positives and can drastically reduce the number of cases requiring manual reviews.

References

[1] Sebastiaan Koning. “Comparing Convolutional neural networks and recurrent neural networks for exoplanet detection,”, 34, December 2018.

[2] Christopher J. Shallue and Andrew Vanderburg "Identifying extra-solar planets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90” The Astronomical Journal, 155:94. (21pp), February 2018.

[3] https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=cumulative

[4] W.J.Borucki & J.D.Scargle. “Detectability of extrasolar planetary transits”. Theoretical studies branch NASA research centre, 1985.

[5] Yiannis Tsapras, “Microlensing Searches for Exoplanets”. The Astronomical Calculation Institute, Heidelberg, Germany, 2018.