Habitability of Exoplanets using Deep Learning

Rutuja Jagtap, Unzela Inamdar,
Shivani Dere, Maziya Fatima
Computer Science and Engineering
MIT School of Engineering, MIT ADT University
Pune, India

Prof. Nikhilkumar B Shardoor
Assistant Professor, Dept. of CSE,
MIT School of Engineering, MIT ADT University
Pune, India

2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) | 978-1-6654-4067-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/IEMTRONICS52119.2021.9422571

Abstract - In observational astronomy, researchers have reached a point where large datasets from sky surveys are regularly published. Sending probes to land on asteroids, collect material, and ship it back to Earth is one strategy for space exploration. Astronomical data is rapidly increasing in size and complexity as space and ground-based telescopes are developed and deployed. Machine learning has gained popularity among astronomers in recent years, and it is now used to solve a variety of tasks, including classification, regression, clustering, outlier detection, time series analysis, association law, and so on. The study on previous work of papers on exoplanets and their habitability show machine learning methods including Support Vector Classification, Random forest, K- Nearest Neighbor, and so on were used in the majority of the papers. Until now, only a few deep learning methods have been studied. As a result, this research work overcomes different challenges faced by astronomers dealing with large data and seeking relevant knowledge for each objective using Deep Learning techniques. Furthermore, deep learning techniques are capable of managing complex data with ease. To begin, this paper proposes ASTRONET, a deep learning architecture, to look for exoplanets that are habitable based on their planet gravity, eccentricity, mass, radius, and other characteristics. This research paper explains the entire process of finding exoplanets and categorizing them based on their habitability. Finally, to establish a knowledge base of parameters that affect the habitability of exoplanets.

Keywords – Exoplanets, Deep Learning, Astronet, Classification, Artificial Intelligence, TESS (Transiting Exoplanet Survey Satellite), Astronomy, Machine Learning, Habitability, Convolution Neural Network (CNN)

I. INTRODUCTION

Most of us wonder if there is life outside of our planet? We reside in a universe that is colossal beyond our imagination. The problem of detecting anomalies in massive, high-volume astronomical datasets is discussed in this paper, and a solution based on machine learning algorithms is studied and the most impactful and intense deep learning technique is proposed.

The NASA Astrobiology Roadmap has established an understanding of the origin and distribution of habitable planets and moons in the Galaxy as a major research theme (Des Marais et al., 2008). We require more detailed and cohesive models to determine the conditions for the habitability of planets to understand how sustainable the atmosphere and living conditions are or could be, to generate detectable biosignatures, and determine the similarities to the conditions that support life on Earth.

The habitable zone for exoplanets was first presented and modelled in detail by [14], who explained how physics and chemistry play a vital role in the exoplanet study. The immense diversity of exoplanets, as well as the estimated variation in their atmospheres in terms of mass and composition, has sparked a deep desire to rethink planetary habitability.

So, in this paper, the model gives an overview of existing machine learning techniques for assessing exoplanet habitability and suggests the ASTRONET deep learning approach to achieve our objectives.

II. LITERATURE SURVEY

This paper discusses the effectiveness of various Machine Learning methods in classifying exoplanets into classes of thermal habitability and characterizing them based on possible habitability. The challenges solved are supremacy in the non-habitable planetary sample dataset using under-sampling methods (of dominant class samples) and over-sampling methods [1].

They proposed to build a Machine Learning model to automate Kepler cumulative object of interest data classification and then deploy it. KNN, Random Forest, and SVM models were used, validated, and checked for precision, accuracy, and recall to create a better model for classification. And for model Deployment- Flask API, Azure Cloud were used. It

978-1-6654-4067-7/21/$31.00 ©2021 IEEE

has a Comprehensive ML pipeline: Engineer data, train, and test models. A robust model is required to handle errors [2].

The purpose of this project was to use the data as training data and use planetary and stellar characteristics to construct a machine learning model that can anticipate habitable planets. Data source: From the explanation of NASA Exoplanet Archive Attribute of Kepler data, 14 stellar and planetary features were discovered. Few of them are "Planetary Radius", "Isolation Flux", "Equilibrium Temperature", "Orbital Period", "Distance from parent Star", "Stellar Temperature" SVM with rbf kernel used. But a robust model needed to cope with huge and complex data [3].

In this paper, they used various supervised learning algorithms to predict the habitability of recently observed exoplanets. Used various models like CART, SVM, FNN, Random Forest, Logistic Regression, and Naïve Bayes. A Regression tree was created to anticipate the value of ESI for a planet. It was not suitable for unlabelled data and image data. Further optimization of the model is possible [4].

Applied Deep Learning algorithms to identify Kepler candidate transit cases. Used Kepler DR24, TESS. CNN models used are Astronet: Baseline Model, Exonet: Revised Model, and Exonet-XS: Decreased Model Size. Stellar parameters include metallicity, radius, mass, stellar effective temperature, surface gravity, and density. The model uses two views of phase-folded light curves as inputs, which is followed by completely connected layers that output a value between 0 and 1 that predicts whether transit is a planet or not. Astronet was prone to overfitting [5].

The model can differentiate between false positives and genuine exoplanets, with good accuracy. The training set was derived from the NASA Exoplanet Archive's Autovetter Planet Candidate List. It requires a high training period and high computational power. But it requires a high training period and high computational power [6].

The probability of an observation being an exoplanet is estimated using a number of datasets and classification models. For predicting exoplanets' existence, the proposal aims at using three classifiers. By using different classification algorithms like SVM, ANN, and Naive Bayes Classifiers can obtain different results. It is not suitable for unlabelled data [7].

It concentrates on the implementation of various algorithms of Machine Learning on NASA's Kepler data for the prediction of exoplanet habitability disposition. The proposed model will be able to operate on data generated by different ground and space observatories and classify exoplanet candidates as habitable or non-habitable. It involves the execution of supervised Machine Learning algorithms which include K Nearest Neighbor, Logistic Regression Naive Bayes, Decision Tree, and Random Forest. The binary classification of objects as “FALSE POSITIVE” or “CONFIRMED” exoplanets is taken into account by the model. The decision tree algorithm is a suitable model for predicting habitability [8].

The traditional Machine Learning models like SVM, Random Forest, Decision Tree, etc. used to predict the habitability of exoplanets had certain drawbacks like time-consuming and variation in outcomes.

So, the proposed Deep Learning model i.e “Astronet” is a modern technique backed with high computation capabilities that will provide error-free and unbiased outcomes.

III. Problem Statement

“Exoplanet Habitability” A planet outside the Solar System is known as an exoplanet or extrasolar planet. Habitability refers to an area or region where life can sustain. Every day, a large amount of astronomical data is produced. NASA is a well-known space agency that has similar data. Satellites, radio telescopes, orbital telescopes, space stations, and other instruments are used to collect this information. The instruments assist in the search for new exoplanet information. Additionally, they gather data on the physical states of the exoplanet. The characteristics of exoplanets will be studied using deep learning and machine learning. Further, the model will be able to discover what the key features of an exoplanet are that make it ideal for sustaining life.

IV. Objective

To determine the habitability of exoplanets based on different properties of exoplanets using deep learning algorithms and techniques.

V. Proposed system

There is numerous research work done in machine learning techniques for the classification of habitability of exoplanets. In this paper, the research proposes a deep learning model for the detection and classification of planets based on their habitable behavior. ASTRONET deep learning model is used to predict anomalies in astronomical bodies by classifying whether the studied newly existing planet is habitable or not. The design of the overall module including process flowchart, algorithm, detailed architecture of ASTRONET is explained.

Prediction of habitability of exoplanets using Deep Learning algorithm: Habitability here refers to the quality of having similar properties of being adequate enough to live in. Habitability can be determined using the exoplanet properties like mass, radius, eccentricity, orbital inclination, gravity, metallicity. Exoplanets are planets that do not orbit around our sun but do orbit around a star. So, we put forward the idea of predicting the habitability of exoplanets using the previous huge amount of data present with the help of deep learning techniques.

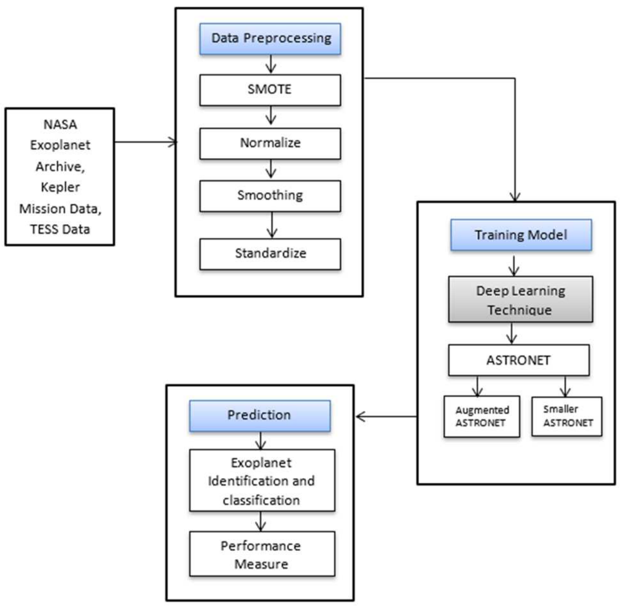

Fig 1. Flow Chart explains the total flow of Exoplanet Classification

ASTRONET:

The proposed ASTRONET architecture to identify and detect potentially habitable exoplanets that support a variety of planet and star characteristics. NASA's Exoplanet Program, NASA's Kepler mission results, and data from the Transiting Exoplanet Survey Satellite (TESS)[9].

These data must be compelled to be efficient, which means they must be reviewed for false-positive signals, such as those caused by stellar eclipses and instrumental noise, which outnumber true planet transit signals.

Let's start with the information gathered by Kepler's telescope, which was used to detect the presence of a planet. A light-weight curve is a graph that displays the brightness of a star (as determined by Kepler's photometer) over time. When a planet moves in front of a star, it blocks some of the suns, causing the measured brightness to drop and then rise again.

Now let's understand the model flowchart (Fig 1) in detail.

A. DATA GATHERING

We start with gathering information from various available data sources. We fetch the data from NASA Archive, Kepler Mission Data, and TESS (Transiting Exoplanet Survey Satellite) data.

B. DATA PREPROCESSING

It is essential to perform data preprocessing before feeding our model with data. This will result in high-quality data or valuable knowledge, which will have a direct effect on our model's ability to learn.

1.SMOTE

SMOTE stands for Synthetic Minority Oversampling. It is a method for eliminating data imbalances. It is introduced to minimize dependency on majority class values.

2.Normalization

The input dataset includes several features with varying ranges and normalization assists in getting them all to a similar scale. The values in the range [0,1] are rescaled.

3. Smoothing

Filters are used to smooth things out. Filters provide robustness in the face of noisy data. This paper aims to increase the accuracy of the data without distorting the signal tendency by averaging out neighboring data points.

4. Standardization

Standardization helps in generalizing the data. By generalizing the numerical conditions of inconsistent data, standardizing the dataset makes the training process more well behaved.

C. TRAINING MODEL

The ASTRONET is proposed which is similar to Deep Convolutional Network [10] to find exoplanets. The diagram below illustrates the Astronet Architecture briefly.

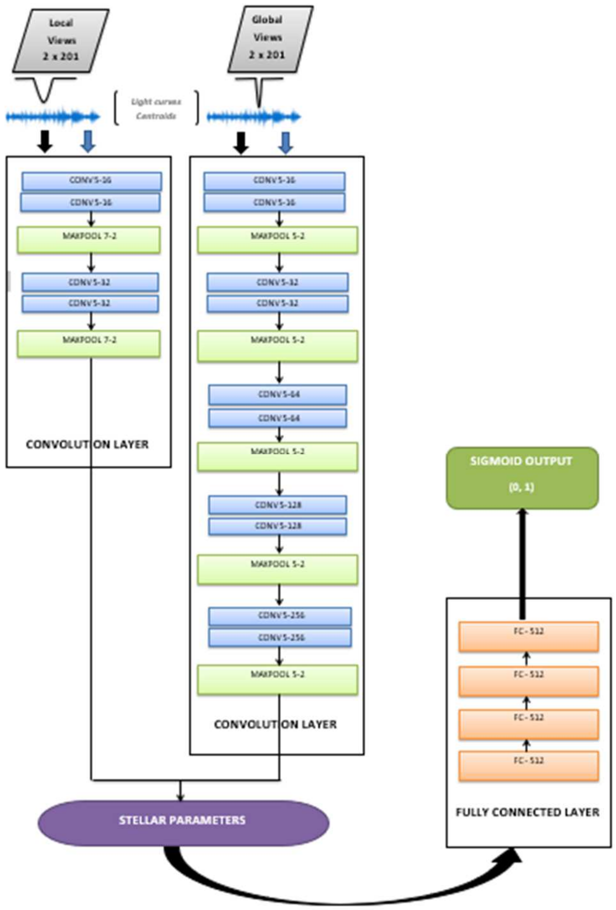

Fig 2. Astronet Model Architecture

The basic architecture is similar to the Convolutional Neural Network.

Step 1: Light Curves

The global and local views of each phase-folded TCE of Light curves [13] are used as the initial input. Each TCE represents a possible exoplanet transit with a particular period, epoch, and length.

Step 2: Convolution Layers

This section contains filters that assist in feature extraction. Every hidden layer of ASTRONET architecture uses ReLU. ReLU is a linear rectifier activation function that produces a linear graph by eliminating all negative values.

Step 3: Max Pooling

By proving an abstract form of representation, Max

Pooling reduces overfitting. It also decreases the number of parameters to learn and provides basic translation invariance to the internal representation, lowering the computational cost.

Step 4: Stellar Parameter

The performance of Max Pooling is combined and fed into stellar parameters. Stellar parameters are important concepts in science. Efficient temperature (Teff), surface gravity (log g), mass (m), density (d), metallicity, and other parameters are among them. These parameters form a set of standardized parameters for data validation.

Step 5: Fully Connected Layer

Stellar Parameter output is given as input to fully connected layers. It compiles the data extracted by previous layers to give the final output.

Step 6: Sigmoidal Output

In CNNs, the sigmoid function is the most common activation function. Finally, the astronet architecture finishes with a sigmoidal feature that generates a number of outputs (0,1). It either classifies the input as true positive exoplanet transit or as true negative exoplanet transit. The graph of the sigmoid function is in the form of an 'S,' and it has a finite limit.

S(x) = $1+\frac{1}{e^{-x}}$, where S(x) is the mathematical representation of a sigmoidal function.

Two Astronet variants can be designed and created during the training process.

1. *Augmented Astronet*
2. *Smaller Astronet*

These variants have fewer convolution layers and max-pooling layers. As a result, the model size is reduced, requiring less training time and computing resources.

D. PREDICTIONS

Hyperparameters: Hyperparameters are used to refine the output metrics' values. The required learning rate, epochs, and batch size are all taken into account here. The value may differ from model to model based on the amount of noise present in the dataset. In some cases, drop-out is also added to reduce model overfitting.

Once the model is built it is very important to evaluate how good the model performs. We will evaluate our model using the below matrix and then finally predict in which category the studied exoplanet belongs to i.e habitable or not habitable.

Performance Matrix or Confusion Matrix: This matrix is generated to evaluate the model performance of classification models. It shows the following values-

		PREDICTED CLASS	
		Class = Yes	Class = No
ACTUAL CLASS	Class = Yes	TP	FN
	Class = No	FP	TN

Fig 3. Confusion Matrix

True Positive (TP) - These are correctly predicted positive values that means prediction of actual class is yes and predicted class is also yes. E.g if the actual class says this exoplanet is habitable and the predicted class tells you the same thing.

True Negative (TN) - These are the correctly predicted negative values which means that the value of the actual class is no and value of predicted class is also no. E.g., if the actual class says this exoplanet is not habitable and the predicted class tells you the same thing.

False Positives (FP) – When actual class is no and predicted class is yes. E.g., if the actual class says this exoplanet is not habitable and the predicted class tells you it is habitable.

False Negatives (FN) – When actual class is yes but predicted class in no. E.g., if the actual class says this exoplanet is habitable and the predicted class tells you it is not habitable.

1. *Accuracy*—the number of accurate classifications made in a given period of time. It is the ratio of correctly predicted observations to the total observations.

Accuracy = TP+TN/TP+FP+FN+TN

2. *Recall or sensitivity* — Ratio of true planets identified correctly. Recall is the ratio of correctly predicted positive observations to all observations in actual class.

Recall = TP/TP+FN

3. *Exoplanet precision or positive predictive value*—the percentage of input TCEs identified as exoplanets that are actually planets. Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

Precision = TP/TP+FP

4. F1 score - F1 Score is the weighted average of Precision and Recall. It measures the test accuracy. It scores maximum value when it tends to reach 1 and minimum value when it tends to reach 0.

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

Optimizer: To render an algorithm's cost function as minimal as possible. The ASTRONET model recommends using Adam Optimizer (Adaptive Moment Estimation). Adam is a combination of two optimization techniques i.e RMSprop and Stochastic Gradient Descent with momentum. Adam optimizer is chosen as it has less memory requirement, more computational efficiency, easy to implement and well suited for problems with large datasets.

Mathematical approach: Adam algorithm first computes the gradient g_t w.r.t parameters Ө, then computes and stores first and second order moments of gradient, m_t and v_t respectively, as

$$mt = \beta 1 * mt + (1 - \beta 1) * gt$$

$$vt = \beta 2 * vt - 1 + (1 - \beta 2) * gt_2$$

Where β1 and β2 are hyper-parameters that ϵ [0,1], β1 = 0.9 and β2=0.999 are ideally considered. At t=1 m_0 and v_0 are zero. Since initially the values are biased towards zero, we counter it by updating it mt' and vt' as

$$mt' = mt/(1 - \beta t1)$$

$$vt' = vt/(1 - \beta t2)$$

Finally, the parameters updated are computed as

$$\theta t = \theta t - 1 - \alpha * mt'/(\sqrt{vt'} + \epsilon)$$

where is a small stability constant, with standard value of $\epsilon = 10^{-8}$.

Classification: Finally, the model would classify whether the studied planet is a habitable exoplanet or not.

Conclusion

Our ASTRONET model would be an automatic approach to examine catalogs of interplanetary objects. By learning to recognize abnormalities in these bodies in any set, you will be able to recognize anomalies in any set. This model will save researchers a lot of time by supplying them with useful knowledge in just a few seconds. Our mission is to dig deep into planetary bodies in order to discover factors that will enable life to sustain on those planets, as well as to collect a vast amount of data for research purposes.

The approach of deep learning has been accomplished to obtain the major objectives of Exoplanet classification and detection. Firstly, a complex deep learning model is proposed that will give accurate predictions. Secondly, Temperature, humidity, eccentricity, radius, weight, metallic materials, transit signals, and a number of other characteristics can all be used to determine whether or not an exoplanet is habitable and lastly, various performance measures are stated to evaluate the model predictions. To achieve these goals, a comprehensive ASTRONET will be developed and implemented.

References

[1] Suryoday Basak, Surbhi Agrawal, Snehanshu Saha, Abhijit Jeremiel Theophilus, Kakoli Bora, Gouri Deshpande, Jayant Murthy, "Habitability Classification of Exoplanets: A Machine Learning Insight ",2018

[2] Brychan Manry, George Sturrock, Sohail Rafiqi, "Machine Learning Pipeline for Exoplanet Classification", 2019

[3] Rajeev Mishra, "Predicting habitable exoplanets from NASA's Kepler mission data using Machine Learning", 2017

[4] Karan Hora, "Classifying Exoplanets as Potentially Habitable Using Machine Learning", 2018

[5] Megan Ansdell, Yani Ioannou, Hugh P. Osborn, Michele Sasdelli, (2018 NASA Frontier Development Lab Exoplanet Team), "Scientific Domain Knowledge Improves Exoplanet Transit Classification with Deep Learning", 2018

[6] Christopher J. Shallue1, Andrew Vanderburg, "Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90", 2018

[7] Piyush Gawade, Akshay Mayekar, Ashish Bhosale, Dr. Sanjay Jadhav, "Finding New Earths Using Machine Learning & Committee Machine", 2020

[8] Shivam Pratap Singh, Devendra Kumar Misra, "Exoplanet Hunting in Deep Space with Machine Learning", 2020

[9] Ricker, Winn, Vanderspek, et al. (2014), "Transiting Exoplanet Survey Satellite," Journal of Astronomical Telescopes, Instruments, and Systems, 014003.

[10] Shallue & Vanderburg (2017), "Identifying Exoplanets with Deep Learning," The Astronomical Journal, 155, 94.

[11] Saha, S., Agrawal, S., Manikandan, R., Bora, K., Routh, S., Narasimhamurthy, A.: ASTROMLSKIT: A New Statistical Machine Learning Toolkit: A Platform for Data Analytics in Astronomy. arXiv:1504.07865 [cs.CE] (April 2015)

[12] Classifying Exoplanets as Potentially … 211 10. Smith, M.R., Martinez, T.: Improving classification accuracy by identifying and removing instances that should be misclassified. In: The 2011 International Joint Conference on Neural Networks (IJCNN), pp. 2690–2697. (2011)

[13] Wolszczan, A., "Searches for planets around neutron stars", Celest. Mech. Dyn. Astr., vol. 68, p. 13, 1997.

[14] J. F. Kasting, D. P. Whitmire, R. T. Reynolds, Icarus 101, 108 (1993).

[15] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. Smote: Synthetic minority over-sampling technique. J. Artif. Int. Res. 16, 321–357

[16] Peng, N., Zhang, Y., Zhao, Y., 2013. A SVM-kNN method for quasar-star classification. Science China Physics, Mechanics and Astronomy 56, 1227–1234.

[17] Borucki, William J., et al.: Kepler planet-detection mission: introduction and first results. Science. 327(5968), 977–980 (2010)

[18] Bloom, J., Richards, J.: Data mining and machine-learning in time-domain discovery & classification. Adv. Mach. Learn. Data Min. Astron. (2011)

[19] Schulze-Makuch, D., Méndez, A., Fairén, A.G., von Paris, P., Turse, C., Boyer, G., Davila, A.F., António, M.R.D.S., Irwin, L.N., Catling, D.: A Two-Tiered Approach to Assess the Habitability of Exoplanet (2011)

[20] Petigura, E.A., Howard, A.W., Marcy, G.W.: Prevalence of Earth-size planets orbiting Sun-like stars. In: Proceedings of the National Academy of Sciences of the United States of America. arXiv: (October 31, 2013)