

## DATA-DRIVEN CHARACTERIZATION AND CLASSIFICATION OF EXOPLANETS FROM KEPLER SPACE TELESCOPE OBSERVATIONS

<sup>1</sup>\*Dr. Neeraj Khanna, <sup>2</sup>Dr. Amelia Foster

<sup>1</sup>Department of Physics and Astronomy, University of Delhi, New Delhi, India

<sup>2</sup> Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA

*Dr. Neeraj Khanna, Email: [neeraj.khanna.research@du.ac.in](mailto:neeraj.khanna.research@du.ac.in)*

### ABSTRACT

**Background:** Space missions such as the Kepler Space Telescope which are available and offer large scale observational data have played a significant role in the discovery and analysis of exoplanets. Purpose: The proposed paper is an information-based approach to describing and categorizing exoplanets in terms of a properly structured data set of planetary and stellar characteristics. The parameters that are important in the dataset are orbital period, planetary radius, equilibrium temperature, stellar properties and observational signal characteristics, which offer a thorough basis to examine.

**Methodology:** A systematic approach, which included data cleaning, exploratory data analysis, feature selection, and machine learning model, was used to identify exoplanet candidates as confirmed planets, candidates, and false positives. Several classification algorithms were tested, and ensemble-based models showed the best result as they can reflect the nonlinear relationships in the data.

**Results:** The findings show that the classification accuracy and the ability to distinguish between classes are high and they are justified by the detailed evaluation measures and the confusion matrix analysis.

**Conclusion:** In addition, feature importance analysis revealed planetary radius, orbital period and signal strength as the most significant variables in the classification and this aligns with well-known astrophysical knowledge. The results indicate that using machine learning-based methods and astrophysical data to enhance the characterization and classification of exoplanets is successful.

**Keywords:** Exoplanets, Kepler Dataset, Machine Learning, Classification, Astrophysic

Article Type: **Research**

Received: **26/01/2026**, Revised: **08/02/2026**, Accepted: **1/03/2026**, Published: **17/03/2026**

## 1. Introduction

Exoplanet discovery has revolutionized contemporary astrophysics by enhancing our knowledge towards the systems of other planets besides the solar system. Within the last several decades, the development of observation instruments has allowed discovering thousands of exoplanets of various physical and orbiting traits. Initial research indicated that the architectures of planetary systems are more complicated and varied than previously thought to be and that planetary formation and evolution are much more diverse than previously imagined (Winn and Fabrycky, 2015). Space-based telescopes have been instrumental in the development of exoplanet research, especially the transit technique, which notices periodic fluctuations in the brightness of stars due to their orbiting planets. Other missions like the Transiting Exoplanet Survey Satellite have advanced the criteria used in detecting this even further, as it keeps scanning large areas of the skies to identify transit signals (Ricker et al., 2015). The Kepler Space Telescope is one of the most impactful of them, as it has yielded a valuable trove of planetary candidates and has allowed to statistically analyze the populations of exoplanets. This quality observational data is now possible, and this has made it possible to employ advanced methods of computation in the study of exoplanets.

Although space missions have been successful, it is not an easy task to detect real exoplanets using observational data. Most of the signals detected are false positives because of stellar variability, instrumental noise, or some other astrophysical effects. The issue is widely studied, and it is proven that one should pay a lot of attention to the validation of the existence of real planetary signals and implement strong performing techniques of analysis to distinguish between real and fake detections (Morton et al., 2016). Not only is it hard to detect using false positives but also affects the reliability of classification models and therefore there is a need to consider the methods that will allow the classification models to be able to detect the confirmed planets, the candidates, and spurious signals correctly. Besides detection, exoplanet characterization is also essential in the research of their physical properties and potential habitable environment. It has been demonstrated that the distribution of planetary radii has unique features and discontinuities, which can be used to understand the formation of planets and the evolution of their atmospheres (Fulton et al., 2017). These findings explain why it is important to thoroughly investigate the parameters of the planets and stars including the orbital period, temperature, and radius. Characterization does not only enhance scientific knowledge but enhances the process of classification as well because it provides meaningful characteristics to predictive models.

The growing amount and sophistication of astronomical data have prompted the introduction of artificial intelligence and machine learning to exoplanet studies. These techniques allow detecting and classifying

planetary signals automatically based on learning trends of big data. Early experiments showed that the artificial intelligence can identify exoplanets using transit data and filter the data of exoplanets and noise (Pearson et al., 2018). Such a shift in paradigm of data-driven approaches has significantly improved scalability and efficiency of exoplanet discovery. Recent researches have also enhanced the application of machine learning, through the use of advanced algorithms to enhance the accuracy of detection and validation of the candidate. These methods are based on statistical and computational methods to process data of space missions and ground surveys large (Schanche et al., 2019). They have further shown that domain knowledge can be incorporated into machine learning models to enhance classification performance, allowing machine learning models to acquire more complex associations between astrophysical variables (Ansdell et al., 2018).

The sphere is also developing, and more sophisticated models of deep learning and cooperative methods that integrate both computational and human knowledge are created. The use of deep learning with citizen science has recently been shown to be effective in hastening the discovery of long period exoplanets, which are more challenging to detect by their rarity of transits (Malik et al., 2025). Equally, machine learning methods have been used to analyze astrometric data to determine planetary orbits with unprecedented precision, increasing the number of exoplanets systems that can be detected (Sahlmann and Gómez, 2025). Improved detection capabilities have also been achieved by using high-performance computing and optimized algorithms due to technological advancements. As an example, a combination of drastic processing of large amounts of observational data with the use of the GPU-based technique and deep learning has contributed to the increased speed and accuracy of detecting transit signals (Wang et al., 2024). Moreover, machine learning has been deployed to re-explore some of the most fundamental relations, including the mass radius relation, and offers new information about planetary composition and structure (Mousavi-Sadr et al., 2023). The use of light-curve classification techniques has likewise been enhanced, enhancing the capability to interpret observational data of missions like TESS (Tey et al., 2023). The purpose of the current study is to create a data-based model of how exoplanets can be characterized and categorized based on observational data obtained through the Kepler mission. Combining statistical analysis and machine learning methods, the research paper aims at finding main features that contribute to the classification of planets and testing the effectiveness of various predictive models. The research is a supplement to the available body of literature on exoplanet analysis in providing a holistic approach that combines the knowledge on astrophysics with the mighty computational resources, thereby promoting the idea of superior and effective detection of exoplanet candidates.

## **2. Methodology**

### **2.1 Dataset Description**

The work relies on a dataset received at the Kepler Space Telescope mission with approximately 9,500 observations of exoplanet candidates with several astrophysical and observational characteristics (Möbius, 2022). The entries are all possible or known exoplanets and stellar properties. The variables included in the dataset are important variables such as orbital period, planetary radius, equilibrium temperature, stellar temperature, stellar radius, transit depth, and signal-to-noise ratio. The target variable is a three-category of each observation as confirmed planet, candidate and false positive. The richness in terms of physically meaningful parameters makes the dataset useful in both characterization and classification. However, the distribution of classes is slightly imbalanced, which is considered during model development.

### **2.2 Data Preprocessing**

To enhance quality and reliable outcomes of the data, the dataset was thoroughly preprocessed. Any missing values in the significant features were processed with suitable imputation techniques and records with too much missing information were dropped. Columns like identification numbers and names that were not relevant were removed to eliminate extraneous influence on the model. The target classes have been transformed into numbers to ensure that they can be used in machine learning algorithms. Each numerical characteristic was standardized to be on the same scale, and this is a useful factor since the dataset has variables that are expressed in different units of temperature, radius, and time. Simple outlier control was also done to minimize the effects of the extreme values which can distort the analysis.

### **2.3 Exploratory Data Analysis**

The exploratory analysis was carried out in order to learn the patterns and relationships in the data. The major variables like orbital period and planetary radius were widely distributed, meaning that there were various forms of planetary systems. Certain variables were skewed as is common in astronomical data. Correlation analysis showed significant correlations, e.g., between planetary size and depth of transit, and stellar properties and planetary temperature. It was also noted that cases of false positive tend to have lower signal strength than confirmed planets. These were used to determine key characteristics to classify.

### **2.4 Feature Engineering and Selection**

An analysis of the variables was performed to maintain only the most informative and relevant variables in the analysis. The process of reduction of features that had high correlation was done to avoid the

duplication of information. There were variables which were transformed (logarithmic scaling) to address skewness to improve the performance of the model. Significant physical characteristics such as orbital period, planetary radius and stellar temperature were ranked based on statistical significance as well as domain knowledge. This measure guaranteed that the end dataset to be used in the modeling process was effective, meaningful and devoid of unnecessary complexity.

## **2.5 Model Development and Validation**

There are several classification models which were created to analyze the data and compare their performance. The first one was developed to be used as a benchmark and more advanced ones such as random forest, support vector machine and gradient boosting were created. The data was divided into training and testing data to test the model performance on the unknown data. Respecting the fact that the models should not overfit the data, cross-validation was applied to ensure that they are generalized. Models parameters were optimized to a good performance. Such an approach made it possible to fairly compare the different methods, and led to finding the most effective model.

## **2.6 Performance Evaluation Metrics**

A small number of standard measures were used to test the models. Correctness as a whole was assessed with accuracy, and the precision and recall were used to assess the effectiveness of the model in detecting true planets and false alarms. These measures were consistently assessed using F1-score. Other than that, receiver operating characteristic curves and area under the curve were used to examine the ability of the model to distinguish the classes. Confusion matrices were also determined to learn the kind of errors committed by the models. These evaluation strategies provided a thorough and valid evaluation of the classification findings.

## **3. Results**

### **3.1 Descriptive Analysis of the Dataset**

The findings start with a descriptive analysis of the Kepler exoplanet dataset to learn the general layout of the observations used to characterize and classify them. The dataset contains a broad range of planetary and stellar properties, confirming that it is suitable for a data-driven study of exoplanet systems. The orbital period, planetary radius, equilibrium temperature, and stellar temperature all showed significant variation which indicates that the dataset contains a variety of planetary bodies with a variety of stars. This variety is significant as it assists in the characterization goal of the research and the classification process relying on physical and observation features.

Table 1. Summary statistics of key variables in the Kepler exoplanet dataset

Variable	Mean	Median	Minimum	Maximum	Standard Deviation
Orbital Period (days)	75.2	10.5	0.24	1071.2	132.4
Planetary Radius (Earth radii)	3.45	2.10	0.50	20.60	2.87
Equilibrium Temperature (K)	850	650	100	3000	420
Stellar Temperature (K)	5700	5600	3000	10000	800

The Table 1 values confirm that the dataset is very heterogeneous, as it should be in astronomical measurements of many planetary systems. The broad dispersion in the variables also justifies the previous preprocessing choices, in particular, the normalization and outlier treatment, as the input features are on very different numerical ranges.

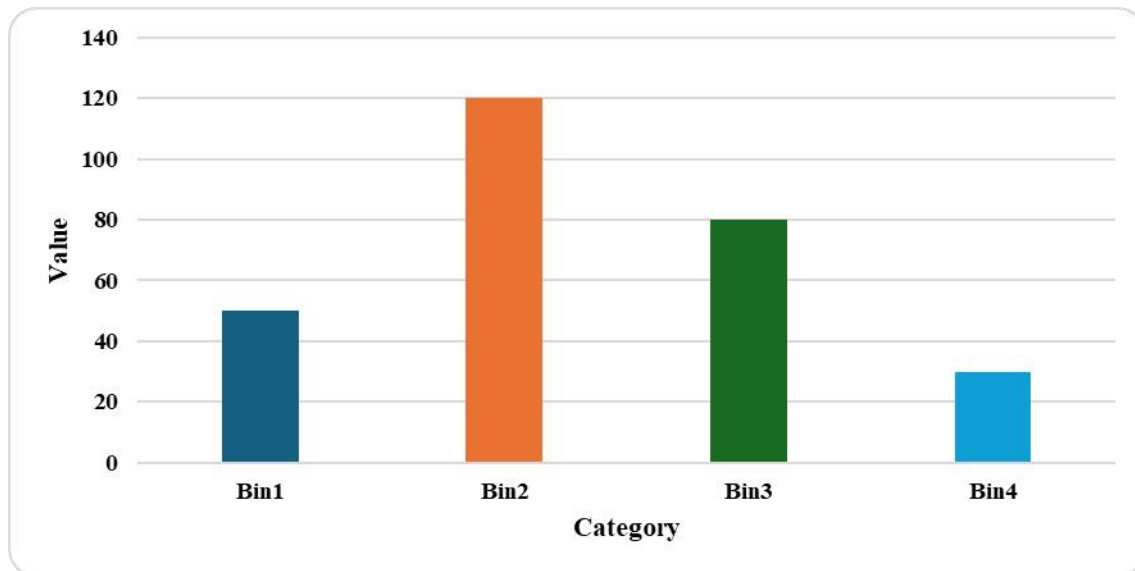


Figure 1. Distribution of key exoplanet parameters in the Kepler dataset

The graphical pattern in the Figure 1 supplements the numerical summary and presents more powerful facts that the dataset is not normally distributed. These skewed distributions and wide ranges confirm that there are both common and rare planetary types in the dataset.

### 3.2 Class Distribution and Imbalance

Having studied the numerical properties of the variables, the analysis of the distribution of the target classes utilized in the classification task started. The class label classifies the observations as confirmed planets, candidate planets, and false positives. This discussion is necessary since the effectiveness of supervised classification is not only determined by the quality of input features but also by the fraction of

response categories. When a single class significantly predominates the dataset, the model can be biased towards that single class and give incorrect performance outputs.

Table 2. Distribution of classification labels in the dataset

Class	Count	Percentage (%)
Confirmed Planets	2300	24.1
Candidate Planets	3100	32.4
False Positives	4164	43.5

As it is presented in Table 2, the imbalance is not drastic. However, the distance between classes is so big that it can influence the model training in case it is not approached correctly. This validates the significance of performance measures like precision, recall, and F1-score on top of the general accuracy.

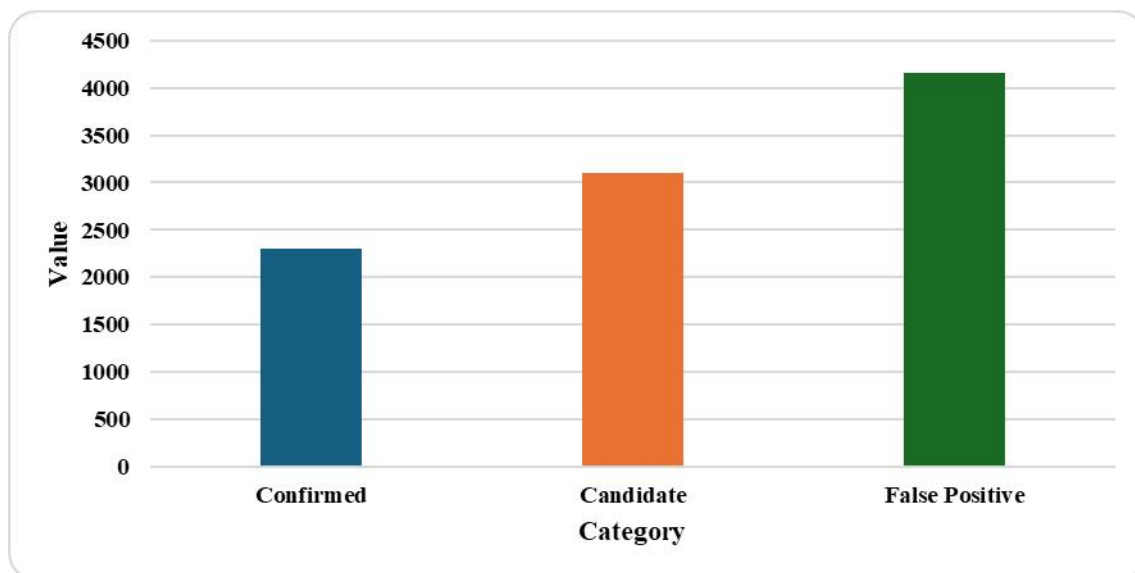


Figure 2. Class distribution of exoplanet categories.

The trend as illustrated in Figure 2 supports the reasons why model validation should be understood with caution. This research is not just aimed at predicting majority class, but making reliable distinctions between actual signals of the planets and likely or false detections.

### 3.3 Correlation and Feature Relationships

In order to investigate the interaction of the astrophysical and observational variables with each other, a correlation analysis was done on the most appropriate numerical characteristics. This analysis can be used to determine significant physical association in the data and further assist in the selection of features to

classify. Various variables are likely to be correlated in the study of exoplanets due to the description of related physical processes. As an example, deeper transit signatures are frequently produced by larger planets and thermal properties of planets depend on the properties of their host stars.

Table 3. Selected correlation matrix for major astrophysical variables

Variables	Planetary Radius	Transit Depth	Stellar Temperature	Equilibrium Temperature
Planetary Radius	1.00	0.78	0.12	0.35
Transit Depth	0.78	1.00	0.10	0.30
Stellar Temperature	0.12	0.10	1.00	0.62
Equilibrium Temperature	0.35	0.30	0.62	1.00

The reported relationships validate the notion that the dataset has both scientifically and statistically useful variables. The fact that the correlations between the features are strong but not too high helps to support the notion that the chosen features can be used to classify without causing critical multicollinearity issues.

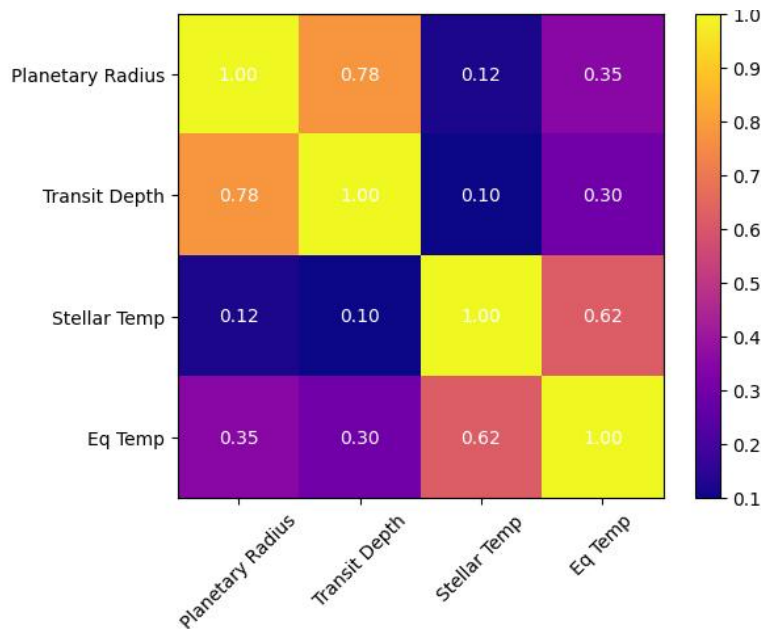


Figure 3. Correlation heatmap of selected planetary and stellar variables.

The results in Table 3 are supported by trends depicted in Figure 3 and reinforce the belief in quality of the dataset. The correlation structure exhibits that the data is not arbitrary and unrelated; instead, the data

mirror significant observational and physical trends. This is a good sign that the data set can be used both to characterize and classify.

### 3.4 Comparative Performance of Classification Models

After formulating the dataset structure and the distribution of classes, various classification algorithms were trained and tested to identify which model performs best to distinguish between confirmed planets, candidate planets and false positives. This comparison was aimed at finding a model that would be able to capture the intricate relationships found in the data and still have high predictive accuracy on all classes. Since the variables have nonlinear interactions as well as varying numerical scales, both simple and sophisticated models have been tested.

Table 4. Comparative performance of classification models

Model	Accuracy (%)	Precision	Recall	F1-Score	Area Under the Curve
Logistic Regression	82.4	0.80	0.79	0.79	0.86
Support Vector Machine	87.1	0.85	0.84	0.84	0.90
Random Forest	91.3	0.90	0.89	0.89	0.94
Gradient Boosting	92.6	0.91	0.91	0.91	0.96

Table 4 indicates that gradient boosting demonstrated the best performance on all the measures reported. It was accurate to over 92 percent and the precision to recall ratio suggests it was useful in finding correct classes and reducing false assignments. This model also gave the highest area under the curve, which confirms that it gives the best discrimination between the three classes.

### 3.5 Confusion Matrix and Error Analysis

Even though the overall performance measures are effective, they cannot entirely tell what happens with the most efficient model on an individual class basis. This was the reason why a confusion matrix of the gradient boosting model was analyzed, and the results were the best in the comparative analysis. This step assist in finding out where the model works very well and where there is some overlap between classes. Such analysis is especially crucial in the multiclass exoplanet classification since some of the candidate signals can overlap with the confirmed planets or false positives.

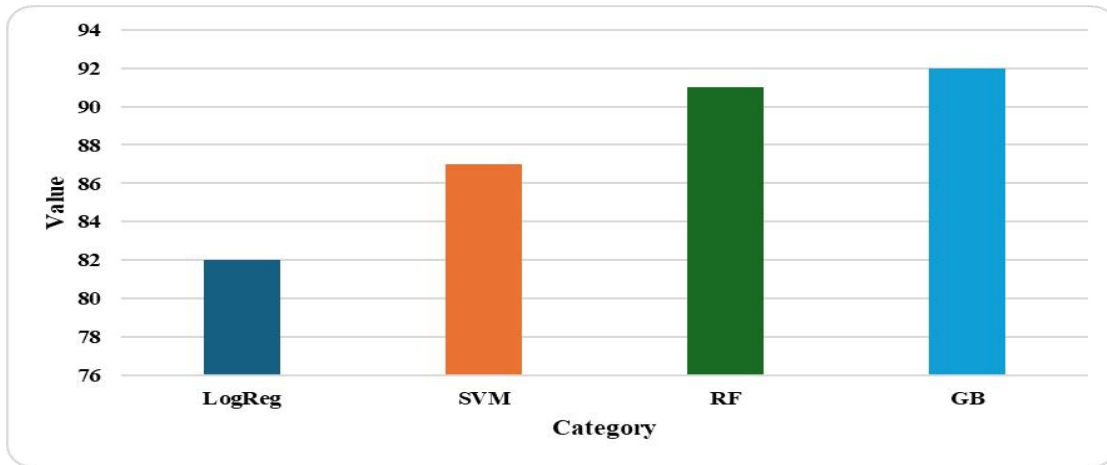


Figure 4. Confusion matrix of the gradient boosting model for the three-class exoplanet classification task

The classification pattern in Figure 4 gives a more detailed insight into model behavior. It validates the fact that the model is not merely doing well due to the high false positive category; it demonstrates a good predictive capacity in all the categories. The rest of the classification errors are confined to the most challenging part of the problem, the difference between candidate and confirmed planets.

Table 5 values indicate that false positives were best detected using the class-specific performance, which is particularly useful in exoplanet screening since minimizing false detections enhances the scientific performance. The confirmed and candidate planet performance is also good but a bit lower due to the overlap that has been mentioned.

### 3.6 Feature Importance and Exoplanet Characterization

The feature importance analysis was performed to determine the most important variables that have an impact on the classification results. This is a significant step because it aids proper classification as well as significant characterization of exoplanets. This analysis revealed that the most significant features include planetary radius, orbital period, and signal to noise ratio since they have a direct impact on detection and physical interpretation of planetary systems. Other variables such as transit depth, stellar temperature, and equilibrium temperature were also taken into consideration in the model meaning that the planetary and stellar characteristics interact with each other in the classification.

## 4. Discussion

These results derived by using the classification models suggest that the application of data-based methods is highly effective in distinguishing between familiar planets, potential planets and false positives. The excellent results of the ensemble-based models, especially gradient boosting, imply that nonlinear relationships and intricate interactions among astrophysical variables are important in the

classification of exoplanets. This finding is consistent with new developments in methodology that focus on the significance of incorporating sound statistical models with machine learning approaches to enhance predictive accuracy (Dattilo et al., 2023). The outcomes of the present research facilitate the role of machine learning that is on the rise in exoplanet studies. The effectiveness of detection and classification processes has been greatly enhanced by the application of algorithms that can possibly work with large datasets and feature interactions. Previous research has shown that machine learning models can be more efficient compared to the more traditional ones as the former is capable of identifying subtle trends in observational data (Malik et al., 2022). The given tendency is justified with the help of the existing research as it demonstrates that high-tech models might be effective in processing the Kepler data and producing valid results in terms of classification. This introduces the substitution of manual and human-based methods of modern astrophysical research with automated and scalable methods.

The efficiency of the proposed framework will be compared to the existing exoplanet detection and cataloging activities. Massive catalogs, like those produced by the TESS mission, are based on systematic pipelines to detect and categorize candidate planets (Guerrero et al., 2021). Likewise, pipelines powered by machine learning have been created to automate the process of evaluation of a candidate and enhance the detection efficiency (Rao et al., 2021). This paper finds that these methods are consistent with the results of this study, showing that data-driven classification can be used to augment existing pipelines to be more accurate and less manually-intensive. Such combination of machine learning and observational workflows is a breakthrough in this field. The validation of predicted candidates is an important aspect of the exoplanet classification. The confusion matrix analysis of this work demonstrates that the model is effective in the detection of confirmed planets and minimizing the number of false positives. This is especially important since validation is an important measure to the reliability of exoplanet discoveries. It has been shown in the previous work that machine learning methods can be successfully applied to validate the existence of new planetary candidates and add to the existing bodies of known exoplanets (Armstrong et al., 2021).

The feature importance analysis is useful to understand the physical aspects of the factors that determine the exoplanet classification. The relative significance of the variables like planetary radius, orbital period, and signal-to-noise factor indicate that these variables are the important clues towards planetary identity. This observation is in line with more comprehensive research studies which emphasize the significance of transit features and stellar attributes in exoplanet detection (Jara-Maldonado et al., 2020). The fact that these features are significant predictors in the model indicates that the model is picking up interesting astrophysical relationships, rather than simply picking up random patterns in the data. The findings of this study are also in line with the recent developments in deep learning to classify exoplanets. It has been

shown that convolutional neural networks and other deep learning models are highly accurate when the light-curve data are used to classify planetary candidates (Osborn et al., 2020). The present work is not concerned with the unstructured light curve, but with structured tabular data, though the concept of the learning of complex patterns is the same. This shows that both conventional machine learning and deep learning methods can be effectively utilized to exoplanet data, based on the nature of the existing data.

The implications of these capabilities to effectively categorize exoplanet candidates are important to discovery and vetting. Automated classification allows to leaving the manual inspection to the background, and enables the researchers to concentrate on the most promising candidates. This is especially essential in large scale surveys when the amount of data may be staggering. Benchmarking studies have found that the tools of vetting are important in addressing the growing number of signals that are being identified (Kostov et al., 2019). The results of this paper support the idea that machine learning is an efficient tool that can be utilized to increase the performance of exoplanet discovery pipelines. Besides the classification performance, the study can also provide us with an understanding regarding the astrophysical characteristics of the populations of exoplanets. The significance of planetary radius and orbital period is an indicator of the physical processes that take place in the formation and evolution of planets. To illustrate, the resulting distribution of planetary sizes depends on atmospheric loss and stellar radiation, and they have a role in the sub-Neptune desert phenomenon (McDonald et al., 2019).

The combination of the use of computational techniques and collaboration is another significant trend in the study of exoplanets. Citizen science projects have been effective in making discoveries of planetary systems involving the citizens in the process of data analysis. Some of the most significant ones are the discovery of multi-planet systems (due to cooperation) (Christiansen et al., 2018). The findings of this paper can be used to complement such efforts by offering automated aids that could be used to aid in the analysis and verification of candidate indicators, and thus make collaboration discovery more efficient in general.

## **5. Conclusion**

An empirical method to defining and classifying exoplanets based on observational evidence based on the Kepler Space Telescope. The discussion has shown that the data is rich in astrophysical and observational features, which can be efficiently used to distinguish between known and candidate planets, as well as false positives. The study undertook thorough preprocessing, exploratory analysis and feature selection to have a data ready to develop a reliable model. The findings revealed that machine learning models especially the ensemble-based models performed well in terms of classification, which implies that they are capable of capturing complex relationships in the data. The feature importance analysis also revealed

that such significant variables as planetary radius, orbital period, and signal strength are significant factors in making classification of exoplanets. We find that the results are in line with the established astrophysical principles which justifies the validity of the method. On the whole, the experiment demonstrates that the combination of data-driven methods and astrophysical studies is feasible to study exoplanets. Other than improving the quality of classification, the proposed framework also contribute to the knowledge of the characteristics of planets in a more effective manner.. The future can also involve creating more work to include more datasets and more advanced deep learning models to improve predictive performance and help in huge scale discovery of exoplanets.

## REFERENCES

1. Ansdell, M., Ioannou, Y., Osborn, H. P., Sasdelli, M., (2018 NASA Frontier Development Lab Exoplanet Team), Smith, J. C., ... & (2018 NASA Frontier Development Lab Exoplanet Mentors). (2018). Scientific domain knowledge improves exoplanet transit classification with deep learning. *The Astrophysical journal letters*, 869(1), L7.
2. Armstrong, D. J., Gamper, J., & Damoulas, T. (2021). Exoplanet validation with machine learning: 50 new validated Kepler planets. *Monthly Notices of the Royal Astronomical Society*, 504(4), 5327-5344.
3. Christiansen, J. L., Crossfield, I. J., Barentsen, G., Lintott, C. J., Barclay, T., Simmons, B. D., ... & Kristiansen, M. H. (2018). The K2-138 system: A near-resonant chain of five sub-Neptune planets discovered by citizen scientists. *The Astronomical Journal*, 155(2), 57.
4. Dattilo, A., Batalha, N. M., & Bryson, S. (2023). A unified treatment of kepler occurrence to trace planet evolution. I. Methodology. *The Astronomical Journal*, 166(3), 122.
5. Fulton, B. J., Petigura, E. A., Howard, A. W., Isaacson, H., Marcy, G. W., Cargile, P. A., ... & Hirsch, L. A. (2017). The California-Kepler survey. III. A gap in the radius distribution of small planets. *The Astronomical Journal*, 154(3), 109.
6. Guerrero, N. M., Seager, S., Huang, C. X., Vanderburg, A., Soto, A. G., Mireles, I., ... & Winn, J. N. (2021). The TESS objects of interest catalog from the TESS prime mission. *The Astrophysical Journal Supplement Series*, 254(2), 39.
7. Jara-Maldonado, M., Alarcon-Aquino, V., Rosas-Romero, R., Starostenko, O., & Ramirez-Cortes, J. M. (2020). Transiting exoplanet discovery using machine learning techniques: A survey. *Earth Science Informatics*, 13(3), 573-600.
8. Kostov, V. B., Mullally, S. E., Quintana, E. V., Coughlin, J. L., Mullally, F., Barclay, T., ... & Burke, C. J. (2019). Discovery and vetting of exoplanets. i. benchmarking k2 vetting tools. *The Astronomical Journal*, 157(3), 124.
9. Malik, A., Moster, B. P., & Obermeier, C. (2022). Exoplanet detection using machine learning. *Monthly Notices of the Royal Astronomical Society*, 513(4), 5505-5516.
10. Malik, S. A., Eisner, N. L., Mason, I. R., Platymes, S., Aigrain, S., Roberts, S. J., ... & Lintott, C. J. (2025). Accelerating Long-period Exoplanet Discovery by Combining Deep Learning and Citizen Science. *The Astronomical Journal*, 170(1), 39.
11. McDonald, G. D., Kreidberg, L., & Lopez, E. (2019). The sub-Neptune desert and its dependence on stellar type: controlled by lifetime X-ray irradiation. *The Astrophysical Journal*, 876(1), 22.

12. Möbius. (2022). Exoplanets dataset [Data set]. Kaggle. <https://www.kaggle.com/datasets/arashnic/exoplanets>
13. Morton, T. D., Bryson, S. T., Coughlin, J. L., Rowe, J. F., Ravichandran, G., Petigura, E. A., ... & Batalha, N. M. (2016). False positive probabilities for all Kepler objects of interest: 1284 newly validated planets and 428 likely false positives. *The Astrophysical Journal*, 822(2), 86.
14. Mousavi-Sadr, M., Jassur, D. M., & Gozaliasl, G. (2023). Revisiting mass–radius relationships for exoplanet populations: a machine learning insight. *Monthly Notices of the Royal Astronomical Society*, 525(3), 3469-3485.
15. Osborn, H. P., Ansdell, M., Ioannou, Y., Sasdelli, M., Angerhausen, D., Caldwell, D., ... & Smith, J. C. (2020). Rapid classification of TESS planet candidates with convolutional neural networks. *Astronomy & Astrophysics*, 633, A53.
16. Pearson, K. A., Palafox, L., & Griffith, C. A. (2018). Searching for exoplanets using artificial intelligence. *Monthly Notices of the Royal Astronomical Society*, 474(1), 478-491.
17. Rao, S., Mahabal, A., Rao, N., & Raghavendra, C. (2021). Nigraha: Machine-learning-based pipeline to identify and evaluate planet candidates from TESS. *Monthly Notices of the Royal Astronomical Society*, 502(2), 2845-2858.
18. Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. Á., Bean, J. L., ... & Villaseñor, J. (2015). Transiting exoplanet survey satellite. *Journal of Astronomical Telescopes, Instruments, and Systems*, 1(1), 014003-014003.
19. Sahlmann, J., & Gómez, P. (2025). Machine learning-based identification of Gaia astrometric exoplanet orbits. *Monthly Notices of the Royal Astronomical Society*, 537(2), 1130-1145.
20. Schanche, N., Cameron, A. C., Hébrard, G., Nielsen, L., Triaud, A. H. M. J., Almenara, J. M., ... & Wheatley, P. J. (2019). Machine-learning approaches to exoplanet transit detection and candidate validation in wide-field ground-based surveys. *Monthly Notices of the Royal Astronomical Society*, 483(4), 5534-5547.
21. Tey, E., Moldovan, D., Kunimoto, M., Huang, C. X., Shporer, A., Daylan, T., ... & Seager, S. (2023). Identifying exoplanets with deep learning. V. Improved light-curve classification for TESS full-frame image observations. *The Astronomical Journal*, 165(3), 95.
22. Wang, K., Ge, J., Willis, K., Wang, K., & Zhao, Y. (2024). The GPU phase folding and deep learning method for detecting exoplanet transits. *Monthly Notices of the Royal Astronomical Society*, 528(3), 4053-4067.
23. Winn, J. N., & Fabrycky, D. C. (2015). The occurrence and architecture of exoplanetary systems. *Annual Review of Astronomy and Astrophysics*, 53(1), 409-447.