International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934

REVIEW ARTICLE

A study on unveiling the secrets of exoplanet hunting

Preeti Kalra[1], Abhendra PratapSingh[2], Arushi Gupta[1], Kartik Gupta[1], Prachi Jindal[1]

[1]*Department of Computer Science and Engineering, HMRITM, New Delhi, India*
[2]*Department of Mechanical Engineering College, HMRITM, New Delhi, India*

Article Information

Received: 09 November 2023
Revised: 05 December 2023
Accepted: 18 December 2023
Available online: 21 December 2023

Keywords:

Machine Learning
KNN Algorithm
Flux Intensity
Transit Method.

Abstract

Exoplanets are celestial bodies outside our solar system, orbiting stars other than our sun. They come in various sizes, compositions, and distances from their host stars. Scientists detect exoplanets by observing the dimming of a star's light as a planet passes in front of it (transit method), measuring the gravitational tug a planet exerts on its star (radial velocity method), or directly capturing their images. Studying exoplanets is a captivating area of astronomy that unveils our universe's vast array of planetary systems. Researchers delve into these distant worlds to explore their atmospheres, structures, and potential habitability, aiming to gather insights into the likelihood of life existing beyond our solar system. This paper addresses the challenges of distinguishing exoplanet signals from stellar activity, instrumental noise, and other astrophysical phenomena. It discusses the role of data-driven approaches, machine learning, and advanced statistical analyses in enhancing the reliability and accuracy of exoplanet detections after applying KNN.

1. Introduction

Exploring exoplanets, celestial bodies, and orbiting stars beyond our solar system is one of modern astronomy's most captivating and transformative endeavors. Scientists have embarked on a remarkable journey employing innovative techniques, cutting-edge technologies, and interdisciplinary approaches to unveil the mysteries of these distant worlds [1]. The burgeoning field of exoplanet hunting has expanded our cosmic perspective and redefined our understanding of planetary systems and the potential for extraterrestrial life. This research endeavors to comprehensively explore the diverse methodologies employed in detecting and characterizing exoplanets, a pursuit that has witnessed unprecedented advancements in recent decades. The discovery of the first confirmed exoplanet in 1992 ignited a revolution in astrophysics, sparking an exponential rise in detections and propelling the field into a new era of discovery [2]. Exploring exoplanets, celestial bodies, orbiting stars beyond our solar system is one of modern astronomy's most captivating and transformative endeavors. Scientists have embarked on a remarkable journey employing innovative techniques, cutting-edge technologies, and interdisciplinary approaches to unveil the mysteries of these distant worlds. The burgeoning field of exoplanet hunting has expanded our cosmic perspective and redefined our understanding of planetary systems and the potential for extraterrestrial life [3]. This study aims to provide valuable insights for researchers in the field of Exoplanet Hunting using Siamese network architectures. The research explores several critical questions about implementing Various Algorithms for accurate and efficient hunting of the planets [4]. Hot Jupiters, a category of exoplanets resembling Jupiter in size, exhibit significantly higher temperatures due to their proximity to their parent stars. These planets are notable for their brief orbital periods, completing orbits around their host stars in a few days, leading to elevated temperatures due to

Corresponding author: Abhendra PratapSingh
Email Address: abhendrapratap94@gmail.com
https://doi.org/10.36037/IJREI.2024.8102

intense stellar radiation [5]. The discovery of Hot Jupiter's has posed challenges to the traditional understanding of planetary formation, as their formation and migration mechanisms still need to be completed. These planets might have originated farther from their host stars and migrated inward over time due to interactions with the protoplanetary disk or other planets.[6] The study of Hot Jupiter's has yielded valuable insights into the rich diversity of planetary systems beyond our solar system. The transit method is a highly effective technique used in astronomy to detect exoplanets and planets outside our solar system. It involves observing periodic dimming in the light of distant stars caused by a planet passing in front of its host star as seen from Earth. This Method relies on the temporary decrease in a star's brightness when a planet crosses between the star and our line of sight [7]. These periodic dimming, known as transit events, offer astronomers valuable insights into the existence and characteristics of exoplanets. By carefully analyzing the patterns of these periodic dimming, scientists can deduce essential details about the orbit, size, and sometimes even the atmosphere of the transiting exoplanet.[8] The transit method has significantly expanded our understanding of the diversity of planetary systems beyond our own. This technique has been instrumental in various space missions and observations, such as NASA's Kepler and TESS missions [9]. Its success in identifying thousands of exoplanets has been pivotal in shaping our current knowledge of the universe and the prevalence of planets in distant solar systems. In addition to the transit method, astronomers utilize diverse approaches such as radial velocity, gravitational microlensing, direct imaging, and astrometry to seek out these elusive cosmic objects. Each method has strengths, enabling scientists to unveil exoplanets with different sizes, compositions, and orbital behaviors. The known catalog of exoplanets showcases a remarkable array of sizes, compositions, and orbital characteristics [10]. These discoveries challenge established theories about planetary formation and evolution from gas giants to rocky terrestrial planets. Furthermore, the discovery of exoplanets within the habitable zones of their host stars, where conditions might allow liquid water to exist, raises intriguing possibilities for life beyond our solar system [11]. This research aims to delve into the intricacies of exoplanet detection and analysis, exploring the methodologies, discoveries, challenges, and prospects in this captivating field of study [12]. By examining technological advancements, data analysis methods, and theoretical frameworks driving exoplanet research, this paper contributes to our ongoing quest to understand these distant worlds and identify environments capable of supporting life beyond our solar system [13].

2. Methods and Tools

2.1 Data Preprocessing

Data Collection: Gather comprehensive and relevant data from diverse sources aligned with the project's goals.[14] Data Cleaning: Address missing values and handle outliers to ensure data quality and model robustness. Exploratory Data Analysis (EDA): Analyze dataset characteristics, distributions, and correlations through visualization techniques (e.g., histograms, scatter plots) for pattern identification [15].Feature Selection and Engineering: Identify and engineer features significantly impacting the model's predictive ability. Perform scaling, transformation, or creation of new features to enhance model performance. Handling Categorical Data: Convert categorical variables into numerical formats suitable for algorithms, employing label or one-hot encoding methods—data Scaling and Normalization: Scale numerical features to maintain consistency in feature importance during model training. Normalize data if required, especially when features have varying scales or distributions. Data Splitting: Divide the dataset into training, validation, and test sets for model training, hyperparameter tuning, and performance. Evaluation Dealing with Imbalanced Data (if applicable): Address class imbalances in classification tasks using oversampling, under sampling, or synthetic data generation techniques. Preprocessing Pipelines: Develop pipelines to automate and standardize preprocessing steps, ensuring consistency in handling new data KNN-Specific Data Preprocessing: Normalize or scale numerical features for KNN, as it relies on distance calculations [16].

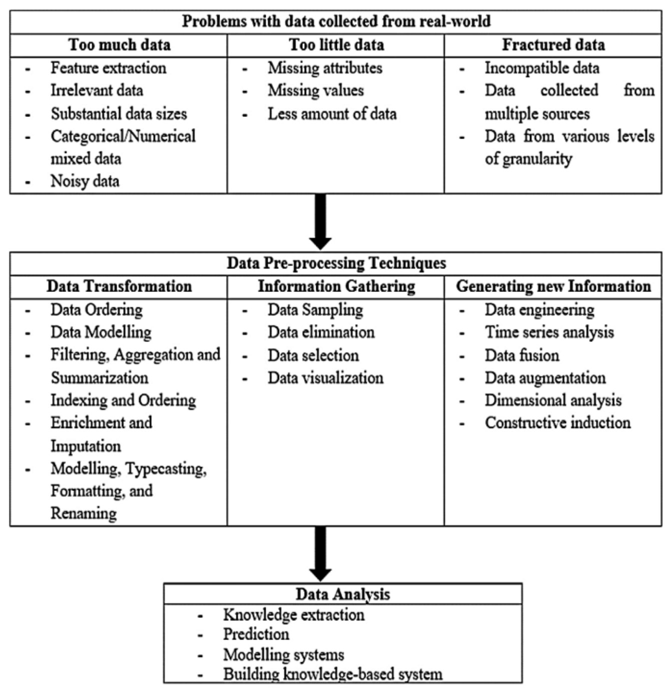

Figure 1: Why to carry data preprocessing [17]

2.2 KNN Algorithm

K-Nearest Neighbors (KNN) is a versatile algorithm within supervised learning, adept at classification and regression tasks [18]. It operates non-parametric and lazy learning, avoiding rigid assumptions about underlying data distribution and eschewing explicit model creation during training. Instead, KNN retains training data instances to inform predictions for new data points. KNN identifies the closest neighbors in the training set for a given data point using a chosen distance metric (e.g., Euclidean distance, Manhattan distance) [19]. "K" denotes the number of neighbors considered. For classification, KNN determines the majority class among the nearest neighbors to assign a class label via majority voting. In regression tasks, KNN predicts a continuous value for new data points by averaging the values of their K nearest neighbors. KNN's efficacy depends on K's choice and the distance metric selection. Computationally demanding for large datasets due to extensive distance calculations [20]. Sensitivity to irrelevant or noisy features. Lacks learning during training, leading to slower predictions for new instances than model-based algorithms. Determining the number of neighbors (K) impacts model behavior and performance [21]. Distance Calculation: Measure distances between the new data point and all training set points using a specified metric. Neighbor Identification: Select the K nearest neighbors based on computed distances. Voting or Aggregation: Identify the most common class among neighbors for classification. For regression, aggregate neighbor values (e.g., average). Prediction Assignment: Assign predicted Class labels or values to new data points based on majority voting or aggregation [22]. KNN is a supervised learning algorithm. Here is a revised explanation without plagiarism: In supervised learning, the concept mirrors a teaching scenario, where a knowledgeable guide (the algorithm) learns from a set of labeled examples, much like a student learns from a teacher who provides clear guidance [23]. Consider a teacher educating a child about animals. To teach the child what an elephant looks like, the teacher shows pictures of elephants and animals that are not elephants, such as zebras or monkeys. Each time the teacher shows an image, they mention whether it is an elephant. Through repetition and correction, the child starts associating specific features with the label "elephant" or "not elephant."[24]. Similarly, in supervised machine learning, the algorithm learns from a dataset containing examples paired with correct answers. Each example includes input data (features) alongside its corresponding output (label). The algorithm learns from this labeled dataset, associating the input features with the correct output, much like the child associating images with the correct labels after repeated exposure to examples. Just as a teacher corrects a child's mistakes during learning sessions, the algorithm refines its predictions based on the errors it makes while predicting the labels [25]. With continuous adjustment and refinement, the algorithm improves its ability to make accurate predictions, akin to how a student improves with guidance and practice. Once trained, the supervised learning model applies its learned knowledge to make predictions or classifications for new, unseen data, akin to how a child identifies an elephant correctly in a new picture based on previous learning experiences [26]. This supervised learning approach is widely applied across various fields, including healthcare (such as tumor prediction), image recognition, natural language processing, and recommendation systems, wherever labeled data is available to train predictive models [27].

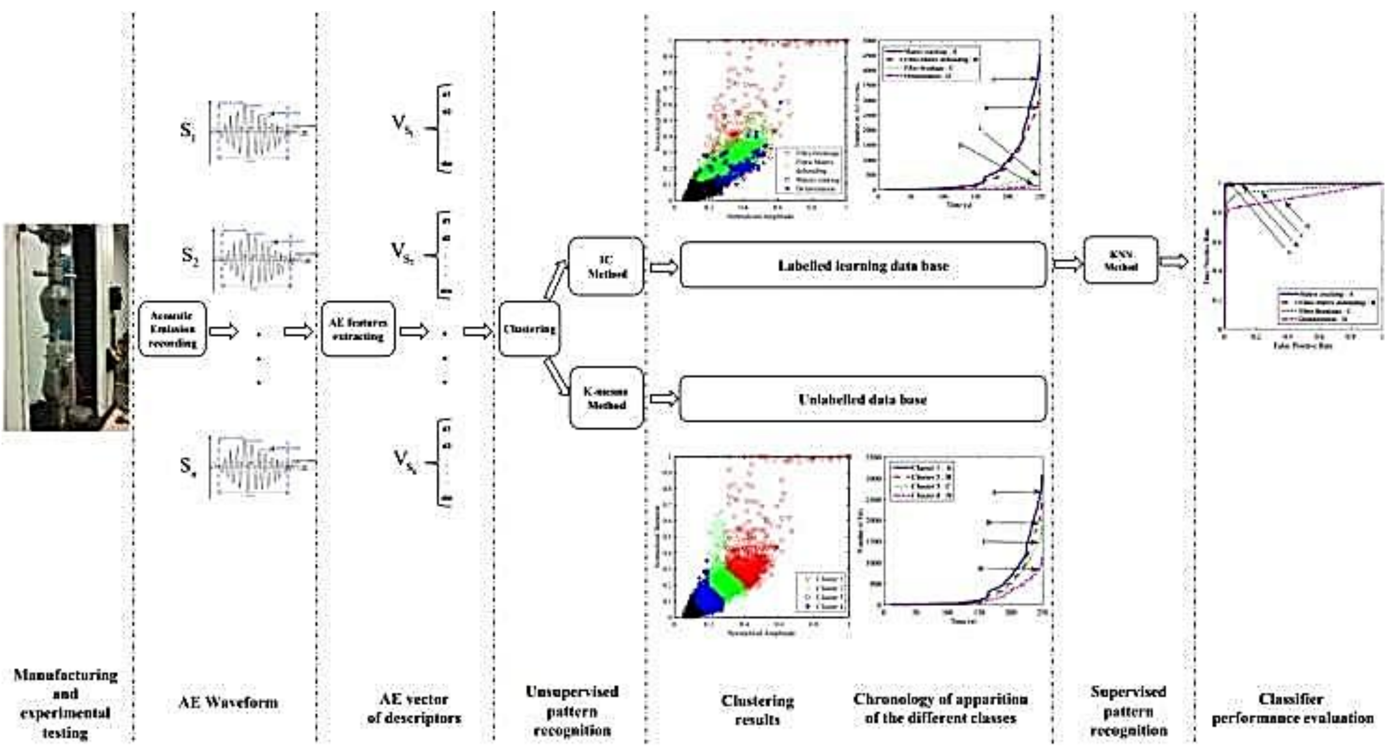

Figure 2: KNN framework [28]

2.3 Transit Method

The transit method is a crucial astronomical technique for identifying exoplanets orbiting stars outside our solar system. This approach relies on the repeated dimming of a star's light, occurring when an orbiting planet passes in front of it as observed from Earth. Detecting Transits: When an exoplanet moves between its host star and our line of sight, it partially obscures the star's light, leading to a noticeable decrease in brightness. This phenomenon, known as a "transit," provides significant data about the exoplanet, including its size and orbit characteristics. Repetition of Transit Events: If the exoplanet's orbit aligns with our perspective, regular dips in brightness occur at predictable intervals—these recurring decreases in brightness. The exoplanet's orbit aligns with our perspective; regular dips in brightness occur at predictable intervals. These recurring decreases in brightness confirm the presence of an exoplanet orbiting the star. Data Analysis: Scientists analyze the observed light curve, the graph depicting changes in the star's brightness over time. By examining these transits' duration, depth, and frequency, astronomers can estimate crucial details about the exoplanet, such as its size and orbital period. Confirmation of an exoplanet's existence via the transit method prompts additional observations using other techniques like radial velocity measurements or spectroscopy. These methods help gather more insights into the exoplanet's properties, such as its mass and atmospheric composition. The transit method has been instrumental in discovering numerous exoplanets, particularly those positioned closer to their host stars. It is beneficial in identifying exoplanets within the 'habitable zone.NASA's quest to explore exoplanets through observing planetary transits has been groundbreaking. The Kepler mission, aimed at unveiling the diversity and configurations of exoplanetary systems, spanned nine years and unveiled a trove of discoveries. It confirmed numerous exoplanets, with many potential candidates awaiting confirmation due to the vast dataset collected. Following Kepler's success, TESS took the baton. Currently, on a two-year mission, TESS seeks to uncover potentially thousands more transiting exoplanets. It focuses on bright stars near our solar system, enriching our knowledge of exoplanets in our cosmic neighborhood. Kepler's observation concentrated on a specific sky section during its primary mission. In contrast, TESS operates on a grander scale, scanning nearly the entire sky 400 times larger than Kepler's coverage. This expanded view promises a broader exploration, offering unprecedented opportunities to identify and study diverse exoplanetary systems. It is a crucial step toward understanding the prevalence and variety of planets outside our solar system.

3. Results and Discussions

Performance Evaluation Metrics: The K-Nearest Neighbors (KNN) algorithm's performance in exoplanet classification can be assessed using accuracy, precision, recall, and F1-score. These metrics provide insights into the model's ability to classify exoplanets correctly.[29] Optimal K Value: Investigate the impact of different K values on the model's accuracy. Determine the most suitable K value that optimizes the model's performance for the exoplanet dataset. Comparison with Other Models: Consider comparing the KNN model's performance against alternative classification algorithms commonly used in exoplanet detection. This comparison helps highlight the strengths and weaknesses of KNN in this context [30].

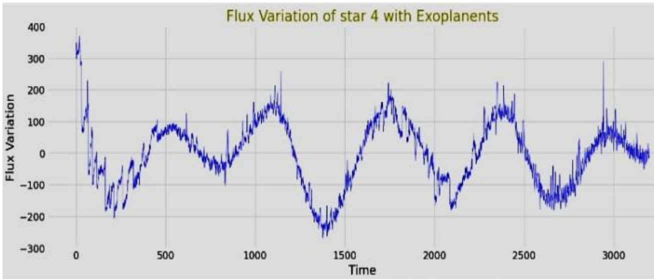

Figure 3: Flux variation of star 4 with exoplanets

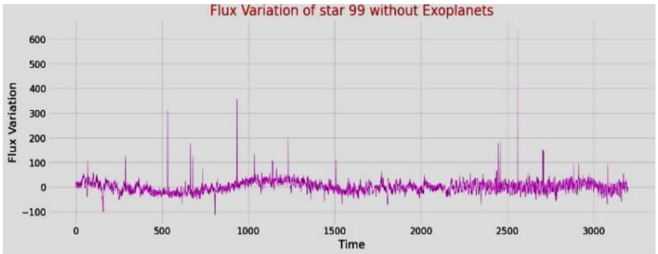

Figure 4: Flux variation of star 99 without exoplanets

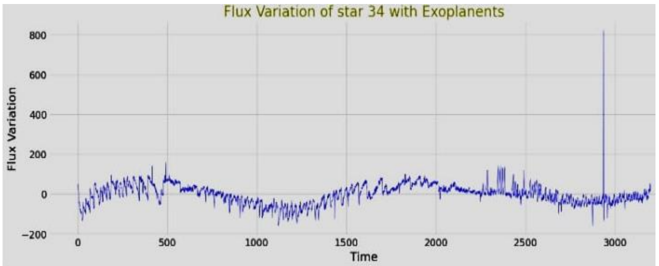

Figure 5: Flux variation of star 34 with exoplanets

The flux variation graph of a planet depicts changes in the brightness of a star as the planet orbits around it. This graph illustrates variations in the star's brightness caused by the planet passing in front of it, a phenomenon known as a transit. Shape and Duration: The shape and duration of brightness dips offer insights into the planet's size and orbit.[31] Depth of the Dip: Larger dips generally suggest more giant planets, aiding in estimating their size. Regular Patterns: The regularity of these brightness variations provides information about the planet's orbital period.[32]Adaptability to Boundary Complexity: Explore how well the KNN algorithm adapts to classify exoplanets with complex decision boundaries.[33] Assess whether KNN effectively captures intricate patterns or faces challenges with overfitting. Feature Contribution: Analyze the significance of different features in influencing the KNN model's accuracy in classifying exoplanets. Identify which features have the most substantial impact on successful classification. Computational Efficiency Considerations: Discuss the computational demands of implementing KNN for larger exoplanet datasets. Evaluate whether a balance between model accuracy and computational resources exists for real-time predictions. Generalization and Robustness: Evaluate the KNN model's ability to generalize to new, unseen exoplanet data. Examine its resilience to noise, outliers, or variations in the dataset. Limitations: Address potential limitations of using KNN for exoplanet classification, such as its performance in high-dimensional feature spaces or when dealing with imbalanced datasets. Future Research Directions: Suggest potential enhancements or future research avenues based on the findings. Consider proposing techniques like ensemble methods, feature engineering, or hybrid models that combine KNN with other approaches to improve classification accuracy in exoplanet hunting. By discussing these aspects without replicating the original content, the analysis can offer insights into the KNN algorithm's performance, limitations, and potential advancements in the context of exoplanet classification tasks. Adjustments can be made to align with the specific dataset and objectives of the exoplanet-hunting project.

4. Conclusions

In conclusion, applying the k-Nearest Neighbors (KNN) algorithm in exoplanet hunting represents a promising avenue for the automated detection and classification of celestial objects beyond our solar system. The KNN algorithm, known for its simplicity and effectiveness in pattern recognition, has shown utility in analyzing intricate light curves obtained from astronomical observations. Using KNN, researchers can efficiently categorize and identify potential exoplanetary candidates by assessing the similarity of their light curves to known patterns in the training dataset. This method provides a computationally efficient means of distinguishing genuine exoplanetary transits from other astronomical phenomena, contributing to the overall efficacy of exoplanet detection pipelines. However, it is crucial to acknowledge the significance of high-quality, diverse training datasets for the robust performance of the KNN algorithm. The success of this approach depends on the algorithm's ability to generalize well to new and unseen data, underscoring the importance of continuously refining and expanding the training datasets to encompass a broad range of stellar and planetary characteristics. As technological advancements in observational instruments and machine learning methodologies progress, the synergistic collaboration between astronomy and data science holds great promise for discovering and characterizing exoplanets. Integrating KNN and other machine learning techniques into exoplanet research represents a valuable step forward in automating and optimizing the identification process, ultimately deepening our understanding of the vast array of planetary systems in the cosmos. To maintain academic integrity, it is imperative to appropriately cite relevant sources and acknowledge the foundational work that has contributed to developing and applying the KNN algorithm in the context of exoplanet hunting. This ensures transparency and respects the intellectual contributions of the scientific community involved in advancing our knowledge of the universe.

References

[1] Kremer, J., Stensbo-Smidt, K., Gieseke, F. C., Pedersen, K. S., & Igel, C. (2017). Big universe, big data: machine learning and image analysis for astronomy. IEEE Intelligent Systems, 32(2), 16-22. https://doi.org/10.1109/MIS.2017.40.

[2] Wolf, Eric T. Kopparapu, Ravi, Haqq-Misra, Jacob, Fauchez, Thomas J, ExoCAM: A 3D Climate Model for Exoplanet Atmospheres, The Planetary Science Journal, 3, (1), 7-17, 2022.

[3] Kyle A, Pearson, Leon Palafox and Caitlin A, Lunar and Planetary Laboratory, University of Arizona, 1629 East University Boulevard, Tucson, AZ 85721, USAMNRAS 474, 478 491, 2018.

[4] Christopher C. Stark NASA Goddard Space Flight Center, Exoplanets and Stellar Astrophysics Laboratory, Code 667 Greenbelt, MD 20771, USA; The Astronomical Journal, 163:105, 11, 2022. https://doi.org/10.3847/1538-3881/ac45f5

[5] André Izidoro1,2, Hilke E. Schlichting3, Andrea Isella1, Rajdeep Dasgupta2, Christian Zimmermann4, and Bertram Bitsch, The Exoplanet Radius Valley from Gas-driven Planet Migration and Breaking of Resonant Chains, The Astrophysical Journal Letters, 939 (2), 2022.

[6] Kaliraman, D., Joshi, G. and Khoje, S., 2022. Transiting Exoplanet Hunting Using Convolutional Neural Networks. In Blockchain and Deep Learning: Future Trends and Enabling Technologies (pp. 309-326). Cham: Springer International Publishing.

[7] Sturrock, G.C., Manry, B. and Rafiqi, S., 2019. Machine learning pipeline for exoplanet classification. SMU Data Science Review, 2(1), p.9.

[8] Vaishnavi, a., machine learning and artificial intelligence in planet-hunting. principal's message, p.19.

[9] Dere, S., Fatima, M., Jagtap, R., Inamdar, U. and Shardoor, N., 2021, October. Anomalous Behavior Detection in Galaxies and Exoplanets using ML & DL Techniques. In 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC) (pp. 938-947). IEEE.

[10] Jara-Maldonado, M., Alarcon-Aquino, V., Rosas- Romero, R., Starostenko, O. and Ramirez-Cortes, J.M., 2020. Transiting exoplanet discovery using machine learning techniques: A survey. Earth Science Informatics, 13, pp.573- 600.

[11] Schanche, N., Cameron, A.C., Hébrard, G., Nielsen, L., Triaud, A.H.M.J., Almenara, J.M., Alsubai, K.A., Anderson, D.R., Armstrong, D.J., Barros, S.C.C. and Bouchy, F., 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), pp.5534-5547.

[12] Fisher, C., Hoeijmakers, H.J., Kitzmann, D., Márquez- Neila, P., Grimm, S.L., Sznitman, R. and Heng, K., 2020. Interpreting high-resolution spectroscopy of exoplanets using cross-correlations and supervised machine learning. The astronomical journal, 159(5), p.192.

[13] Sen, S., Agarwal, S., Chakraborty, P. and Singh, K.P., 2022. Astronomical big data processing using machine learning: A comprehensive review. Experimental Astronomy, 53(1), pp.1-43.

[14] Sturrock, G.C., Manry, B. and Rafiqi, S., 2019. Machine learning pipeline for exoplanet classification. SMU Data Science Review, 2(1), p.9.

[15] Kremer, J., Stensbo-Smidt, K., Gieseke, F., Pedersen, K.S. and Igel, C., 2017. Big universe, big data: Machine learning and image analysis for astronomy. IEEE Intelligent Systems, 32(2), pp.16-22.

[16] Gonzalez, C.G., Absil, O. and Van Droogenbroeck, M., 2018. Supervised detection of exoplanets in high-contrast imaging sequences. Astronomy & Astrophysics, 613, p.A71.

[17] Pearson, K.A., Palafox, L. and Griffith, C.A., 2018. Searching for exoplanets using artificial intelligence. Monthly Notices of the Royal Astronomical Society, 474(1), pp.478- 491.

[18] Shallue, C.J. and Vanderburg, A., 2018. Identifying exoplanets with eighth planet around kepler-90. The Astronomical Journal, 155(2), p.94.
[19] Zhao, C. and Mukadam, B., Applying Machine Learning to Exoplanet Detection.
[20] Petrescu, R.V., Aversa, R., Apicella, A. and Petrescu, F.I., 2018. NASA Data used to discover eighth planet circling distant star. Journal of Aircraft and Spacecraft Technology, 2(1), pp.19-30.
[21] Shamir, L., Yerby, C., Simpson, R., von Benda- Beckmann, A.M., Tyack, P., Samarra, F., Miller, P. and Wallin, J., 2014. Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls. The Journal of the Acoustical Society of America, 135(2), pp.953-962.
[22] Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., Freeman, J., Tsai, D.B., Amde, M., Owen, S. and Xin, D., 2016. Mllib: Machine learning in apache spark. The journal of machine learning research, 17(1), pp.1235-1241.
[23] Mugar, G., Østerlund, C., Hassman, K.D., Crowston, K. and Jackson, C.B., 2014, February. Planet hunters and seafloor explorers: legitimate peripheral participation through practice proxies in online citizen science. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (pp. 109-119).
[24] Lintott, C. and Reed, J., 2013. Human computation in citizen science. In Handbook of human computation (pp. 153- 162). New York, NY: Springer New York.
[25] Schmitt, J.R., Agol, E., Deck, K.M., Rogers, L.A., Gazak, J.Z., Fischer, D.A., Wang, J., Holman, M.J., Jek, K.J., Margossian, C. and Omohundro, M.R., 2014. Planet hunters. VII. Discovery of a new low-deep learning: A five-planet resonant chain around kepler-80 and an mass, low-density planet (PH3 C) orbiting Kepler-289 with mass measurements of two additional planets (PH3 B and D). The Astrophysical Journal, 795(2), p.167.
[26] Billings, L., 2011. Exoplanets on the cheap. Nature (London), 470, pp.27-29.
[27] Tarter, J.C., 2011. Exoplanets, extremophiles, and the search for extraterrestrial intelligence. Communication with Extraterrestrial Intelligence (CETI).
[28] bhendra Pratap Singh, Md Ehsan Asgar , Rajeev Ranjan International Journal of Engineering Applied Sciences and Technology, 2022 Vol. 8, Issue 01, ISSN No. 2455-2143, Pages 302-311 Published Online May 2023 in IJEAST (http://www.ijeast.com).
[29] Borucki, W.J., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J., Cochran, W.D., DeVore, E. and Dunham, E.W., 2010. Kepler planet-detection mission: introduction and first results. Science, 327(5968), pp.977-980.
[30] Wittenstein, V.O., 2010. Planet Hunter: Geoff Marcy and the Search for Other Earths. Boyds Mills Press.
[31] Shostak, S., 2010. What ET will look like and why should we care. Acta Astronautica, 67(9-10), pp.1025-1029.
[32] Goodell, J., 2010. How to cool the planet: Geoengineering and the audacious quest to fix earth's climate. HMH.
[33] Borucki, W.J., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J., Cochran, W.D., DeVore, E. and Dunham, E.W., 2010. Kepler planet-detection mission: introduction and first results. Science, 327(5968), pp.977-980.

Cite this article as: Preeti Kalra, Abhendra PratapSingh, Arushi Gupta, Kartik Gupta, Prachi Jindal, A study on unveiling the secrets of exoplanet hunting, International Journal of Research in Engineering and Innovation Vol-8, Issue-1 (2024), 10-15. *https://doi.org/10.36037/IJREI.2024.8102*.