galaxies

MDPI

Article

Analysis of Habitability and Stellar Habitable Zones from Observed Exoplanets

Jonathan H. Jiang [1,*], Philip E. Rosen [2], Christina X. Liu [3], Qianzhuang Wen [4] and Yanbei Chen [5]

1 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
2 Independent Researcher, Vancouver, WA 98662, USA; philip.e.rosen@gmail.com
3 Lakeside School, Seattle, WA 98125, USA; christinaliu2026@gmail.com
4 Pacific Academy, Irvine, CA 92604, USA; qianzhuang.wen@gmail.com
5 Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA 91109, USA; yanbei@tapir.caltech.edu
* Correspondence: jonathan.h.jiang@jpl.nasa.gov

Abstract: The investigation of exoplanetary habitability is integral to advancing our knowledge of extraterrestrial life potential and detailing the environmental conditions of distant worlds. In this analysis, we explore the properties of exoplanets situated with respect to circumstellar habitable zones by implementing a sophisticated filtering methodology on data from the NASA Exoplanet Archive. This research encompasses a thorough examination of 5595 confirmed exoplanets listed in the Archive as of 10 March 2024, systematically evaluated according to their calculated average surface temperatures and stellar classifications of their host stars, taking into account the biases implicit in the methodologies used for their discovery. Machine learning, in the form of a Random Forest classifier and an XGBoost classifier, is applied in the classification with high accuracies. The feature importance analysis indicates that our approach captures the most important parameters for habitability classification. Our findings elucidate distinctive patterns in exoplanetary attributes, which are significantly shaped by the spectral classifications and mass of the host stars. The insights garnered from our study both inform refinement of existing models for managing burgeoning exoplanetary datasets, and lay foundational groundwork for more in-depth explorations of the dynamic relationships between exoplanets and their stellar environments.

Keywords: exoplanet; habitable zone

check for updates

Citation: Jiang, J.H.; Rosen, P.E.; Liu, C.X.; Wen, Q.; Chen, Y. Analysis of Habitability and Stellar Habitable Zones from Observed Exoplanets. *Galaxies* **2024**, *12*, 86. https://doi.org/10.3390/galaxies12060086

Academic Editor: Alberto C. Sadun

Received: 19 August 2024
Revised: 23 October 2024
Accepted: 22 November 2024
Published: 3 December 2024

1. Introduction

The search for habitable worlds beyond our Solar System is one of the most profound and rapidly advancing areas of astrophysics and astrobiology. Since the first confirmed exoplanet discovery in 1992 [1], advancements in observational technology—such as the Kepler Space Telescope, Transiting Exoplanet Survey Satellite (TESS), and other ground- and space-based missions—have led to the identification of nearly 5800 exoplanets by 4Q2024 [2]. Among the most compelling aspects of exoplanet research is the quest for habitable worlds, specifically those located within the circumstellar habitable zone (HZ) of their host stars. The HZ is defined as the region around a star where a planet could potentially maintain liquid water on its surface—a key criterion for habitability as we understand it [3,4].

Early definitions of the HZ were relatively simplistic, based primarily on the amount of stellar radiation a planet receives [5,6]. Over time, these models evolved to incorporate more complex parameters, such as the impact of planetary albedo, greenhouse gas effects, stellar luminosity variations, and orbital dynamics [3,7]. Kopparapu et al. (2013) [4] further refined the classical HZ definition by updating the boundaries for different stellar types, providing a more accurate framework for assessing exoplanet habitability across a broader range of stellar environments. In addition to planetary distance from the host star, factors such as atmospheric composition, magnetic field strength, and axial tilt play critical roles in determining a planet's potential to support life [8,9]. While these advancements have significantly improved the theoretical understanding of planetary habitability, observational challenges persist.

One major obstacle in exoplanetary habitability studies is the inherent bias in detection methods, which skews the distribution of discovered planets. The Transit method, responsible for the majority of exoplanet discoveries, is particularly biased toward detecting planets with short orbital periods, leading to an overrepresentation of "Too Hot" planets located close to their host stars [10,11]. Similarly, the Radial Velocity (RV) method preferentially detects massive planets with strong gravitational influences on their stars, resulting in a dataset that is biased toward large, Jupiter-like planets [12]. These biases complicate efforts to accurately assess the true distribution of habitable planets and necessitate careful consideration when interpreting observational data.

To address these challenges, recent studies have begun employing more sophisticated statistical techniques, including machine learning (ML), to improve habitability predictions and mitigate detection biases [13,14]. ML algorithms can analyze large datasets and identify patterns that may be difficult to detect using traditional methods, thereby enhancing the ability to classify planets based on their potential for habitability. For instance, unsupervised learning techniques have been used to cluster exoplanets by stellar and planetary characteristics, providing a more nuanced understanding of which types of planets are more likely to be habitable [15]. Such approaches complement existing models by incorporating a wider range of planetary properties, such as atmospheric conditions and stellar activity, into habitability assessments [8,9].

Building on these foundational studies, our research presents a large-scale statistical analysis of 5595 confirmed exoplanets listed in the NASA Exoplanet Archive as of 10 March 2024, with a specific focus on their potential habitability. Unlike many previous studies that center on exoplanets detected by a single mission, such as Kepler or TESS, our analysis incorporates data from multiple sources, offering a broader and more diverse dataset. This more comprehensive approach allows for a richer analysis of planetary habitability and helps mitigate the biases introduced by relying on a single detection method. By applying a detailed filtering methodology, we categorize exoplanets based on their estimated average surface temperatures, host star characteristics, and proximity to the habitable zone. Our study also confronts the biases introduced by detection methods head-on, critically analyzing how these biases influence the observed distribution of potentially habitable planets.

This work distinguishes itself from prior research in several key ways. First, while we acknowledge the importance of sophisticated habitability models—such as those proposed by Kopparapu et al. (2013) [4] and Kasting et al. (1993) [3]—we adopt a more empirical approach that accounts for conventional habitability metrics (e.g., surface temperature ranges) while addressing the observational limitations of current exoplanet detection techniques. Our approach involves an equation to calculate exoplanet average surface temperature based on three features: the host's effective surface temperature, the host's radius, and the exoplanet's distance from its host star. Then, the exoplanet's habitability is determined based on the calculated average surface temperature. A Random Forest [16] and an XGBoost [17] classifier are trained based on the combined data from the NASA Exoplanet Archive and the Habitable Worlds Catalog (HWC), Planetary Habitability Laboratory (PHL) @ UPR Arecibo [18]. The feature importance analysis on the Random Forest classifier shows that our approach indeed captures the most importance features for habitability classification. Second, we emphasize the role of host star mass and spectral type in shaping the properties of exoplanets within the HZ. Much exoplanet research has concentrated on G-type stars similar to the Sun, yet M-type and K-type stars are more abundant in the galaxy and often less explored in habitability studies due to their generally smaller HZs and unique stellar activity patterns [8,9,19]. By broadening our scope to include these often-overlooked stellar types, we aim to offer a more holistic understanding of planetary habitability.

Our analysis addresses several important questions: How does the distribution of habitable zone planets vary across different stellar classes? What biases are introduced by the Transit and Radial Velocity methods, and how do these biases affect our understanding of planetary habitability? By answering these questions, we aim to refine current models of habitability and provide novel insights into the dynamic relationships between exoplanets and their stellar environments. Moreover, by critically assessing the limitations of the data and methods used, our study highlights the need for future observational missions with improved sensitivity and coverage to better assess the true prevalence of habitable worlds. In particular, missions with advanced imaging and spectroscopic capabilities could help overcome the detection biases we face today, offering more accurate insights into exoplanetary atmospheres and surface conditions.

The remainder of this paper is structured as follows: In Section 2, we provide a detailed description of our methodology, including the criteria for selecting exoplanets, the calculation of average surface temperatures, and our approach to addressing observational biases, followed by result summary of the Random Forest and XGBoost classifiers and the feature important analysis. Section 3 presents our findings, focusing on the distribution of exoplanets in relation to their host stars' habitable zones, with an emphasis on differences across stellar types. Section 4 discusses the broader implications of our results for exoplanetary science, particularly in the context of future observational missions. Finally, we offer conclusions and propose avenues for further research that can expand our understanding of planetary habitability beyond current limitations.

2. Methodology

The primary dataset was sourced from the NASA Exoplanet Archive, a comprehensive repository of exoplanet data. The focus was on the 5595 confirmed exoplanets listed in the Archive's *Planetary Systems Composite Data* section as of 10 March 2024. This dataset was organized to facilitate the efficient application of computational analyses and the interpretation of complex patterns. Observational specifics, such as exoplanet orbital parameters and host star radiative characteristics, were systematically accounted for. The approach leverages detailed descriptions within the dataset, enabling the extraction of statistically significant relationships based chiefly on primary observables such as host star mass, type, and exoplanet semi-major axis.

2.1. Addressing Observational Biases

Acknowledging and accounting for the biases inherent in common exoplanet detection methods is crucial to enabling the extraction of meaningful results and conclusions. We critically analyzed how the Transit method, predominantly detecting short-period exoplanets, and the Radial Velocity method, favoring massive exoplanets, might skew our dataset. To counter these biases and refine our dataset's reliability, we considered using exclusively Kepler mission data, known for its comprehensive and diverse exoplanet discoveries. While this strategic selection aimed to provide a more balanced view of the exoplanet population, excluding non-Kepler derived data would introduce its own set of biases simply through the absence of valuable confirmed exoplanetary data from other sources. Alternatively, quantitative contrasts were drawn between the population distribution of stellar classes in general throughout the Milky Way galaxy and those comprising host stars contained in the Archive. Further, our analysis focused on single (known) host exoplanetary systems in order to provide consistency in considering the key habitability factors of radiative environments and orbital mechanics.

As every observational technique imparts its own intrinsic biases on the data collected and the number of methodologies for detecting exoplanets are limited (e.g., the Transit and Radial Velocity methods account for over 90% of the Archive's confirmed exoplanets), avoiding all bias is not possible with present day technology. Attempts to apply correction factors would introduce new sources of bias, along with the layering in of additional uncertainty. Accordingly, an internalized approach to countering bias, in as much as can be done, is called for. This is substantially achieved by selection of the Archive's composite database for statistical analysis, which itself is a quantitative blending of multiple observations that have been made for many of the 5595 exoplanets individually listed. While redundancy in reporting, in aggregate averaging approximately six per confirmed exoplanet, is generally derived from the same observational technique (e.g., Transiting, Radial Velocity, etc.), the benefits from applying differing methodologies for analysis and interpretation are directionally corrective. Further, by examining relationships between a small set of key parameters and from a simplified perspective (e.g., host stellar class, first principles-based habitable zone calculation), variances were deemphasized which enabled useful conclusions to be extracted.

2.2. Habitability Zone Determination

The habitable zone of a given exoplanet was determined based on its calculated average surface temperature, denoted as $T_{surf,\ ave}$, a critical indicator of potential liquid water presence. We categorized exoplanets as "Too Hot" ($T_{surf,\ ave}$ > 100 °C), "Too Cold" ($T_{surf,ave}$ < 0 °C) or within the habitable zone ("In HZ") for $T_{surf,\ ave}$ between the benchmark temperature range of 0 and 100 °C. This classification was vital to identifying exoplanets that could potentially support life under the assumed necessary precursor of surface-accessible liquid phase H_2O.

Utilizing the basic Radiative Equilibrium equation as derived from first principles of radiative heat transfer, the exoplanet average surface temperature calculation considered several characterizing factors. These include the observed exoplanet's distance from its host star (d), the host's effective surface temperature (T) and radius (R), along with an assumed exoplanet albedo (A) and additional scaler to account for bulk atmospheric greenhouse gas effect (k):

$$T_{surf,ave} = kT(1-A)^{0.25}(R/(2d))^{0.5} \tag{1}$$

Applying Equation (1) to each listing in the NASA Exoplanet Archive enabled selective sifting of the database to produce quantified results based on the aforementioned exoplanet HZ status categories. Additionally, where the Archive had no entry for the observable parameters T and/or R and/or d, a designation of "N/A" was made for the associated exoplanet to denote insufficient information for determining HZ status in those cases. In determining exoplanet HZ status, by intent a wide net was cast, as observable parameters for distinguishing gas giants from rocky worlds, most specifically planetary mass and radius, did not enter into the above calculation. The basis for taking this broad approach was four-fold:

1. Exoplanet mass and radius (and thus density) are bulk descriptors, hence limited in detailing the particulars of composition and topography of a given world.
2. Quantitative categorization between larger rocky planets (e.g., "super-Earths") and relatively small gas planets (e.g., "mini-Neptunes") remains elusive—i.e., there is no precise cut-off point from which to delineate one from the other.
3. Habitability, aside from the singular requirement of dependence on liquid phase H_2O, was interpreted in a comprehensive sense. Driven by logical necessity to include the environments which play host to the wide variety of known biological life (let alone any speculations on extraterrestrial native life), this definitional elasticity directly lays open for occupation the atmospheric, surface and/or sub-surface realms.
4. As a high-level (i.e., first pass) statistical treatment of habitability for confirmed exoplanets, it is preferrable to avoid any borderline case exclusions. More granular analysis can be readily applied to these findings, further refining the definition of habitability itself and the ramifications for exoplanets subsequentially described as such.

2.3. Stellar Classification and Exoplanet Distribution

We examined how exoplanetary formation correlates with various stellar classes. This analysis involved sorting habitable zone exoplanets by their host stars' spectral types and investigating the variations in HZ boundary distances in relation to host star mass. This part of our methodology was designed to uncover trends in exoplanet distribution across different types of stars, providing insights into the likelihood of habitable planets around various stellar classes. Further elucidation of aforementioned observational bias was sought by challenging conventional assumptions regarding the propensity for hosting planetary systems based on stellar type.

2.4. Non-Applicable and Additional Factors

The study acknowledged, but did not focus on, several complexing factors which can influence habitability, such as orbital eccentricity, axial obliquity, atmospheric composition, and the presence (or lack thereof) of exoplanetary magnetic fields and/or exomoons, due to data limitations. Some insights into the temporal variability of habitable zones—i.e., the impact of stellar evolution on the habitability of orbiting planets over time—can be deduced qualitatively from statistical treatment of the data.

2.5. Computational Approach

The *Planetary Systems Composite Data* from the NASA Exoplanet Archive as of 10 March 2024 was the primary dataset, chosen for its consolidation of multiple observations for a given exoplanet into single line data entries. Custom calculations for thermal equilibrium and habitable zone estimations were added directly into the resultant spreadsheet. This enhanced the functionality of the spreadsheet by folding-in analytical tool capability, enabling extraction of meaningful patterns and relationships from the nearly 5600 confirmed exoplanet entries. Additional relevant data, such as host star age, was also incorporated to further extend the analysis.

2.6. Assumptions and Limitations

Our analysis was underpinned by several assumptions, notably adopting Earth's albedo (A = 0.306) as a baseline for exoplanets, as well as accounting for the atmospheric greenhouse gas effect through the bulk temperature factor (k = 1.13), again using Earth as the standard. Recognizing that our empirical relationships, based on a large but nonetheless limited dataset, might introduce certain biases, we were careful to frame our findings within these constraints. Where assumptions were necessary to complete calculations, rational bracketing conditions were applied accordingly.

2.7. Habitability Classification via Random Forest and XGBoost

Machine learning has been increasingly adopted in exoplanet habitability study to identify the relationships between stellar and planetary parameters that influence habitability and to predict exoplanet habitable status. Saha et al. (2018) [20] trained an XGBoost model based on a dataset from PHL @ UPR Arecibo [18] with high accuracy. The model was validated through testing against Proxima b and the TRAPPIST-1 system. Basak et al. (2021) [21] constructed a novel neural network architecture for the exoplanet habitability classification task. Ghadekar et al. (2024) [22] applied causal learning to find the most influential factors in the data and then trained habitability classification models with Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM).

As part of our study, we trained a Random Forest classifier and an XGBoost classifier to predict exoplanetary habitability. We intentionally chose these decision tree-based models as they provide insights into the importance of each feature's contribution to the prediction, so we can use this information to cross check our empirical approach mentioned in Section 2.2. With regards to the results, both models achieved high accuracies, with the Random Forest at 0.95 and the XGBoost at 0.97. Appendix A provides details about the training process of these two classifiers (e.g., training data processing, hyperparameter tuning) and further breakdown of the models' performance (precision, recall, and F1 score).

To understand which features influence exoplanet habitability the most, a feature importance analysis was conducted on the Random Forest classifier. Figure 1 shows the feature importance for each feature, sorted in descending order. The host star's radius, host star's effective temperature, host star's surface gravity, and planet's orbit semi-major axis (i.e., distance from its host star) are the top four features in terms of feature importance for the habitability predication. This analysis validates our empirical approach described in Section 2.2 as three out of the top four features are captured for habitability determination. This result also calls out the importance of the host star's surface gravity as a feature for habitability prediction. Future iterations of our empirical approach could potentially take the host star's surface gravity into consideration.

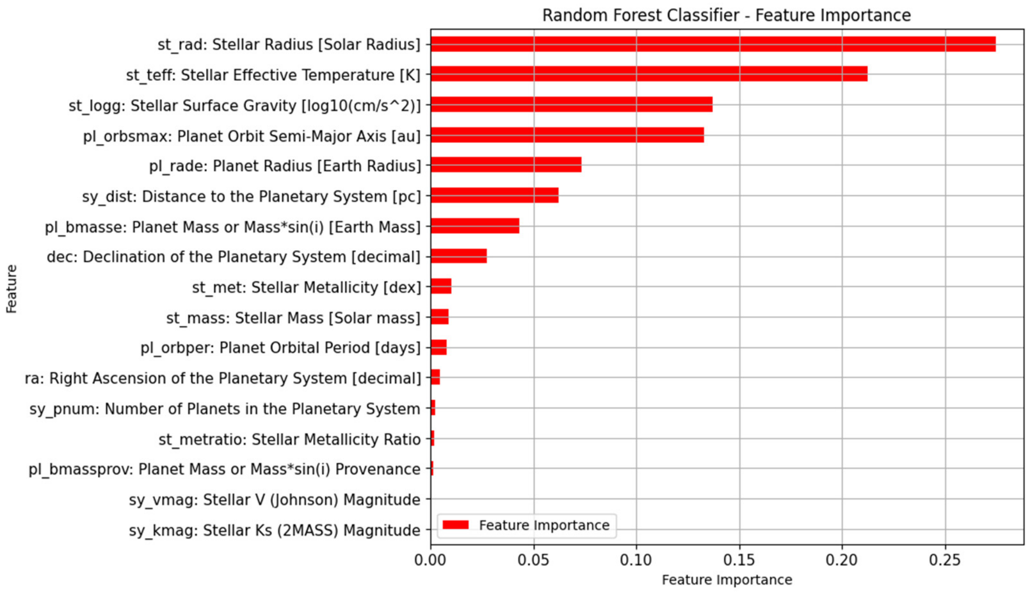


Figure 1. Feature importance analysis results of the Random Forest classifier. Features are sorted in descending order in terms of importance of contribution to the habitability prediction. Asterisk (*) notates multiplication, and caret (ˆ) notates power. sin(*i*) denotes the sine of the planet orbital inclination.

3. Results

Analysis of exoplanet habitability within circumstellar habitable zones reveals several critical insights, visualized through a series of figures that illustrate various aspects of the data. By linking these figures together, a narrative is constructed that enhances the understanding of exoplanetary habitability and aligns it with existing literature. Systematic examination of 5595 confirmed exoplanets from the NASA Exoplanet Archive, applying Equation (1) to calculate the average surface temperatures ($T_{surf,\ ave}$) of each and categorize their habitable zone status, was performed.

Figure 2 depicts the cumulative count history of all confirmed exoplanet discoveries, highlighting their habitable zone status. The exponential growth in exoplanet discoveries, particularly since the launch of the Kepler Space Telescope, underscores significant advancements in detection technologies and methodologies [23]. This trend illustrates our increasing ability to identify potentially habitable exoplanets, reflecting ongoing refinements in search strategies and expanding observational capabilities. Data for 2024, though incomplete, is indicative of ongoing discoveries, suggesting continued growth in the counts of habitable and non-habitable zone exoplanets, driven by new technologies enabling more advanced data analysis techniques and observation missions.

To understand the spatial distribution of exoplanet discoveries as it relates to habitable zone status, Figure 3 illustrates the distance-wise distribution of all confirmed exoplanets from the Solar System, categorized also by their HZ status. The majority of exoplanet detections remain concentrated within 1000 light-years while HZ exoplanets are all but completely contained within 10,000 light-years, highlighting observational biases where nearer stars are more frequently surveyed due to limitations in current detection technologies. This bias underscores the challenge of detecting more distant worlds and emphasizes the need for next-generation telescopes capable of probing deeper into the galaxy to identify more distant, potentially habitable exoplanets [24]. The clustering of discoveries within 1000 light-years also reflects the limitations in the sensitivity and resolution of current instruments, as well as the prioritization of closer stars for detailed observation [11].

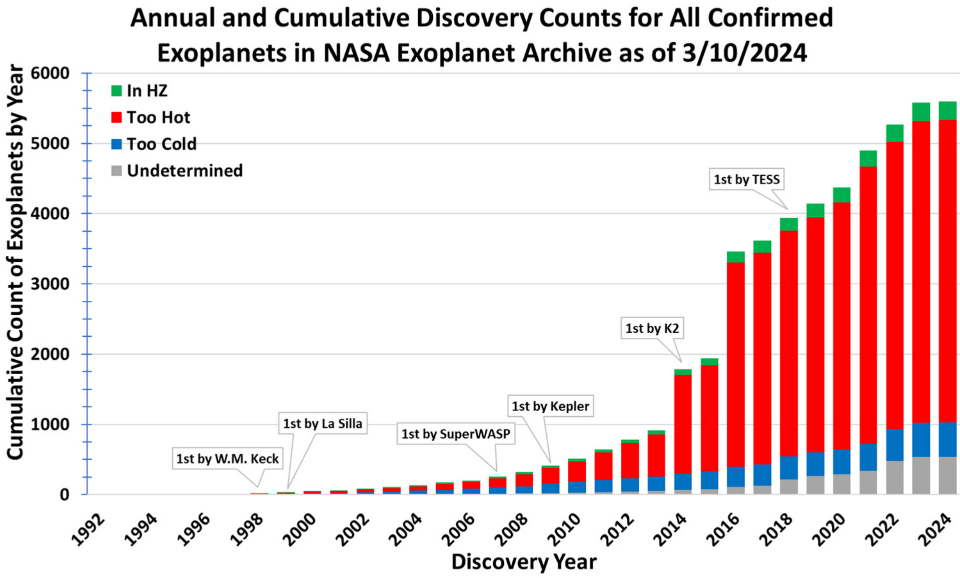


Figure 2. Cumulative count history of all confirmed exoplanet discoveries showing habitable zone status per the chosen parameters of this analysis. Note: Data for 2024 is incomplete.

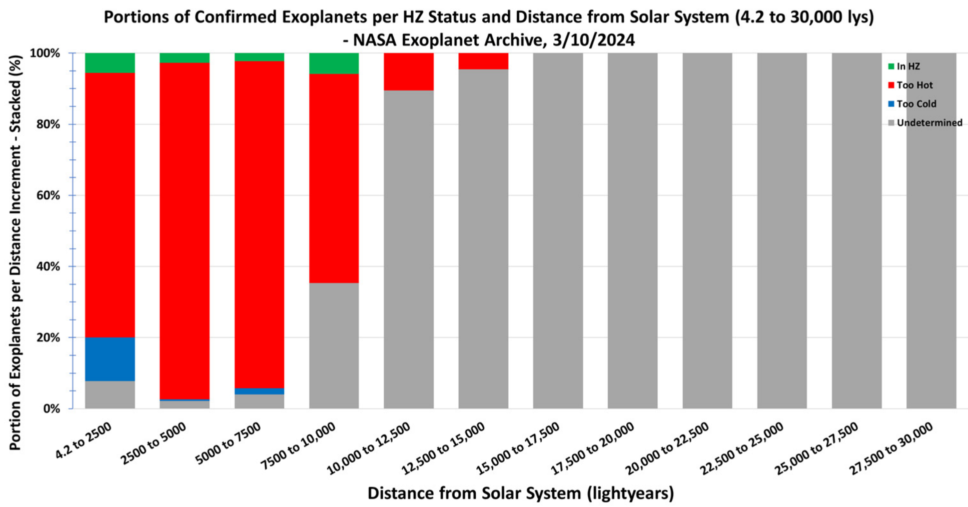


Figure 3. Distance-wise distribution of all confirmed exoplanets from Solar System and their habitable zone status per the chosen parameters of this analysis.

Figure 4 presents the distribution of confirmed single-hosted exoplanets based on their habitable zone status. Among these exoplanets, 77.75% are categorized as "Too Hot," 8.04% as "Too Cold," 4.48% are within the HZ, and 9.73% have indeterminate status (N/A). The significant proportion of "Too Hot" exoplanets suggests an observational bias, as closer-in planets with shorter orbital periods are easier to detect using methods such as the Transit method. This finding aligns with previous studies that highlight the detection bias towards short-period exoplanets, often leading to an underrepresentation of planets orbiting within more distant habitable zones [25]. The skew towards "Too Hot" exoplanets also reflects the challenges in detecting cooler, potentially habitable planets that lie farther from their host stars, where longer orbital periods and lower transit probabilities complicate their detection [10].

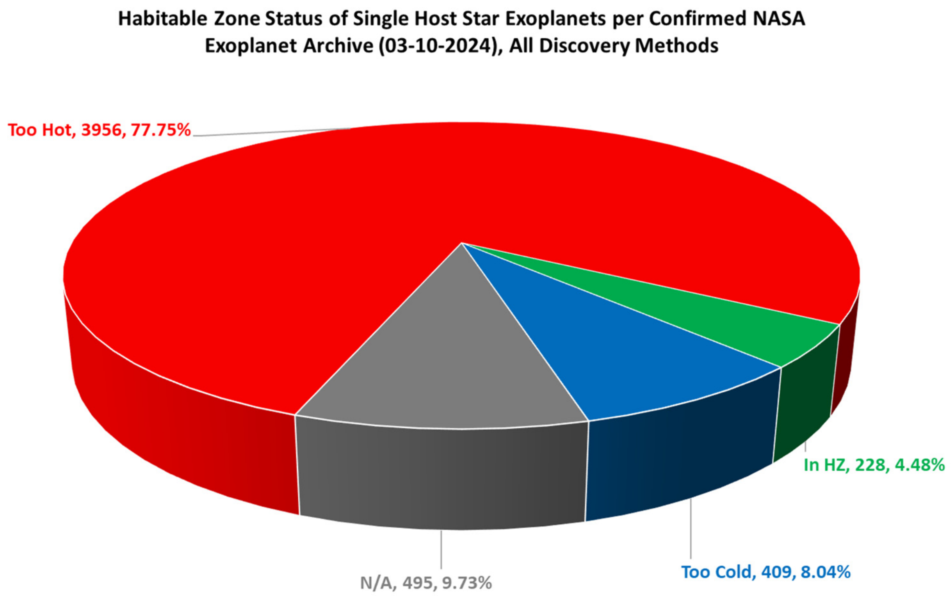
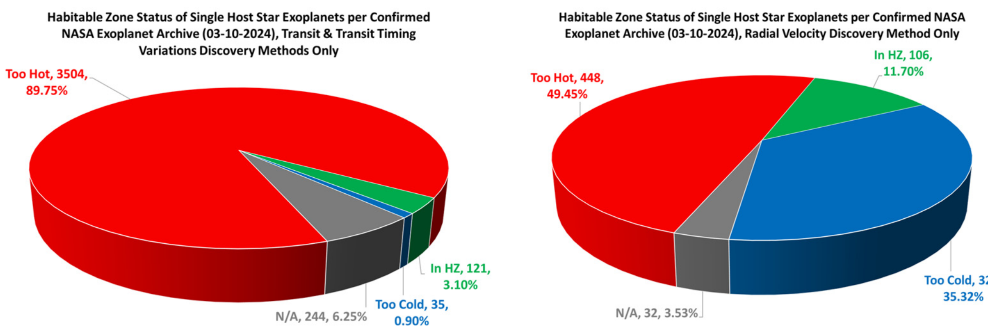


Figure 4. Count-up and portion of confirmed single-hosted exoplanets' habitable zone status per the chosen parameters of this analysis.

Delving deeper into the detection methods, Figure 5a shows the habitable zone status of single-hosted exoplanets discovered via the Transit and Transit Timing Variations methods. Specifically, 89.75% are "Too Hot," 0.90% are "Too Cold," 3.10% are within the HZ, and 6.25% are N/A. The overwhelming majority of "Too Hot" exoplanets discovered through these methods underscores the inherent observational bias towards detecting planets with shorter orbital periods, which are more likely to transit their host star frequently. This bias reflects the limitations of the Transit method in identifying habitable zone exoplanets, as it is more sensitive to planets that orbit closer to their stars [14]. The underrepresentation of "Too Cold" and HZ planets emphasizes the need for more sensitive instruments and extended observation periods to detect planets in wider orbits.



(**a**) (**b**)

Figure 5. (**a**) Left: Habitable zone status of single hosted exoplanets discovered by way of the Transit and Transit Timing Variations methods; (**b**) Right: Habitable zone status of single hosted exoplanets discovered by way of the Radial Velocity method.

Figure 5b presents the habitable zone status of single-hosted exoplanets discovered using the Radial Velocity method. The distribution is more balanced compared to the Transit method, with 49.45% "Too Hot", 35.32% "Too Cold", 11.70% within the HZ, and 3.53% N/A. The Radial Velocity method's sensitivity to planets at various distances from their host stars provides a broader view of exoplanetary systems, although it still shows a detection bias towards larger planets. This method's ability to detect planets in a wider range of orbits highlights its complementary role in identifying habitable zone exoplanets, addressing some of the limitations inherent in the Transit method [12]. The broader distribution of HZ exoplanets detected by Radial Velocity indicates its potential in revealing more distant, possibly habitable planets that are missed by transit surveys. Note: Appendix B, Figure A1 extends the analysis presented in Figure 5a,b to the Imaging method, revealing this technique's detection capability for HZ exoplanets still to be quite limited in comparison.

Moving to the stellar classifications of host stars, Figure 6 shows the proportions of host stars across all single-hosted exo-systems containing at least one HZ exoplanet. The breakdown is as follows: 26.32% are M-type stars, 29.82% are K-type stars, 35.53% are G-type stars, and 7.89% are F-type stars. This distribution indicates a higher prevalence of HZ exoplanets around G-type and K-type stars, aligning with the fact that these stars are prime targets for habitability studies due to their stable lifetimes and favorable conditions for liquid water [3]. The relatively lower proportion of HZ exoplanets around M-type stars, despite their abundance in the galaxy, reflects the challenges in detecting potentially habitable planets around these dimmer, cooler stars [8,9]. M-type stars, while plentiful, have smaller HZs closer to the star, making planets within these zones more susceptible to stellar activity and tidal locking, which could hinder habitability [19].

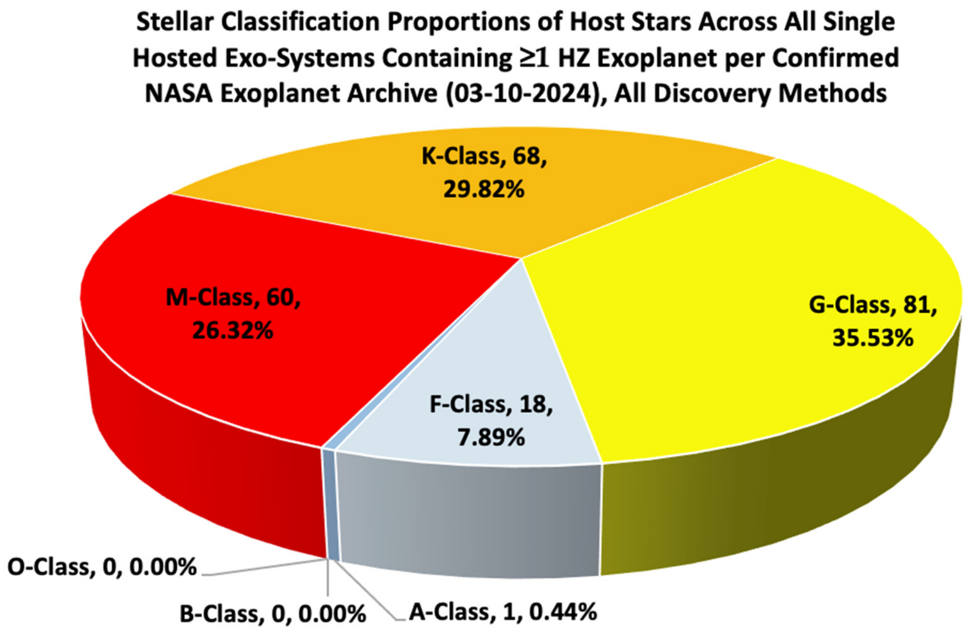


Figure 6. Host stellar class proportions for all single host star exo-systems which contain at least one habitable zone exoplanet. Note: Single hosts of more than one HZ exoplanet are counted multiply according to HZ exoplanet count in those exo-systems.

Figure 7 compares the relative abundance of M, K, G, F, A, and B stellar classes in the Milky Way galaxy to the stellar classes of confirmed exoplanet host stars in single-host systems. For exoplanet host stars, the proportions are 7.8% M-type, 24.7% K-type, 47.4% G-type, 19.4% F-type, 0.6% A-type, and 0.2% B-type. In contrast, the overall abundance in the Milky Way is 76.5% M-type, 12.1% K-type, 7.6% G-type, 3.0% F-type, 0.6% A-type, and 0.13% B-type, this based on the Harvard Spectral Classification system (Las Cumbres Observatory) [26]. This discrepancy highlights a selection bias towards G-type stars, which are similar to our Sun and are often prioritized in exoplanet searches due to their potential for habitability. The underrepresentation of M-type star hosts further underscores the observational challenges and biases in current exoplanet surveys. The preference for G-type stars also reflects historical biases and the assumption that solar analogs are more likely to host habitable planets [27].

Examining exoplanet discovery methods further, Figure 8 shows the proportion of single-hosted exoplanets categorized by these various methodologies. The breakdown is as follows: Transit 76.20%, Radial Velocity 17.81%, Pulsation Time Variation 0.04%, Pulsar

Timing 0.12%, Orbital Brightness Modulation 0.18%, Microlensing 3.95%, Imaging 1.12%, Astrometry 0.04%, Disk Kinematics 0.02%, and Transit Timing Variations 0.53%.

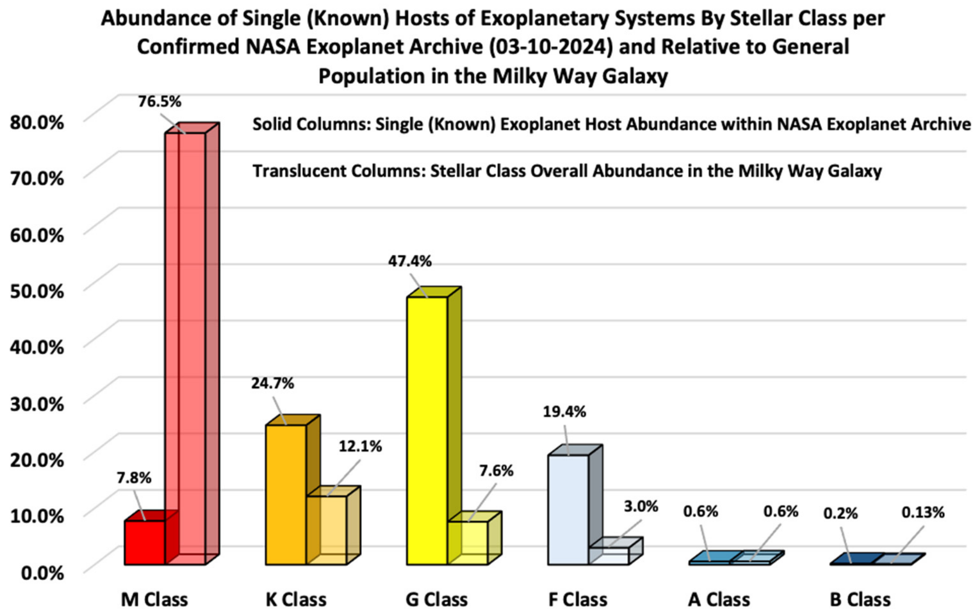


Figure 7. Relative abundance of M, K, G, F, A and B stellar classes in the Milky Way galaxy to the stellar classes of confirmed exoplanet host stars of single host systems. Note: The host star of each singly-hosted exoplanetary system is counted only once, regardless of how many exoplanets in that system are hosted, or their HZ status.

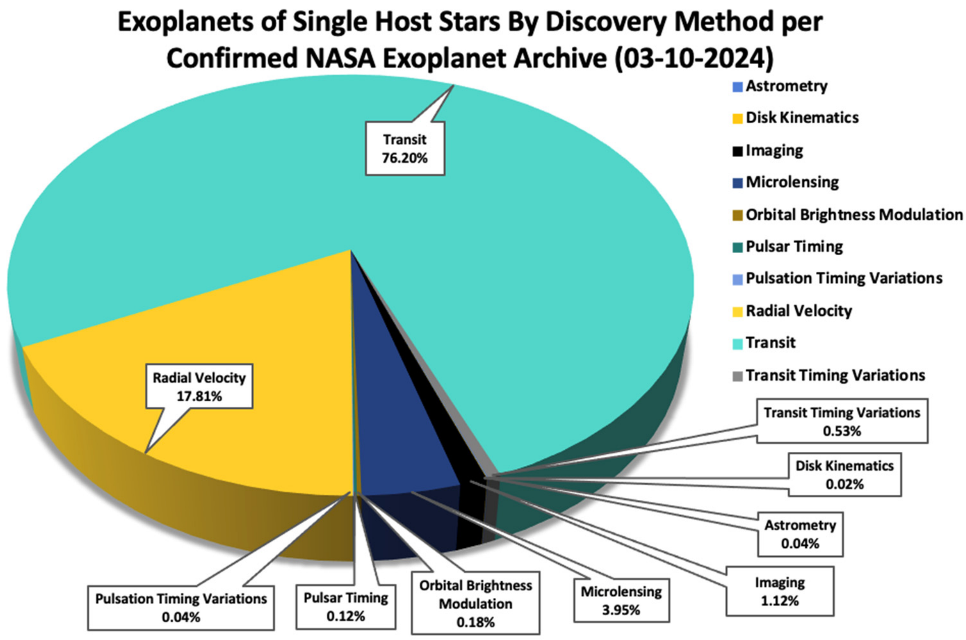


Figure 8. Single-hosted exoplanets apportioned by discovery method.

The dominance of the Transit method reflects its efficiency in detecting exoplanets, particularly those close to their host stars. This method's prevalence in exoplanet discovery highlights its strengths in surveying large areas of the sky and detecting numerous exoplanets, although it also emphasizes the need for complementary methods to provide a more complete picture of exoplanetary systems [11]. The significant presence of exoplanets discovered via microlensing and imaging methods showcases their role in identifying planets at greater distances and in different stellar environments [28]. Note: Appendix B,

Figure A2 focuses the analysis of Figure 7 specifically to HZ exoplanets, illustrating the co-dominance and near parity in terms of overall counts between the Transit and Radial Velocity methods.

Figure 9 shows the habitable zone width as a function of host star mass, including only those hosts with mass known to ≤10% uncertainty. The Sun is included for reference as a large yellow dot. Consistent with HZ calculations elsewhere in this analysis, the Radiative Equilibrium based Equation (1) was utilized along with application of the $T_{surf,\ ave}$ limits of 100 °C and 0 °C for defining the inner and outer HZ boundaries, respectively. Using the Sun as a standard host for model comparison, our finding of 0.51 AU for HZ width is substantially narrower than the 0.71 AU width predicted by the Kopparapu et al. (2013) [4] model or that of the Kasting et al. (1993) [3] model, 0.72 AU. A closer look at the more recent of the two notes that the Kopparapu et al. (2013) [4] model places the geometric midpoint of its HZ at an orbital distance of 1.35 AU for Earth-like planets hosted by Sun-like stars, while our model's midpoint of 0.85 AU more closely aligns with that of the actual Earth-Sun distance. Not surprisingly, all models' general reliance on necessarily elastic assumptions for exoplanetary atmospheric parameters plays a central role in explaining their differing results. As detailed exoplanet atmospheric data is still largely pending collection from more sensitive instrumentation and analysis, it is anticipated these discrepancies will continue to narrow towards eventual resolution as observational capability further improves.

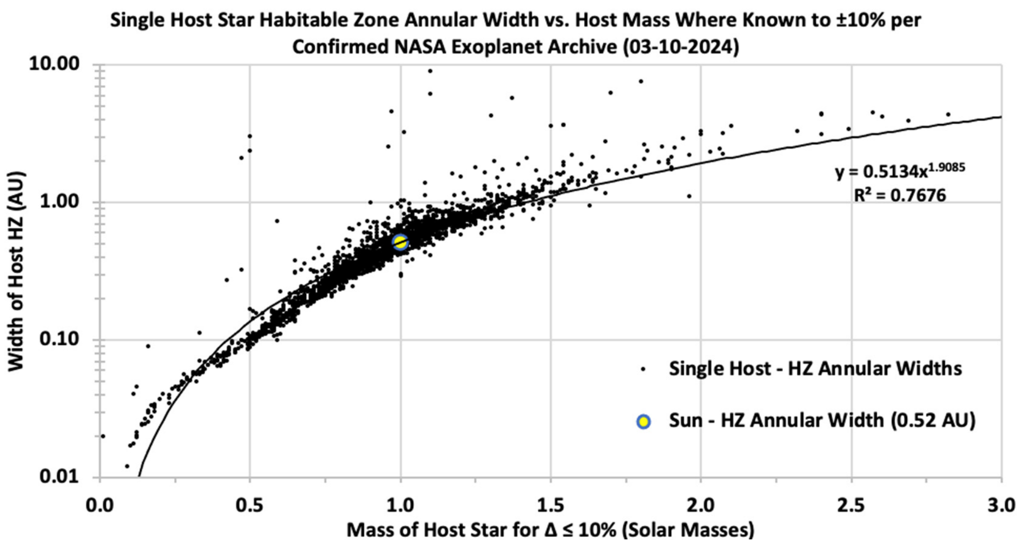


Figure 9. Habitable zone width of exoplanet single host stars as a function of host mass, including only hosts with mass known to ≤10% uncertainty. Sun (large yellow dot) included for reference.

The figure indicates that the width of the habitable zone increases with the mass of the host star and can be empirically approximated as a simple power function that varies as 0.51 × (Host Star Mass in Solar Masses)$^{1.91}$ and is statistically fit correlated to 0.77. This relationship aligns with theoretical models where more massive stars have broader habitable zones due to their higher luminosities, which affect the range of distances at which liquid water could exist on a planet's surface [3]. This finding suggests that more massive stars may offer wider potential zones for habitability, though they also present challenges such as shorter lifespans and higher levels of stellar activity [29]. The broader HZ around massive stars implies that planets can orbit at greater distances while still maintaining surface conditions conducive to liquid water, but the increased stellar radiation and shorter stellar lifetimes may limit long-term habitability [4].

Figure 10 illustrates the variation of habitable zone boundary distances from the host star as a function of host mass, overlaid with exoplanet semi-major axes corresponding to host mass. The inner (red dots) and outer (blue dots) boundaries of the HZ are shown alongside the semi-major axes of exoplanets. The Solar System's planets, as well as the Sun's similarly calculated HZ inner and outer HZ boundaries, are depicted for reference.

As in the above analysis, Equation (1) is used to calculate HZ boundaries with the assumed limitations of 0 to 100 °C on average exoplanet surface temperature for liquid H_2O to exist again being imposed. This visualization demonstrates how the HZ boundaries expand outward with increasing host star mass, while the distribution of exoplanet semi-major axes implies a tendency for planets to reside closer to their stars in lower-mass systems and further away in higher-mass systems. This pattern reflects the influence of stellar mass on planetary formation and orbital dynamics, providing insights into the distribution of potentially habitable exoplanets across different stellar environments [4]. The data suggest that in lower-mass systems, HZ planets are more likely to be found closer to the star, within the narrower HZ, while in higher-mass systems, HZ planets can be situated further out within the broader HZ, potentially offering a more stable environment for life [7].

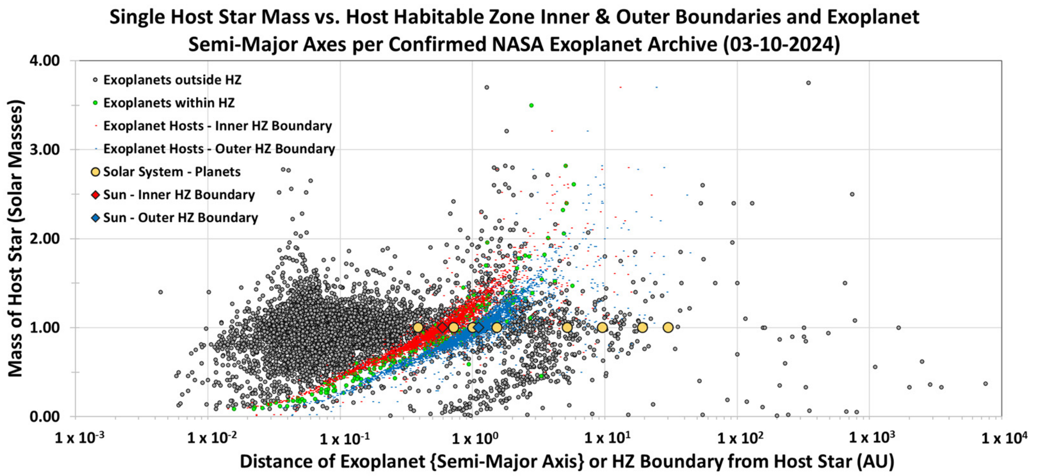


Figure 10. Variation of habitable zones' boundary distances from host star for single hosts as functions of host mass, overlaid with exoplanet semi-major axis and corresponding host mass.

Figure 11 illustrates the relationship between the effective temperatures of single host stars and the surface temperatures of their corresponding exoplanets, with exoplanets grouped by type: Gas Giants, Neptunian Planets, Super-Earths, and Terrestrial Planets. The Earth's position is labeled for reference, along with those for the Solar System's other planets, with the habitable zone range indicated within the vertical green dashed lines. The size of each point represents the relative size of the exoplanet. An exoplanetary system of particular note is TRAPPIST-1, consisting of seven confirmed planets and the host itself, an ultra-cool (effective surface temperature of 2566 K) red dwarf much smaller in radius than the Sun. Notwithstanding, the TRAPPIST-1 system contains two HZ planets and three of terrestrial size, one of which is among the HZ pair as calculated using Equation (1). Given its intriguing array of exoplanets, this system has been one of the more closely studied since its discovery in 2016.

This figure provides a comprehensive visualization of the relationship between stellar temperatures and exoplanet surface temperatures. It shows that few exoplanets fall within the HZ range, indicating the rarity of conditions suitable for surface liquid water. Higher stellar temperatures generally correspond to higher planetary surface temperatures, as evident from the upward trend of data points. Gas Giants and Neptunian Planets predominantly lie outside the HZ, while Super-Earths and Terrestrial Planets show a greater propensity to occupy or approach the HZ. This visualization underscores the need for advanced observational technologies and methodologies to discover and characterize in greater detail (e.g., atmospheric composition) exoplanets within the HZ. Future missions equipped with direct imaging capabilities and improved sensitivity are essential to identifying and studying potential Earth-like planets in habitable zones of a broader range of stellar types.

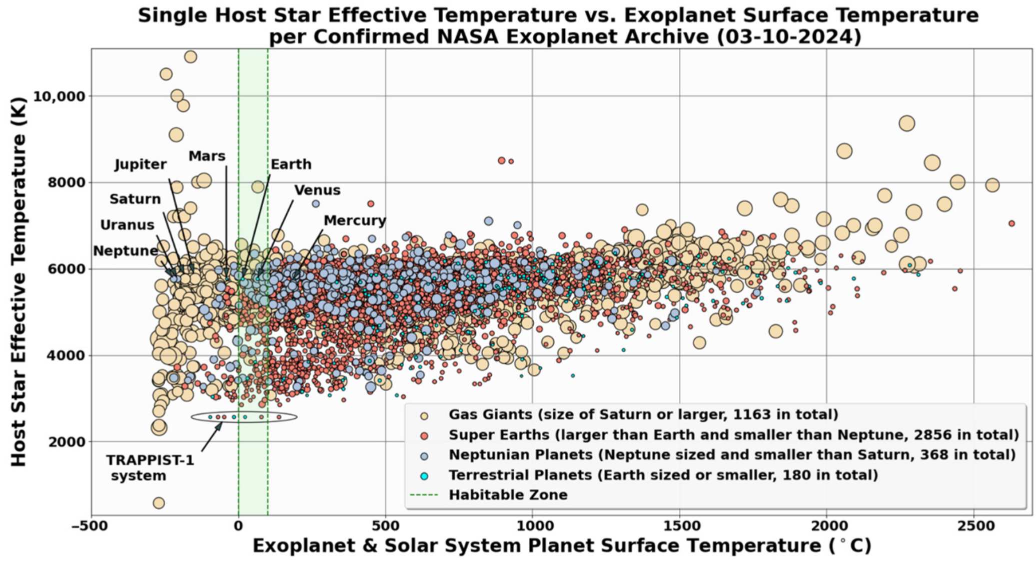


Figure 11. Single host star effective temperature vs. exoplanet and solar system planet surface temperature. This scatter plot shows the effective temperature of single host stars versus the surface temperature of their corresponding exoplanets. Exoplanets are grouped by types: Gas Giants (size of Saturn or larger), Neptunian Planets (Neptune-sized, smaller than Saturn), Super-Earths (larger than Earth, smaller than Neptune), and Terrestrial Planets (Earth-sized or smaller). The positions of the Earth and all other solar system planets are labeled for reference, along with the habitable zone (HZ) range indicated by green dashed lines. The size of each point represents the relative size of the exoplanet. Note: The seven known planets of the TRAPPIST-1 system are indicated at lower left.

This figure also brings attention to assumption limitations introduced earlier in this paper. For the planets in our solar system, surface temperatures calculated with Equation (1) generally align with known mean temperatures (https://science.nasa.gov/resource/solar-system-temperatures/ accessed on 22 July 2024) with a 6–37% difference. However, Venus is an exception. Although its surface is too hot for life as we know it, Equation (1) flags the planet as within the habitable zone. This discrepancy arises from the assumption of a standardized bulk temperature factor (k) based on Earth's value of k = 1.13 when accounting for the atmospheric greenhouse effect. In reality, Venus has a very thick atmosphere composed primarily of CO_2, trapping heat and resulting in a much higher bulk temperature factor of k = 3.17. This limitation is discussed earlier in Section 2.6. Bracketing the inner Solar System-based atmospheric greenhouse assumption, this on the cooler end, is Mars. While the atmosphere of Mars is also predominantly composed of CO_2, it is far less dense and accordingly much less capable of trapping solar radiation. The particular exception of Venus indicates that variations in the atmospheric greenhouse effect will need to be observed and further considered to better determine exoplanet surface temperatures.

Overall, the analysis presented above highlights several observational biases inherent in current exoplanet detection methods. The Transit method, which dominates exoplanet discoveries, is more likely to detect planets with shorter orbital periods, leading to an overrepresentation of "Too Hot" exoplanets. This bias is evident in the high percentage of "Too Hot" planets discovered via Transit methods (Figures 4 and 5a). Conversely, the Radial Velocity method, which can detect exoplanets at various distances from their host stars, presents a more balanced distribution of habitable zone statuses (Figure 5b).

The stellar classification of host stars (Figures 6 and 7) shows a preference for G-type stars in exoplanet searches, despite M-type stars being by far the most common in the Milky Way. This selection bias may be due to the more stable and long-lived nature of G-type stars, which are conducive to sustaining life-supporting environments over extended periods.

Figures 9 and 10 provide insights into the relationship between host star mass and habitable zone characteristics. The widening of the habitable zone with increasing host star mass suggests that more massive stars offer a larger range of distances where conditions might support surface liquid water. However, the semi-major axis distribution of exoplanets indicates a concentration of planets closer to lower-mass stars, potentially due to the higher likelihood of detecting such planets through current observational techniques—again echoing the dominance of the Transit method.

Figure 11 adds additional context by illustrating the relationship between the effective temperatures of host stars and the surface temperatures of their corresponding exoplanets. This figure highlights the challenges in finding planets within the HZ and underscores the importance of advanced observational technologies to overcome current limitations.

Overall, these results emphasize the need for continued development and deployment of diverse detection methods to achieve a more comprehensive understanding of exoplanetary systems. Future missions should aim to mitigate observational biases by targeting a broader range of stellar types and distances, thereby enhancing our ability to identify potentially habitable exoplanets and build a more complete understanding of planetary systems in general.

4. Summary and Discussion

The investigation into the habitability of exoplanets within circumstellar habitable zones (HZs) has provided valuable insights and contributed to the broader understanding of exoplanetary science. By analyzing 5595 confirmed exoplanets from the NASA Exoplanet Archive, this study has highlighted several key findings that advance our knowledge of potential habitable worlds beyond our Solar System.

The exponential growth in exoplanet discoveries underscores the advancements in detection technologies and methodologies, especially with contributions from missions such as the Kepler/K2 Space Telescope and the Transiting Exoplanet Survey Satellite (TESS). This growth trend indicates that our capability to identify exoplanets, with an ongoing emphasis on those orbiting within their host HZ, will continue to improve with future technological advancements in both data gathering missions and analysis.

One of the significant findings from this study is the importance of addressing observational biases inherent in different exoplanet detection methods. The Transit method, while efficient in detecting numerous exoplanets, shows a bias towards closer-in, hotter planets. In contrast, the Radial Velocity method provides a more balanced view but still favors larger planets. This highlights the need for employing a combination of detection methods to achieve a comprehensive understanding of exoplanet populations and to identify potentially habitable exoplanets more effectively.

The study also emphasizes the crucial role of stellar classifications in determining exoplanet habitability. G-type and K-type stars are shown to host a higher proportion of habitable zone exoplanets, aligning with their stable lifetimes and favorable conditions for surface liquid water. However, the underrepresentation of M-type stars, despite their general abundance, points to the challenges and potential habitability issues associated with these stars, such as stellar activity and tidal locking.

Moreover, the analysis of habitable zone widths as a function of host star mass reveals that more massive stars have broader habitable zones due to their higher luminosities. This suggests that such stars may offer wider potential zones for habitability, though their shorter lifespans and higher levels of stellar activity could pose challenges for long-term habitability. This finding aligns with theoretical models and provides a deeper understanding of how stellar characteristics influence planetary habitability.

The discrepancy between the stellar class distribution of confirmed exoplanet host stars and that of the Milky Way's general distribution indicates a selection bias towards Sun-like G-type stars. This historical focus on solar analogs for habitability studies underscores the necessity of broadening survey efforts to include a wider range of stellar types, thus providing a more comprehensive view of potential habitable planets across different host star environments.

Despite robust theoretical modeling, more thorough analysis of exoplanetary systems—and particularly the potential habitability of planets therein—requires additional detailed observational data. As Ramirez (2018) [30] suggests, "this situation would vastly improve with technological (e.g., engineering) advancements in observational techniques, including direct imaging. Even with current limitations, we have suggested some observations that can still be made with upcoming and next generation missions". It is vital to continue gathering and studying data that encompass the broad spectrum of possibilities.

Another aspiration of more in-depth analysis is the unification of the Circumstellar Habitable Zone, Galactic Habitable Zone, and Cosmic Habitable Age concepts. The ultimate goal is to formulate a comprehensive methodology for determining if any strongly Earth-like planets exist, such as that proposed by Cai et al. (2021) [31], which suggests that planets most likely to harbor life are located in an annular region approximately four kpc from the galactic center. Integrating each piece of the puzzle leads us towards a better understanding of where life might exist in the galaxy. The years ahead are expected to prove particularly enlightening as more powerful ground and space-based instrumentation comes online, along with advancements in computational techniques for determining the habitability of worlds outside the Solar System.

In conclusion, this study enhances existing models for managing burgeoning exoplanetary datasets and lays foundational groundwork for future explorations into the dynamic relationships between exoplanets and their stellar environments. By addressing observational biases, employing diverse detection methods, and considering the influence of stellar classifications, this research contributes to a more nuanced understanding of exoplanet habitability. The findings underscore the importance of multifaceted approaches and enhanced observational capabilities in the ongoing search for environments which may host extraterrestrial life and/or eventually be colonized by humans. Future research will benefit from these insights, guiding efforts to uncover the full spectrum of potentially habitable exoplanets and advancing our quest to understand our place in the cosmos.

Author Contributions: Conceptualization, J.H.J. and P.E.R.; methodology, J.H.J. and P.E.R.; software, P.E.R. and C.X.L.; validation, J.H.J. and P.E.R.; formal analysis, J.H.J., P.E.R., C.X.L. and Q.W.; investigation, J.H.J., P.E.R. and C.X.L.; resources, J.H.J. and Y.C.; data curation, P.E.R.; writing—original draft preparation, J.H.J., P.E.R. and C.X.L.; writing—review and editing, J.H.J., P.E.R. and C.X.L.; visualization, P.E.R., C.X.L. and Q.W.; supervision, J.H.J.; project administration, J.H.J.; funding acquisition, J.H.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the NASA Exoplanet Research Program NNH22ZDA001N-XRP.

Data Availability Statement: The data underlying this article can be downloaded from the NASA Exoplanet Archive (https://exoplanetarchive.ipac.caltech.edu accessed on 10 March 2024) and the Habitable Worlds Catalog (HWC), PHL @ UPR Arecibo (https://phl.upr.edu/hwc/data accessed on 6 September 2024). The method of data calculation, data analysis, and ML model training and evaluation are fully described in the article.

Acknowledgments: This research was conducted at the NASA sponsored Jet Propulsion Laboratory, California Institute of Technology (Caltech). It has made use of the NASA Exoplanet Archive, which is operated by Caltech, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program. The authors also thank the Pacific Academy at Irvine, California, and Lakeside School at Seattle, Washington for their support.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Random Forest and XGBoost Model Training and Evaluation Results

The Habitable Worlds Catalog (HWC), PHL @ UPR Arecibo dataset contains more than 5599 exoplanets as of 6 September 2024. The potentially habitable exoplanets are identified with the P_HABITABLE data field in the dataset: P_HABITABLE=1 indicates conservative habitable (more likely to be rocky planets capable of surface liquid water) and P_HABITABLE=2 indicates optimistic habitable (might include water worlds or mini-Neptunes), while P_HABITABLE=0 indicates non-habitable. We joined this dataset with the Planetary Systems Composite Data from the NASA Exoplanet Archive and used P_HABITABLE to mark the labels for the training data (P_HABITABLE=1 or 2 as habitable and P_HABITABLE=0 as non-habitable).

The Planetary Systems Composite Data has a data field called pl_controv_flag indicating whether the confirmation status of an exoplanet has been questioned in published literature. As the first step, we filtered out any exoplanets with pl_controv_flag=1. We also limited our training data to only contain planetary systems with one host star. We then removed data fields that are irrelevant to the training; for instance, pl_name (planet name), disc_year (discover year), etc.

For the remaining data fields, we calculated the missing value percentage for each of them and removed fields missing more than 25% of values. For the rest of the data, we filled the missing values with the mode for categorical data fields and with the Multivariate Imputation by Chained Equation algorithm, known as MICE (Khan et al., 2020) [32] for numerical data fields. We then applied Label Encoding to convert categorical features into numerical values.

Our training data was highly imbalanced and dominated by the negative samples (non-habitable exoplanets), which could be challenging for model training and potentially result in poor model performance. To counter this problem, we used a combination of Synthetic Minority Oversampling Technique (SMOTE) and Edited Nearest Neighbors (ENN) for over-sampling minorities and under-sampling majorities to balance our training data.

To deal with correlated data fields, we applied correlation analysis to identify highly correlated features and removed them. Seventeen data fields remained as the features in the final training data. A MinMaxScalar was also applied to standardize the value ranges of all of the features to between 0 and 1.

During the model training, we randomly shuffled training data obtained from the process above and then split data for training and evaluation. We repeated this process over multiple rounds and calculated the mean accuracy as representative of the model's performance. Hyperparameter tuning was applied in the process to identify the hyperparameters that achieve the best model quality.

Both classifiers achieved high overall accuracies, with the Random Forest at 0.95 and XGBoost at 0.97, respectively. Further breakdown of model performance (precision, recall, F1 score) is shown in the Table A1, below. We can conclude that both the Random Forest and the XGBoost classifiers perform the habitability classification remarkably well.

Table A1. Random Forest and XGBoost model evaluation results (precision, recall, F1 score).

Class	Random Forest			XGBoost		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
1 (Habitable)	0.94	0.97	0.96	0.96	0.98	0.97
0 (Non-Habitable)	0.97	0.94	0.95	0.98	0.96	0.97

Appendix B. Additional Figures

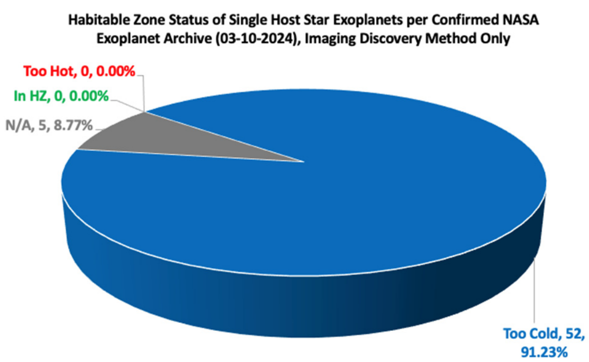


Figure A1. Habitable zone status of single hosted exoplanets discovered by way of the Imaging method.

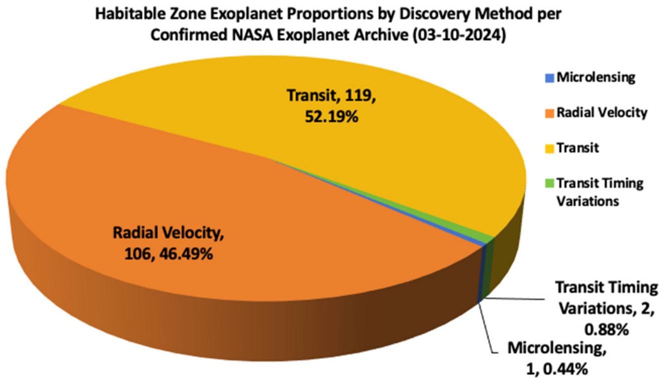


Figure A2. Apportionment of habitable zone exoplanets by discovery method.

References

1. Wolszczan, A.; Frail, D.A. A Planetary System around the Millisecond Pulsar PSR1257+12. *Nature* **1992**, *355*, 145–147. [CrossRef]
2. NASA Exoplanet Archive. NASA Exoplanet Archive: Planetary Systems Composite Data. 2024. Available online: https://exoplanetarchive.ipac.caltech.edu/ (accessed on 10 March 2024).
3. Kasting, J.F.; Whitmire, D.P.; Reynolds, R.T. Habitable Zones around Main Sequence Stars. *Icarus* **1993**, *101*, 108–128. [CrossRef] [PubMed]
4. Kopparapu, R.K.; Ramirez, R.; Kasting, J.F.; Eymet, V.; Robinson, T.D.; Mahadevan, S.; Terrien, R.C.; Domagal-Goldman, S.; Meadows, V.; Deshpande, R.; et al. Habitable Zones around Main-sequence Stars: New Estimates. *Astrophys. J.* **2013**, *765*, 131. [CrossRef]
5. Huang, S.S. Occurrence of Life in the Universe. *Am. Sci.* **1959**, *47*, 397–402.
6. Hart, M.H. Habitable Zones about Main Sequence Stars. *Icarus* **1979**, *37*, 351–357. [CrossRef]
7. Lammer, H.; Bredehöft, J.H.; Coustenis, A.; Khodachenko, M.L.; Kaltenegger, L.; Grasset, O.; Prieur, D.; Raulin, F.; Ehrenfreund, P.; Yamauchi, M.; et al. What Makes a Planet Habitable? *Astron. Astrophys. Rev.* **2009**, *17*, 181–249. [CrossRef]
8. Shields, A.L.; Barnes, R.; Agol, E.; Charnay, B.; Bitz, C.; Meadows, V.S. The Effect of Orbital Configuration on the Possible Climates and Habitability of Kepler-62f. *Astrobiology* **2016**, *16*, 443–464. [CrossRef]
9. Shields, A.L.; Ballard, S.; Johnson, J.A. The Habitability of Planets Orbiting M-dwarf Stars. *Phys. Rep.* **2016**, *663*, 1–38. [CrossRef]
10. Howard, A.W.; Marcy, G.W.; Bryson, S.T.; Jenkins, J.M.; Rowe, J.F.; Batalha, N.M.; Borucki, W.J.; Koch, D.G.; Dunham, E.W.; Gautier, T.N.; et al. Planet Occurrence within 0.25 AU of Solar-type Stars from Kepler. *Astrophys. J. Suppl. Ser.* **2012**, *201*, 15. [CrossRef]
11. Winn, J.N.; Fabrycky, D.C. The Occurrence and Architecture of Exoplanetary Systems. *Annu. Rev. Astron. Astrophys.* **2015**, *53*, 409–447. [CrossRef]
12. Mayor, M.; Queloz, D. A Jupiter-mass Companion to a Solar-type Star. *Nature* **1995**, *378*, 355–359. [CrossRef]
13. Aigrain, S.; Pont, F.; Zucker, S. Machine Learning for Exoplanet Detection: Techniques and Challenges. *Mon. Not. R. Astron. Soc.* **2017**, *467*, 1638–1647. [CrossRef]
14. Petigura, E.A.; Howard, A.W.; Marcy, G.W. Prevalence of Earth-size Planets Orbiting Sun-like Stars. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 19273–19278. [CrossRef] [PubMed]
15. Coughlin, J.L.; Mullally, F.; Thompson, S.E.; Rowe, J.F.; Burke, C.J.; Latham, D.W.; Batalha, N.M.; Ofir, A.; Quarles, B.L.; Henze, C.E.; et al. Planetary Candidates Observed by Kepler. VII. The First Fully Uniform Catalog Based on the Entire 48-month Data Set (Q1–Q17 DR25 Kepler Catalog). *Astrophys. J. Suppl. Ser.* **2018**, *235*, 38. [CrossRef]
16. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
17. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16), San Francisco, CA, USA, 13–17 August 2016.
18. PHL @ UPR Arecibo. The Habitable Worlds Catalog (HWC). 2024. Available online: https://phl.upr.edu/hwc (accessed on 6 September 2024).
19. Dressing, C.D.; Charbonneau, D. The Occurrence of Potentially Habitable Planets Orbiting M Dwarfs Estimated from the Full Kepler Dataset and an Empirical Measurement of the Detection Sensitivity. *Astrophys. J.* **2015**, *807*, 45. [CrossRef]
20. Saha, S.; Basak, S.; Safonova, M.; Bora, K.; Agrawal, S.; Sarkar, P.; Murthy, J. Theoretical validation of potential habitability via analytical and boosted tree methods: An optimistic study on recently discovered exoplanets. *Astron. Comput.* **2018**, *23*, 141–150. [CrossRef]
21. Basak, S.; Mathur, A.; Theophilus, A.J.; Deshpande, G.; Murthy, J. Habitability Classification of Exoplanets: A Machine Learning Insight. *Eur. Phys. J. Spec. Topics.* **2021**, *230*, 2221–2251. [CrossRef]
22. Ghadekar, P.; Chandra, A.; Jiby, B.; Kamble, A.; Arole, R.; Kirar, A.P.S. Analysis of Exoplanet Habitability Using RNN and Causal Learning. In *International Conference on Advancements in Smart Computing and Information Security*; Springer: Cham, Switzerland, 2024; Volume 2037, pp. 71–84.

23. Borucki, W.J.; Koch, D.; Basri, G.; Batalha, N.; Brown, T.; Caldwell, D.; Caldwell, J.; Christensen-Dalsgaard, J.; Cochran, W.D.; DeVore, E.; et al. Kepler Planet-Detection Mission: Introduction and First Results. *Science* **2010**, *327*, 977–980. [CrossRef]
24. Gaudi, B.S.; Seager, S.; Mennesson, B.; Kiessling, A.; Warfield, K.; Cahoy, K.; Clarke, J.T.; Domagal-Goldman, S.; Feinberg, L.; Guyon, O.; et al. The Habitable Exoplanet Observatory (HabEx) Mission Concept Study Report. *arXiv* **2020**, arXiv:2001.06683.
25. Seager, S.; Mallén-Ornelas, G. A Unique Solution of Planet and Star Parameters from an Extrasolar Planet Transit Light Curve. *Astrophys. J.* **2003**, *585*, 1038–1055. [CrossRef]
26. Las Cumbres Observatory. Available online: https://lco.global (accessed on 21 October 2024).
27. Brown, A.G.A. The Gaia Mission: Science, Organization and Present Status. *Proc. Int. Astron. Union* **2015**, *11*, 29–38.
28. Fukui, A.; Gould, A.; Sumi, T.; Bennett, D.P.; Bond, I.A.; Han, C.; Suzuki, D.; Beaulieu, J.P.; Batista, V.; Udalski, A.; et al. OGLE-2012-BLG-0563Lb: A Saturn-Mass Planet Around an M Dwarf with the Mass Constrained By *Subaru* AO Imaging. *The Astrophys. J.* **2015**, *809*, 74. [CrossRef]
29. Selsis, F.; Kasting, J.F.; Levrard, B.; Paillet, J.; Ribas, I.; Delfosse, X. Habitable Planets around the Star Gliese 581? *Astron. Astrophys.* **2007**, *476*, 1373–1387. [CrossRef]
30. Ramirez, R.M. Exploring Exoplanet Surface Habitability at Different Stellar Energies. *Nat. Astron.* **2018**, *2*, 32–37.
31. Cai, X.; Jiang, J.H.; Fahy, K.A.; Yung, Y.L. Statistical Estimation of the Occurrence of Extraterrestrial Intelligence in the Milky Way Galaxy. *Galaxies* **2021**, *9*, 5. [CrossRef]
32. Khan, S.I.; Hoque, A.S.M.L. SICE: An improved missing data imputation technique. *J. Big Data* **2020**, *7*, 37. [CrossRef]