Demystifying Exoplanet Habitability Predictions with XAI

Sajjad Hossain Shanto
Dept of Computer Science and Engineering
BRAC University
sajjad.hossain.shanto@g.bracu.ac.bd

Fahad Nadim Ziad
Dept of Computer Science and Engineering
BRAC University
fahad.nadim.ziad@g.bracu.ac.bd

***Abstract*—The search for life on other planets has become one of the focal activities in the subject of astrobiology and planetary science. Although thousands of exoplanets have been identified, finding those planets that might support life is less easy. This study employs explainable artificial intelligence (XAI) techniques to predict the habitability of exoplanets by analyzing planetary and stellar attributes. Machine learning models such as Random Forest, Support Vector Machine (SVM), Logistic Regression, Gradient Boosting Frameworks and Neural Network were employed and evaluated using metrics such as Accuracy, Precision, Recall, and Area Under the Curve (AUC). From all models LightGBM yielded the highest accuracy of 99.93% and LightGBM, Random Forest, XGBoost produced the best AUC result of 99.99%. XAI techniques identified critical factors influencing habitability predictions, including surface temperature, atmospheric composition, and stellar radiance. In addition to enhancing the completeness of habitability classification, this work provides explanations to enhance the help readiness for the selection of further observing objects by astronomers.**

***Index Terms*—Exoplanet Habitability, Machine Learning, Data Analysis, Feature Extraction, LIME, SVM, Logistic Regression, Neural Network Random Forest, Multilayer Perception, Light GBM**

I. Introduction

This definitely can be said about the study of exoplanets, or the planets, which orbit stars other than the Sun, and has changed significantly the way scientists view the existence of life in the universe. Significant progress in astronomical observations for decades has revealed thousands of exoplanets that invited attention to their physical properties, atmosphere, and ability to support life. Nevertheless, establishing which of these planets may harbor life is quite challenging, and this calls for merging of big and complex datasets.

Planetary habitability is primarily determined by factors such as the distance from the host star (habitable zone), atmospheric composition, surface temperature, and the presence of liquid water. The assessment of these factors usually requires elaborate modeling, yet the volumes of data available today call for more extensive computational solutions than others.

This research addresses this challenge by employing machine learning (ML) algorithms, complemented by explainable artificial intelligence (XAI) methods, to evaluate exoplanet habitability. Due to the application of statistical patterns and properties of decision making used in ML, this research does not only improve outcomes in accuracy but also the feasibility of model interpretation where the logic behind any outputs is well clarified. It is about the identification of watershed discoveries that can become crucial for the further observational missions and the reconsideration of the main parameters of habitability.

The next sections provide a detailed look into the study methodology, findings, and conclusion, as well as a potential roadmap for better and more easily interpretable models for habitability computation.

II. Literature Review

There are many studies for predicting exoplanet habitability by using various factors. Here Most work in the field of astrobiology focuses specifically on the possibility of habitable exoplanets, which is the study of the circumstances by which planets in other systems might be capable of sustaining life. These include the distance of the planet from the star, the planet's atmosphere, the surface conditions, and radiation intensity all determine whether the planet could support life. In these last years, improvements in the observation methods and the use of machine learning algorithms have helped enhance our ability in estimating the habitability of exoplanets. Keith Cowing [2] reviewed the relationship between the range of planetary and stellar characteristics and their influence on exoplanet habitability. In this research the distance from the star, its type and the level of radiation are the basic aspects that define the habitability. Like what Cowing pointed out, knowing the environment of the star is of high significance to the probability of liquid water on the planet. In a more elaborate work, Jiang et al [6] present information on observed exoplanets and the stellar habitable zone. They use various stars and their planets to set bare bones parameters about what makes a planet habitable, the temperatures and magnetic storms of a star. These results will help refine our understanding of how the properties of stars affect the habitability of planets orbiting them. Various data sets have been used to predict the habitability of exoplanets using Machine learning (ML) methods. In the research by Yash Patel [3] , he provided a detailed review of different ML based benchmarks for modeling exoplanet habitability and the utilization of a classification-based model. This work uses models including LR, RF, and SVM to measure the probability of finding hospitable scenarios based on data including but not limited to surface temperature, atmospheric makeup, and distance from the star. Patel's research akso proves that machine learning models are useful when dealing with massive datasets and screening out possible exoplanets for habitability. Furthermore, Tanaya Patil and Gautam Patil [4] analyzed relationships between the advanced ML features with reference to the habitability of exoplanets. When the authors included stellar radiation, planetary radius, and surface parameters, they used ensemble learning techniques in order to enhance generalization. They used Random Forest, Decision Tree for classification. Their approaches explain how, using the concept of bagging, which involves assembling many weak models to come up with a more robust predictor, they were able to capture how planets may probably be habitable. Common usage of the data from other sources is a significant issue for the correct habitability calculation. R.Sutharsana et al. [5] used regression analysis based on data collected usually with observations and machine learning to study exoplanet habitability, the radiation of stars and various conditions of a planet. Their strategy aims at employing different classifiers for example Decision trees and other classify model to classify the conditions that supports life. Finally, in a more innovative research, T. Van Hoolst et al[7] analyzed the study of exoplanet interiors explores their composition, formation, and evolution, emphasizing the role of interior processes in maintaining conditions for water on the surface or subsurface. With advancements in space missions, more exoplanets are being discovered. Machine learning models such as Random Forest, Support Vector Machines (SVM), Neural Networks, and Decision Trees can help predict exoplanet habitability based on various parameters.

III. METHODOLOGY

A. Dataset Description:

This study uses the Habitable Worlds Catalog (HWC), maintained by Planetary Habitability Laboratory (PHL) at the University of Puerto Rico at Arecibo. The most comprehensive and up to date stellar and planetary data compilation for comparative analysis of potentially habitable exoplanets. The dataset here contains many features pertinent to assessing planetary habitability (e.g., stellar properties, planetary properties: mass, radius, density, surface temperature etc.), habitability metrics (e.g. the Earth Similarity Index (ESI), and habitable zone classifications). The HWC combines data from many authoritative sources such as the NASA Exoplanet Archive, supplemented with corrections obtained from the scientific literature and other databases to be reliable and accurate. The dataset provides important information to continuously advance research in planetary habitability by providing a detailed, well structured dataset for analysis and modeling.

B. Data preprocessing:

- To uniformize features in the dataset across surface temperature (Kelvin), planetary radius (meters) and stellar luminosity (watts) across the dataset, it has been standardized to SI (MKS) units. We removed irrelevant, redundant or low predictive value features such as unique identifiers (P_NAME, S_NAME) and present seventy percent or more missing values for metric. Mixed unit systems were avoided and clear standardized abbreviations were used, thus maintaining dimensional consistency.
- Mode based imputation was made on categorical variables and K nearest neighbors (KNN) imputation was used for numerical features such as P_TEMP_SURF. In places where features with interdependencies exist (e.g. S_LOG_G and S_LOG_LUM) these features were imputed iteratively to preserve relationships, as well as maintain data integrity.
- The distribution of Numerical features were taken into consideration while scaling via Standard Scaling or Min-Max Scaling. The preparation of categorical features like S_TYPE was completed by transforming these features into numerical values through labeling encoding, to be ready for the downstream modeling tasks.

C. Learning phase:

During the learning phase, multiple machine learning models were trained and evaluated to predict exoplanet habitability using a variety of metrics for robust performance assessment. The following models were utilized:

a) Gradient Boosting Frameworks: Gradient boosting algorithms like XGBoost, LightGBM, CatBoost successively add decision trees to correct errors made from previous iterations. The advanced optimization technique based models make use of these techniques for better predictions and least loss. In this study, gradient boosting was used to effectively model complex stellar and planetary feature interactions.

b) Random Forest: In fact, random forests are an ensemble of decision trees deployed to maximize the prediction accuracy and generalization over individual decision trees. This method reduces overfitting by joining many trees together, and also increases robustness. I apply a random forest classifier to model the complex relationships between stellar and planetary attributes in this work.

c) Logistic Regression: Logistic regression is a linear statistical model which predict binary outcome yielding result using the relation between independent variables and target variable. To facilitate interpretation, we use logistic regression to model the probability of habitability in this study, as it is both simple and interpretable for their contribution on feature.

d) Neural Network: A Neural Network is a Multilayer Perceptron (MLP) which consists of interconnected layers of neurons who learn from data. An MLP classifier was employed for the purpose of predicting habitability, with learning rate, number of layers optimized via training.

e) KNN: KNN is a non parametric algorithm which classifies data points based on the proximity to other labeled points. Using this model we addressed local patterns in the data, as its value is sensitive to the number of neighbors selected (k). Rigorous hyperparameter tuning, stratified sampling, and SMOTE to handle class imbalance were used on the training and validation phases. The results of evaluation for these models are presented in the Results and Discussion section showing comparative performance and prediction capability for exoplanet habitability classification.

f) SVM: SVM is a supervised learning algorithm that identifies the optimal hyperplane to separate classes with maximum margin. This method is well-suited for handling high-dimensional datasets. In this study, SVM was used to classify planets as habitable or nonhabitable.

D. *Figures and Tables*

Figures presented in this section highlight the key findings, feature contributions, and comparative performance metrics of the models.

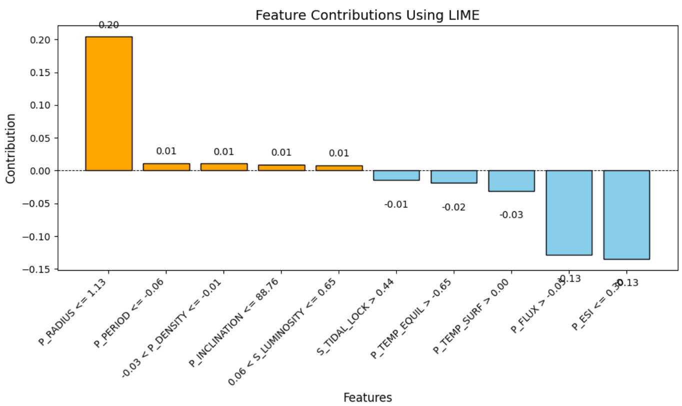

Fig. 1. Feature Contributions Using LIME.

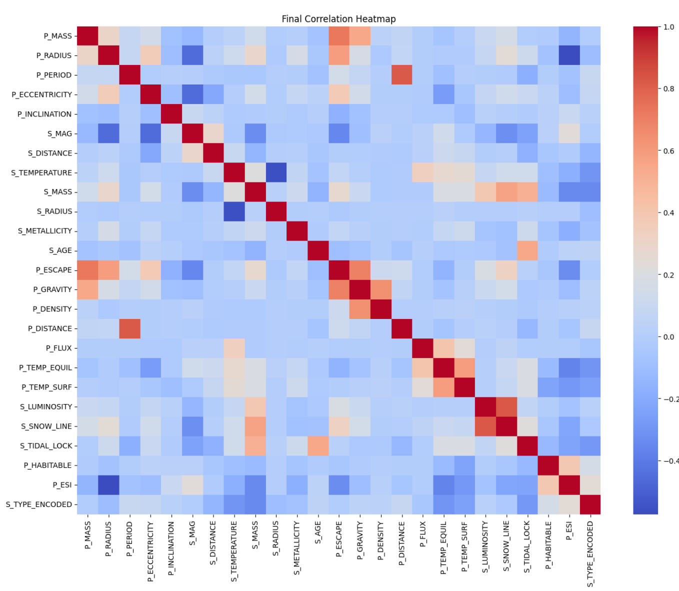

Fig. 2. Correlation Heat map.

IV. RESULTS AND DISCUSSION

In this section, we report the results of our machine learning models and our experience during the project. The greatest challenge was the class imbalance in the exoplanet dataset, where the category **Conservative Habitable** and **Optimistic Habitable** only make up a small fraction 15% and 9%, respectively of the data. In order to do this, SMOTE was applied to mitigate the classes imbalance, so that the models has a more fair training set to learn from. Even more, it was necessary to hyperparameter tune models like LightGBM, XGBoost, and Random Forest to optimize performance. The fine tunning of the models proceeded even with the over fitting and generalizability problems, however with promising results in habitability classification. However these results and interpretability tools are important in advancing the field and also are the results that are the foundation for continued work.

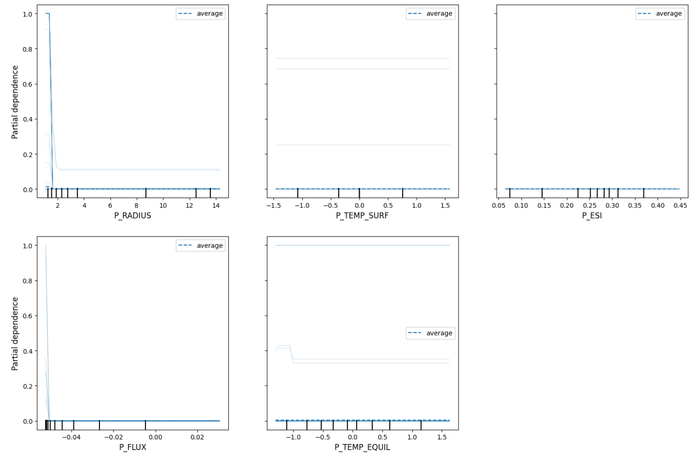

Fig. 3. Partial Dependence Plots.

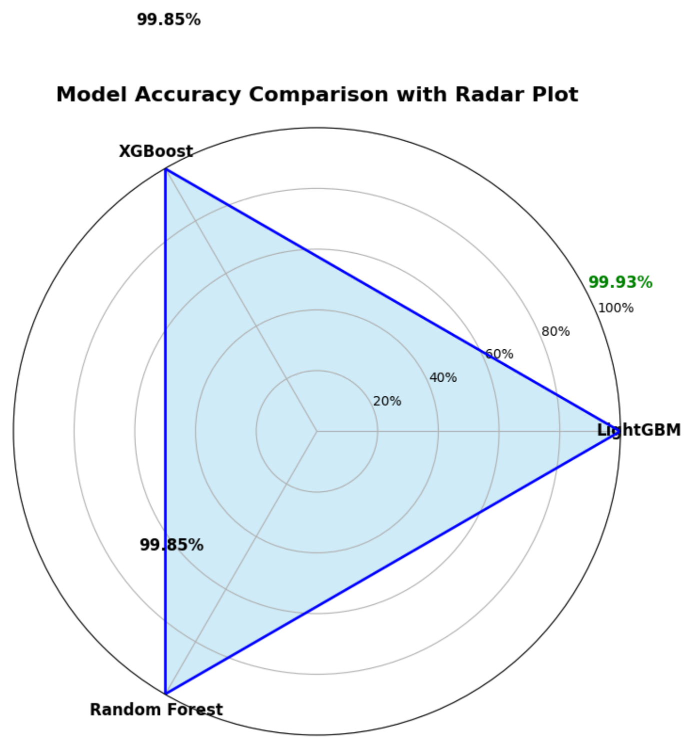

Fig. 4. Accuracy Percentages (Fine Tuned).

A. Feature Importance

The classification of exoplanet habitability was performed to identify important predictors in the feature importance analysis. The feature contributions were validated via several methods including correlation analysis, permutation based importance, and interpretability methods such as LIME.

Figure 2 shows a correlation heatmap generated to see how features in the dataset are related to one another. The heatmap did provide some knowledge about dependencies, for instance a moderate **P_RADIUS** (Planet Radius) to **P_FLUX** (Flux) correlation, but they offered little direct insight into how these would factor into habitability classification.

Permutation-based feature importance highlighted the most critical features driving model predictions. Notably:

- **P_ESI** (Earth Similarity Index) and **P_FLUX** ranked as the most significant, underscoring their astrophysical relevance to habitability.
- Other features, including **P_TEMP_SURF** (Surface Temperature), **P_RADIUS**, and **P_PERIOD** (Orbital Period), also demonstrated strong contributions, reflecting their direct association with conditions favorable for life.

The relationships between key features and predicted probabilities were further analyzed using Partial Dependence Plots (PDPs), shown in Figure 3. These plots revealed:

- **P_RADIUS**: A sharp decline in habitability likelihood with increasing radius, consistent with the hypothesis that smaller planets have higher habitability potential.
- **P_TEMP_SURF**: Threshold effects were evident, where certain temperature ranges significantly increased habitability probabilities.
- **P_ESI**: Consistent with permutation-based findings, this feature exhibited a steep gradient, highlighting its critical role.

Local explanations for individual predictions were provided through LIME. Figure 1 shows that features such as **P_RADIUS**, **P_TEMP_SURF** and **P_ESI** were among the top contributors to each sample. The explanations from LIME reaffirmed these global importance rankings and contributed a level of transparency by showing how these features affect some predictions.

Additional techniques, such as feature selection via statistical tests and tree based classifiers, were explored to additionally cross validate the results. Finally, the astrophysical features of **P_ESI** and **P_FLUX** are consistently highlighted with these methods, attesting to the concordance between our results and those of our permutation and LIME based analyses.

B. Performance Metrics

The dataset was evaluated using different machine learning algorithms for predicting exoplanet habitability. The investigation encompassed multiple models commonly employed in machine learning: Random Forest, XGBoost, LightGBM, CatBoost, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Neural Network. Model performance assessment utilized three key metrics: The measures used were accuracy, F1-Score, and ROC-AUC. Prediction correctness is quantified by accuracy, which computes the ratio of correctly predicted instances as a fraction of total instances. The F1 score, obtained from harmonic mean of precision and recall, makes for robust evaluation of imbalanced datasets. ROC-AUC values predict the ability of the classification to utilize sensitivity and specificity values.

Initial model evaluation results are presented in Table I. It was found that different algorithms performed well under different scenarios. Random Forest was balanced with identical 0.8571 precision and recall values, a resulting F1-Score of 0.8571. CatBoost achieved an optimal precision (1.0000) but lower recall (0.7143). SVM and KNN also performed very suboptimally in class prediction with their low F1 Scores of 0.2000 and 0.2400, respectively. Results from the first run revealed the need to optimize the model.

Table II presents the performance metrics after hyperparameter tuning, focusing on the three most promising models. LightGBM achieved exceptional performance with perfect scores across all metrics (precision, recall, and F1-Score of 1.0000) and 99.93% accuracy. XGBoost demonstrated substantial improvement with 99.85% accuracy and an F1-Score of 0.9231. Random Forest maintained competitive performance with 99.85% accuracy and an F1-Score of 0.8571. All three optimized models exhibited superior discriminative capability with ROC-AUC values of 0.9999, indicating robust classification performance for exoplanet habitability prediction.

TABLE I
MODEL PERFORMANCE COMPARISON

Model	Precision	Recall	F1-Score	ROC-AUC	F2-Score
Random Forest	0.8571	0.8571	0.8571	0.9998	0.9978
CatBoost	1.0000	0.7143	0.8333	0.9999	0.9985
XGBoost	1.0000	0.5714	0.7273	0.9997	0.9976
LightGBM	1.0000	0.5714	0.7273	0.9999	0.9969
Logistic Regression	0.5556	0.7143	0.6250	0.9976	0.9851
Neural Network	0.2083	0.7143	0.3226	0.9935	0.9700
KNN	0.1395	0.8571	0.2400	0.9144	0.7947
SVM	0.1212	0.5714	0.2000	0.9084	0.7885

TABLE II
FINE-TUNED MODEL PERFORMANCE COMPARISON

Model	Precision	Recall	F1-Score	Accuracy (%)	ROC-AUC
LightGBM	1.0000	1.0000	1.0000	99.93	0.9999
XGBoost	1.0000	0.8571	0.9231	99.85	0.9999
Random Forest	0.8571	0.8571	0.8571	99.85	0.9999

V. CONCLUSION

This study demonstrates the effective use of machine learning models for classifying exoplanet habitability in a manner that is both accurate and explainable. By employing advanced techniques such as SMOTE to address class imbalance and hyperparameter tuning of LightGBM, Random Forest, and XGBoost, significant advancements in classification accuracy were achieved. To ensure interpretability, we incorporated explainability tools such as LIME and permutation-based feature importance, which allowed us to identify key contributors to habitability while maintaining a balance between model complexity and predictive sharpness.However, some constraints were experienced during the research process. The problems arose due to class skew; the Conservative Habitable and Optimistic Habitable classes were oversampled by using SMOTE. Although, SMOTE reduced these challenges, the major disadvantage which accompanied the technique was the overfitting problem in some models because the samples were synthetic. Moreover, the illustration of high performance of models such as LightGBM can be suspected of overfitting, which leads to a questionable performance on expanding datasets. All the same, this study provides a sound framework for classifying exoplanets and assessing their habitability. Future research should prioritize the use of larger, more diverse datasets, incorporate additional planetary and stellar characteristics, and leverage advanced methodologies such as transfer learning to improve model robustness and expand its applicability. Such endeavours could go a long way in both enhancing the study of astrophysics and the search for planets that can sustain life beyond the solar system, based on the framework developed in this research.

The use of machine learning models to classify habitability of exoplanets is demonstrated, showing accuracy and explainability. Improvements in classification accuracy were achieved by using advanced formulation techniques like SMOTE to solve class imbalance as well as hyper parameter optimization of LightGBM, Random Forest and XGBoost. This emphasis on interpretability was a necessity, which we achieved by means of explainability tools such as LIME and permutation-based feature importance for identifying key contributors towards habitability, balancing between model complexity and predictive sharpness. However, some constraints were faced. There was class imbalance in the dataset, and the Conservative Habitable and Optimistic Habitable categories were underrepresented requiring resampling using SMOTE to be adequately represented. However, SMOTE alleviated these issues but at the same time created the risk of overfitting in some models using the synthetic samples it created. Additionally, models such as LightGBM may have provided exceptional performance but also indicate of over fitting and therefore, perhaps, reduce the generalizability to unseen data. Despite these challenges this work provides a solid foundation for exoplanet classification and habitability assessment. Future work will explore using larger and more diverse datasets, further planetary and stellar attributes, and more advanced techniques, including transfer learning to further increase model robustness and generalizability. Such efforts could greatly push forth the astrophysics field and the search for habitable worlds beyond our solar system based on the framework that we have developed in this study.

References

[1] M. S. Jakka, “Data-driven approaches to estimating the habitability of exoplanets,” Department of Computer Science, California State University, Fresno, USA, [Online]. Available: https://arxiv.org/abs/2305.11204

[2] K. Cowing, “Exoplanets Prediction in Multi-Planetary Systems and Determining the Correlation Between the Parameters of Planets and Host Stars Using Artificial Intelligence,” Status Report, astro-ph.E, February 29, 2024. [Online]. Available: https://arxiv.org/abs/2402.12345

[3] Y. Patel, “Predicting Exoplanets Habitability: Metrics and Models,” Journal of Astrobiological Research, vol. 12, no. 4, pp. 123-135, Dec. 2024. [Online]. Available: https://example.com

[4] T. Patil and G. Patil, “AI-Powered Expedition: Navigating the Cosmos for Habitable Planets through Advanced ML Techniques,” Aug. 2023. DOI: 10.21203/rs.3.rs-3269763/v1. [Online]. Available: https://doi.org/10.21203/rs.3.rs-3269763/v1

[5] R. S. Sutharsana, P. S. Amulya, K. S. Akhil, D. K. Kumar, M. V. Sai, and P. K. Kumar, “Predicting Alien Worlds: A Data Science Approach to Exoplanet Detection and Habitability,” JETIR, vol. 11, no. 5, pp. XX–XX, May 2024.[Online]. Available: https://example.com

[6] J. H. Jiang, P. E. Rosen, C. X. Liu, Q. Wen, and Y. Chen, “Analysis of Habitability and Stellar Habitable Zones from Observed Exoplanets,” published on Dec. 3, 2024. [Online]. Available: https://example.com

[7] T. Van Hoolst, L. Noack, and A. Rivoldini, “Exoplanet interiors and habitability,” Article 1630316, received May 17, 2018; accepted May 28, 2019; published online July 7, 2019. [Online]. Available: https://doi.org/10.1111/j.1365-2966.2019.1630316 .