Automated Exoplanet Detection and Analysis: Leveraging Machine Learning to unveil Astronomical Insights

Shaik Ayesha Mohsin (11601110) Areeba Hassan (11641013) Mohammed Faisal (11632846)

1. Introduction

In the vastness of the cosmos, a fundamental question persists: Are we alone? Humanity has tirelessly pursued the quest to unravel signs of life beyond our planet. This exploration comprises two vital tasks: identifying exoplanets, including those potentially habitable, and studying their atmospheric properties through spectroscopy to assess habitability. This project focuses on the initial phase — discovering exoplanets. Historically, this has been a laborious endeavor, resulting in the detection of only approximately 4000 exoplanets over two decades. However, leveraging machine learning can automate and expedite this process significantly.

Exoplanets, planets beyond our solar system, reveal themselves through a technique known as transit photometry. As an exoplanet orbits its host star, it intermittently obstructs a portion of the star's light, causing a momentary dip in light intensity. Analyzing the depth and duration of these light dips aids in identifying and confirming exoplanets, providing valuable insights into their presence and characteristics within distant star systems. NASA has diligently collected this data using the Kepler Space Telescope, making it publicly available. This dataset serves as the foundation for automating exoplanet detection through machine learning and deep learning techniques.

2. Motivation

The exploration for exoplanets is a crucial step toward comprehending the possibility of life existing beyond Earth. Researchers traditionally faced a slow exoplanet discovery process due to the manual analysis of vast and complex astronomical data. Analyzing light curves and detecting subtle patterns required meticulous human scrutiny. Machine learning and deep learning algorithms can expedite this process by autonomously sifting through immense datasets, swiftly identifying patterns and anomalies in the light curves, enabling rapid and accurate detection of exoplanetary transits, thus revolutionizing the pace of discovery. By automating and accelerating the detection of exoplanets through advanced technologies like machine learning, it is possible to greatly enhance the ability to identify celestial bodies that may harbor life. This project is fueled by the desire to expedite exoplanet discovery, unlocking a deeper comprehension of our universe and the tantalizing possibility of life existing beyond our home planet.

3. Goals

The project's anticipated goals are detailed below.

- Automated Exoplanet Detection

Develop and train a robust machine learning model(s) capable of automating the detection of exoplanets using the provided light curve data.

- Improved Efficiency
  Increase the efficiency and speed of exoplanet detection compared to traditional manual methods, aiming for a significant reduction in detection time.
- High Accuracy
  Achieve high accuracy in distinguishing exoplanet-stars from non-exoplanet-stars, ensuring reliable identification of potential exoplanetary candidates.
- Exploratory Insights
  Gain exploratory insights into the features and patterns within the light curves that indicate the presence of exoplanets.

4. Literature Survey

The application of machine learning (ML) in exoplanet detection has significantly advanced our ability to explore the universe. In this literature overview, we delve into two research papers that not only exemplify the innovative use of ML but also introduce unique aspects to the field of exoplanet detection. The first paper, "Novel Artificial Intelligence (AI) Technique for Exoplanet Detection" (ScienceDirect, 2021), introduces an AI/ML system developed by ThetaRay, Inc., which marries various ML algorithms to analyze NASA's Transiting Exoplanets Survey Satellite (TESS) dataset, automating the classification of Threshold Crossing Events (TCEs) and uncovering new exoplanetary candidates.

The second paper, "A Machine Learning-Based Approach for Exoplanet Detection" (arXiv, 2020), presents a pioneering technique that employs machine learning to enhance conventional algorithm-based methods for exoplanet detection. What sets it apart is the use of time-series analysis and ML tools to extract and analyze 789 features from light curves, providing a unique perspective. This approach outperforms traditional methods and competes with deep learning models while offering computational efficiency and not requiring folded or secondary views of light curves.

5. Significance

Our project introduces a pioneering approach in the realm of exoplanet detection by seamlessly integrating cutting-edge advancements in machine learning while embracing the paradigm of Explainable AI (XAI). By implementing techniques that shed light on the model's decision-making process, we ensure transparency and interpretability, setting our work apart from existing studies. Furthermore, we emphasize the incorporation of hyper-parameter optimization using methodologies like HyperOpt, foreseeing a future where optimizing neural network architectures becomes standard practice. This proactive approach positions our project at the forefront of machine learning algorithms, ensuring adaptability to forthcoming advancements. Moreover, we are committed to validating our model's performance using attention maps and class activation maps, thus ensuring that our algorithm avoids overfitting and makes informed decisions based on relevant features from the input space. By combining these unique elements, our project stands as a beacon of innovation, promising to significantly advance the accuracy, interpretability, and efficiency of exoplanet detection methodologies.

6.Objectives

To facilitate clarity, a comprehensive list of project objectives is presented below.

- Data Preprocessing
  a. Handle missing values and potential outliers in the dataset.
  b. Scale and normalize the values to ensure uniformity and aid in model training.
- Model Development and Training
  a. Implementing and training a variety of chosen machine learning and deep learning models like SVM, CNN, LSTM, among others.
  b. Train these models on the provided training dataset, optimizing hyperparameters to enhance performance.
- Evaluation and Model Selection
  a. Evaluate models using appropriate metrics such as accuracy, precision, recall, F1-score, and AUC-ROC curve.
  b. Select the best-performing model based on evaluation results for further refinement.
- Fine-Tuning and Optimization
  a. Optimize the selected model further for improved accuracy and efficiency.
  b. Experiment with hyperparameters to achieve the best possible outcomes.
- Testing and Validation:
  a. Apply the optimized model to the provided test dataset for final validation.
  b. Assess the model's generalizability and performance on unseen data.
- Visual Representation:
  a. Develop visualizations to present results, including confusion matrices, ROC curves, and feature importance plots.
  b. Create informative and intuitive visuals to aid in the understanding and interpretation of the model's performance.

7. Features

The different features or modules which will be implemented in the project are as follows:

1. Pre-processing:
   - Handling missing values in the dataset.
   - Scaling the features, potentially using techniques like Min-Max scaling or Standardization.
2. Model Training:
   - Implementing and training a variety of chosen machine learning and deep learning models like SVM, CNN, LSTM, among others.

- Hyperparameter tuning to optimize model performance.

3. Evaluation:
    - Assessing models using appropriate metrics like accuracy, precision, recall, F1-score, and AUC-ROC.
    - Comparing models and selecting the best-performing one.
    - Creation of visual representations illustrating model predictions on light curves, aiding in a clear understanding of the detection process.

8. Expected outcomes

The project aims to achieve the following results:

- A well-trained machine learning model that can accurately identify exoplanets based on the given light curve data.
- Insights into the effectiveness of different models and their potential applications to exoplanet detection.
- Enhanced Model Interpretability, allowing for meaningful insights into how the model operates and makes predictions.
- Insightful Visualization of Predictions.
- Potential new discoveries of exoplanetary candidates.

9. References

[1] Ofman, L., Averbuch, A., Shliselberg, A., Benaun, I., Segev, D., & Rissman, A. G. (2022, February 1). *Automated identification of transiting exoplanet candidates in NASA Transiting Exoplanets Survey Satellite (TESS) data with machine learning methods*. New Astronomy; Elsevier BV. https://doi.org/10.1016/j.newast.2021.101693

[2] Malik, A., Moster, B. P., & Obermeier, C. (2021, December 21). Exoplanet detection using machine learning. Monthly Notices of the Royal Astronomical Society; Oxford University Press. https://doi.org/10.1093/mnras/stab3692