ExoNet - An Explainable AI Pipeline for Accelerating Exoplanet Discovery

CHALLENGE NAME

A WORLD AWAY: HUNTING FOR EXOPLANETS WITH AI

TEAM NAME

SANAD

TEAM MEMBERS

Omar Elsayed Ibrahim Aly
Hazem Mohamed Abdelazim
Omar Mohamed Salama
Ali Tarek Fekry
Abdelrhman Mahmoud

1. Introduction: The Data Deluge

The search for worlds beyond our own has entered a golden age. Missions like Kepler, K2, and TESS have provided an unprecedented flood of data, revealing thousands of exoplanets. However, this success has created a new bottleneck: the sheer volume of data overwhelms traditional manual analysis. Sifting through noisy light curves to distinguish a genuine planetary transit from stellar activity or instrumental noise is a slow, labor-intensive process. This manual effort limits the speed of discovery and risks leaving faint, promising signals buried in the archives.

2. Our Solution: ExoNet

ExoNet is a comprehensive, open-source solution designed to automate, accelerate, and democratize the exoplanet discovery process. It is an end-to-end pipeline featuring a sophisticated AI model and an interactive web interface that allows scientists and citizen-scientists alike to analyze transit survey data with confidence.

Our core features include:

- Multi-Modal AI Model: ExoNet doesn't just look at summary statistics. It uniquely combines tabular data (e.g., orbital period, transit duration, stellar radius) with a deep analysis of the time-series light curve itself, allowing it to learn the subtle morphological signatures of a true planetary transit.

- Interactive Web Interface: Built with Streamlit, our user-friendly dashboard allows users to upload new light curve data, receive an instant classification, and visualize the results in a clear, intuitive manner.

- Explainable & Transparent Predictions: We believe AI in science should not be a "black box." ExoNet uses SHAP (SHapley Additive exPlanations) to show users why the model made a specific decision, highlighting which features (from both the light curve shape and the tabular data) contributed to the classification.

Reproducible & Open: The entire project—from data preprocessing notebooks to model training scripts and the web application—will be open-sourced to ensure transparency and encourage community collaboration.

3. Technical Approach

Our pipeline is built on a robust, multi-modal machine learning architecture.

- **Data Sources:** We train ExoNet on the complete public data catalogs from the Kepler, K2, and TESS missions, utilizing their classifications of Confirmed Planets, Candidates, and False Positives.
- **Feature Engineering:**
  - **Time-Series Features:** For each potential signal, we process the time-series "light curve window." This data is fed into a lightweight **1D Convolutional Neural Network (1D-CNN)**. This network is specifically designed to learn the characteristic shape of a transit—the "U" or "V" shape dip—while being resilient to noise. It outputs a learned feature embedding that represents the signal's morphology.
  - **Tabular Features:** We concurrently process key tabular features such as orbital period, transit depth, duration, signal-to-noise ratio (SNR), and stellar parameters. These are normalized and prepared for the model.
- **Hybrid Model Architecture:** The time-series embedding from the CNN and the processed tabular features are concatenated and fed into a final dense neural network. This hybrid approach allows ExoNet to make a holistic judgment based on both the *shape* of the signal and its physical *context*.
- **Addressing Class Imbalance:** Astronomical datasets are notoriously imbalanced. To ensure our model doesn't simply learn to predict "false positive" every time, we employ two key strategies: **stratified sampling** during training to ensure each batch sees a representative sample of all classes, and the use of a **focal loss function**, which focuses the model's learning on harder-to-classify examples.
- **Output:** The model outputs a calibrated probability score for each of the three classes: **Confirmed Planet**, **Candidate**, and **False Positive**.

4. User Experience & Impact

The ExoNet web application is designed for accessibility and insight. A user can:

1. **Upload Data:** Directly upload a light curve file or manually input key parameters for a signal of interest.
2. **Receive Instant Analysis:** The app runs real-time inference using the pre-trained ExoNet model.
3. **Visualize Results:** The interface displays the folded light curve, clearly showing the potential transit event.

4. **Understand the "Why":** A SHAP force plot visualizes the factors that pushed the prediction toward its final outcome. For example, it might show that a V-shaped transit (characteristic of an eclipsing binary) and a long duration were key drivers of a "False Positive" classification.
5. **Export Reports:** Users can download a summary report for promising candidates, streamlining the process for follow-up observation proposals.

By providing this tool, we empower researchers to rapidly screen vast datasets, validate potential candidates, and ultimately accelerate the pace of exoplanet discovery.