

Physics-Informed Machine Learning for Anomaly Detection and Candidate Transit Identification in TESS Light Curves

Rahul Thakur (BITS Pilani)

Abstract

The Transiting Exoplanet Survey Satellite (TESS) has generated high-cadence photometric observations for millions of stars, creating an increasingly large search space for identifying transit-like signals and unusual stellar variability. While conventional transit-search algorithms such as the Box Least Squares method remain highly effective, automated machine-learning approaches can complement these techniques by identifying patterns that are difficult to specify explicitly and by prioritising unusual light curves for further investigation.

This work investigates a physics-informed machine-learning framework for detecting and characterising anomalous transit-like signatures in publicly available TESS light curves. The pipeline first performs photometric preprocessing, including quality filtering, normalisation, detrending, and segmentation. Candidate periodic signals are identified using conventional period-search techniques, while statistical and astrophysically motivated features describing transit depth, duration, periodicity, local variability, and signal morphology are extracted from each light curve.

These representations are subsequently analysed using unsupervised and semi-supervised anomaly-detection methods to identify observations that deviate significantly from dominant stellar variability patterns. Known exoplanet systems and labelled false-positive events are used as reference populations for evaluating whether the anomaly score provides additional discriminatory information beyond conventional detection statistics.

Particular emphasis is placed on interpretability rather than treating the machine-learning model as an independent discovery mechanism. The objective is to determine which photometric and temporal properties cause individual systems to be identified as anomalous and whether these properties correspond to physically meaningful transit characteristics.

The study demonstrates a computational framework in which machine learning acts as a candidate-prioritisation layer alongside established astronomical techniques, potentially reducing the number of light curves requiring manual inspection while retaining physically interpretable criteria for subsequent validation.