Photometric Signatures of Exoplanet Host Stars Are Detectable Without Transit Signals: Evidence From Self-Supervised Light Curve Embeddings

Manna, Adnan (School: International Independent Schools 2)

Current exoplanet detection relies on identifying periodic transit dips in stellar light curves, explicitly searching for the geometric signature of a planet crossing its host star. This project investigates whether exoplanet-hosting stars exhibit persistent photometric behavior detectable without any transit search. We hypothesize that self-supervised representation learning encodes exoplanet-related information in its embeddings. In Phase I, a custom transformer (AstraScan) trained on 1,000 Kepler exoplanet host light curves via next-flux prediction shows statistical separation between hosts and non-hosts. In Phase II, ASTROMER, a foundation model pretrained on 1.5 million MACHO survey light curves through masked reconstruction, serves as a frozen feature extractor on 1,600 Kepler PDCSAP flux curves, with only a logistic regression trained on the resulting 256-dimensional embeddings. Despite severe domain mismatch — ASTROMER was pretrained on ground-based MACHO R-band data, never saw Kepler data, and was never trained for exoplanet detection — frozen embeddings achieve cross-validated ROC-AUC of 0.615 ± 0.024, confirmed by a 1,000-iteration permutation test ($p < 0.001$). Exoplanet hosts cluster more tightly in embedding space (intra-class distance 1.29 vs 8.05 for non-hosts, centroid separation 3.52). ASTROMER outperforms a 12-feature statistical baseline by +0.046 AUC, and the signal survives confound regression for magnitude and noise (residual AUC 0.582). These results demonstrate that stellar light curves contain latent structure correlated with exoplanet presence, recoverable without transit detection, phase folding, or feature engineering. This opens a path toward detecting planets in the 99.5% of systems where orbital geometry prevents transit observation.

Awards Won: