**CLARA: A Scalable Unsupervised ML Framework for Anomaly Discovery and Morphological Prioritization in TESS**

*Mainak Dasgupta*

*IIT Madras*

We present CLARA (Controllable Learning for Anomaly Recognition in Astrophysics), a modular unsupervised framework for transit detection in TESS light curves. CLARA leverages Unsupervised Random Forests (URFs) trained on systematically designed synthetic datasets, enabling controllable anomaly discovery tuned toward transit-like morphologies.

Our study addresses two central questions: (a) how synthetic training set design influences the generalization of URF models across independent TESS sectors, and (b) whether anomaly scores derived from these models correlate with genuine astrophysical phenomena. We introduce three URF variants optimized under recall, precision, and balanced scoring objectives, and validate their behavior across 384,000 light curves spanning five SPOC sectors. The balanced configuration demonstrates stable cross-sector generalization, while low- and high-$\alpha$ models systematically trade completeness against astrophysical prioritization.

Astrophysical validation incorporates Gaia DR3 parameters and SIMBAD object label based morphological similarity matching, establishing that high-scoring anomalies align with physically meaningful classes. Morphological feature based clustering yields a 14.0\% recovery rate of confirmed transits (16 among 114 candidates), representing a 30× enrichment compared to baseline SPOC yields.

CLARA demonstrates how unsupervised frameworks can be systematically steered and physically contextualized, offering a scalable approach for curve prioritization for exoplanet detection in TESS and future large surveys such as Rubin/LSST.