EXOPLANET CANDIDATE DETECTION THROUGH TESS LIGHT CURVE ANALYSIS

Srishti Hazra
11th Grade at Edison Academy Magnet School

INTRODUCTION

- Transit photometry is responsible for ~75% of confirmed exoplanets as of October 2023 (NASA Exoplanet Archive).

- Missions like Kepler and TESS produce high-cadence photometric data at massive scale
  - Helps enable detection
  - However, requires overwhelming manual classification methods

CHALLENGES

- Light curves often contaminated by astrophysical false positives (eclipsing binaries, blended background stars)
- Class imbalance and random noise
- Manual vetting is infeasible at scale

Current architecture

- Astronet-Triage-v2, developed in 2023 by Tey et al.
  - CNN trained on Kepler light curves
  - One of the first deep learning models to classify planet candidates
  - Architecture has remained largely unchanged

Tey, Evan, et al. "Identifying Exoplanets with Deep Learning. V. Improved Light Curve Classification for TESS Full Frame Image Observations." ArXiv (Cornell University), 3 Jan. 2023, https://doi.org/10.48550/arxiv.2301.01371.

MACHINE LEARNING

ML for exoplanet detection remains underdeveloped especially in exploiting temporal and spatial features

CNNs for temporal analysis?

- CNNs applied to temporal sequences can significantly outperform classical models, leveraging both spatial and temporal information jointly (Zhang et al., 2022).
- Highly parallelizable and computationally efficient, making them practical for large-scale datasets (like TESS)

Huapeng Li, Yajun Tian, Ce Zhang, Shuqing Zhang, Peter M. Atkinson, Temporal Sequence Object-based CNN (TS-OCNN) for crop classification from fine resolution remote sensing image time-series, The Crop Journal, Volume 10, Issue 5, 2022, Pages 1507-1516, ISSN 2214-5141, https://doi.org/10.1016/j.cj.2022.07.005.

PROPOSED ARCHITECTURE

Model: "sequential"

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 198, 32)	128
max_pooling1d (MaxPooling1D)	(None, 99, 32)	0
dropout (Dropout)	(None, 99, 32)	0
conv1d_1 (Conv1D)	(None, 97, 64)	6,208
max_pooling1d_1 (MaxPooling1D)	(None, 48, 64)	0
dropout_1 (Dropout)	(None, 48, 64)	0
lstm (LSTM)	(None, 128)	98,816
dense (Dense)	(None, 64)	8,256
dropout_2 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 1)	65

Total params: 113,473 (443.25 KB)

Trainable params: 113,473 (443.25 KB)

Non-trainable params: 0 (0.00 B)

PROPOSED ARCHITECTURE

- CNN detects localized spatial transit features (e.g., depth, symmetry)
- LSTM distinguishes periodicity (planetary transits) from irregular patterns (stellar activity)
- MaxPooling, Dense, Dropout to ensure robustness of model
- Architecture captures both what a signal looks like and how it changes

DATA PIPELINE

- Obtained light curve data from NASA's Exoplanet Archive
  - Confirmed exoplanets discovered via transits
  - Kepler certified false positive table

- Preprocessing using gaussian process regression & median filtering
- Labeled flux drops ≥10% of segment mean as transit (1), otherwise non-transit (0).

- X: (12,195, 200, 1) › (samples, time steps, features)
- Y: (12.195) › Binary labels

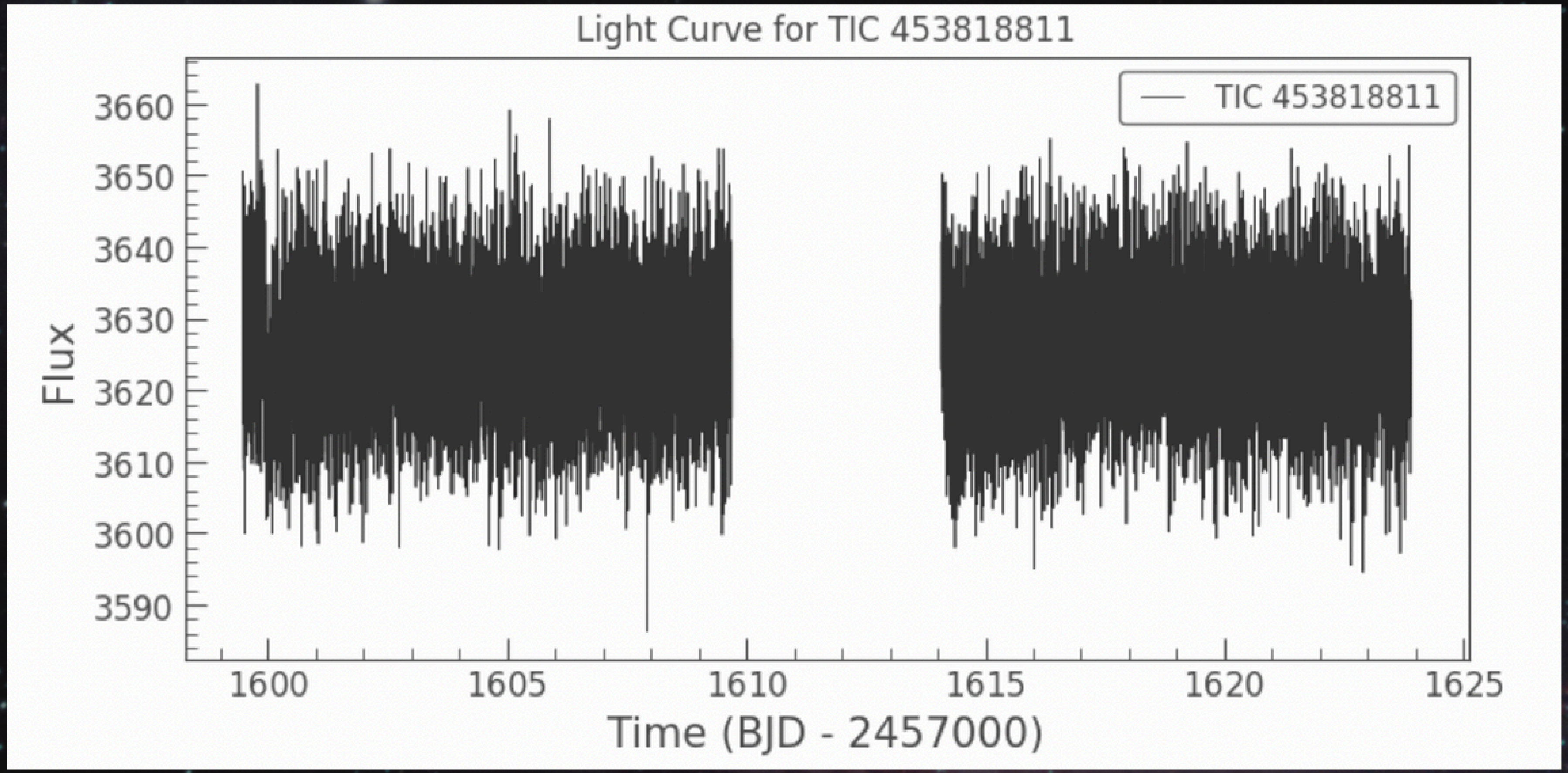
**very imbalanced dataset; focal loss was used to mitigate this as much as possible

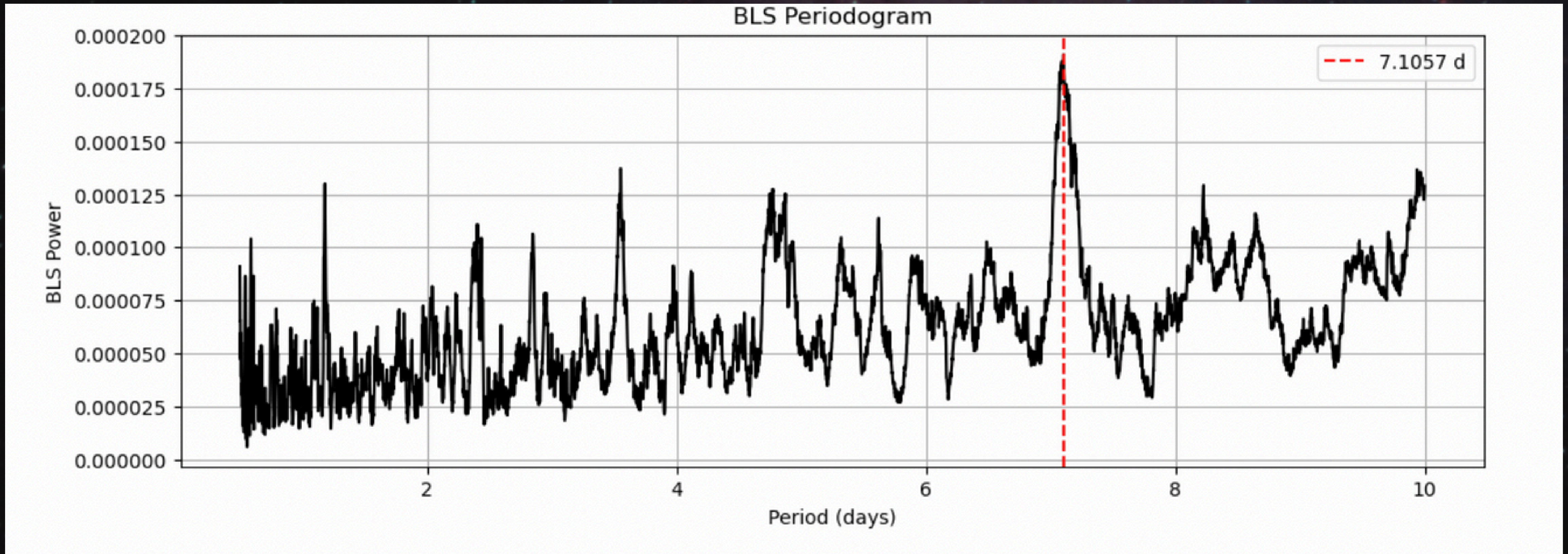
DATA PIPELINE

- Model trained on over 38k data points from NASA Exoplanet Archive
  - Adequate loss / accuracy / mAP
- 12195 segmented TESS light curves then evaluated
  - Model's output is a softmax activation function (percentage value)
  - Light curves with an output of >0.9 were flagged for further evaluation

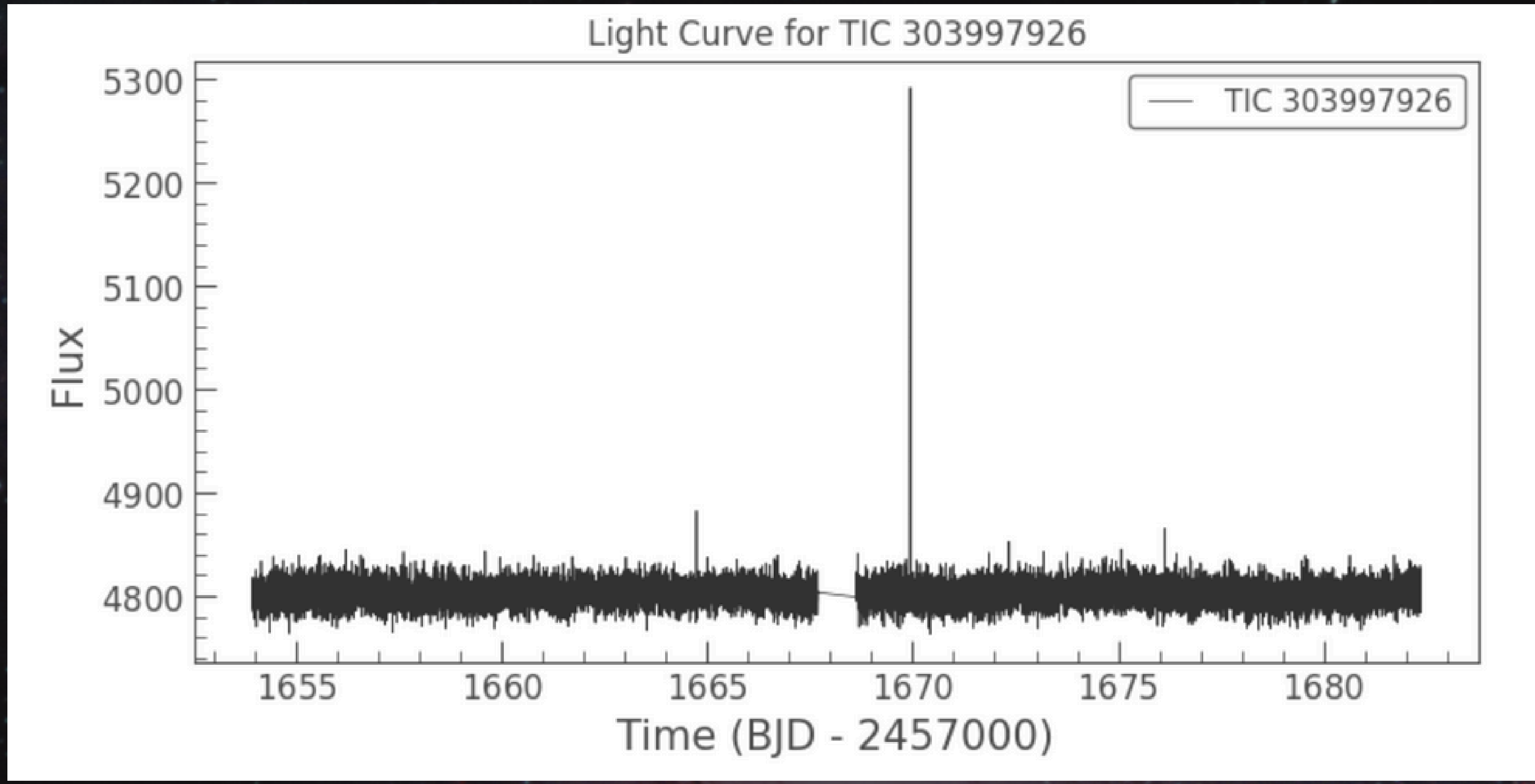
- These were cross referenced with Eclipsing Binary Catalogs (NASA & Gaia DR3). Any V-shaped dips were removed (more indicative of binary stars).

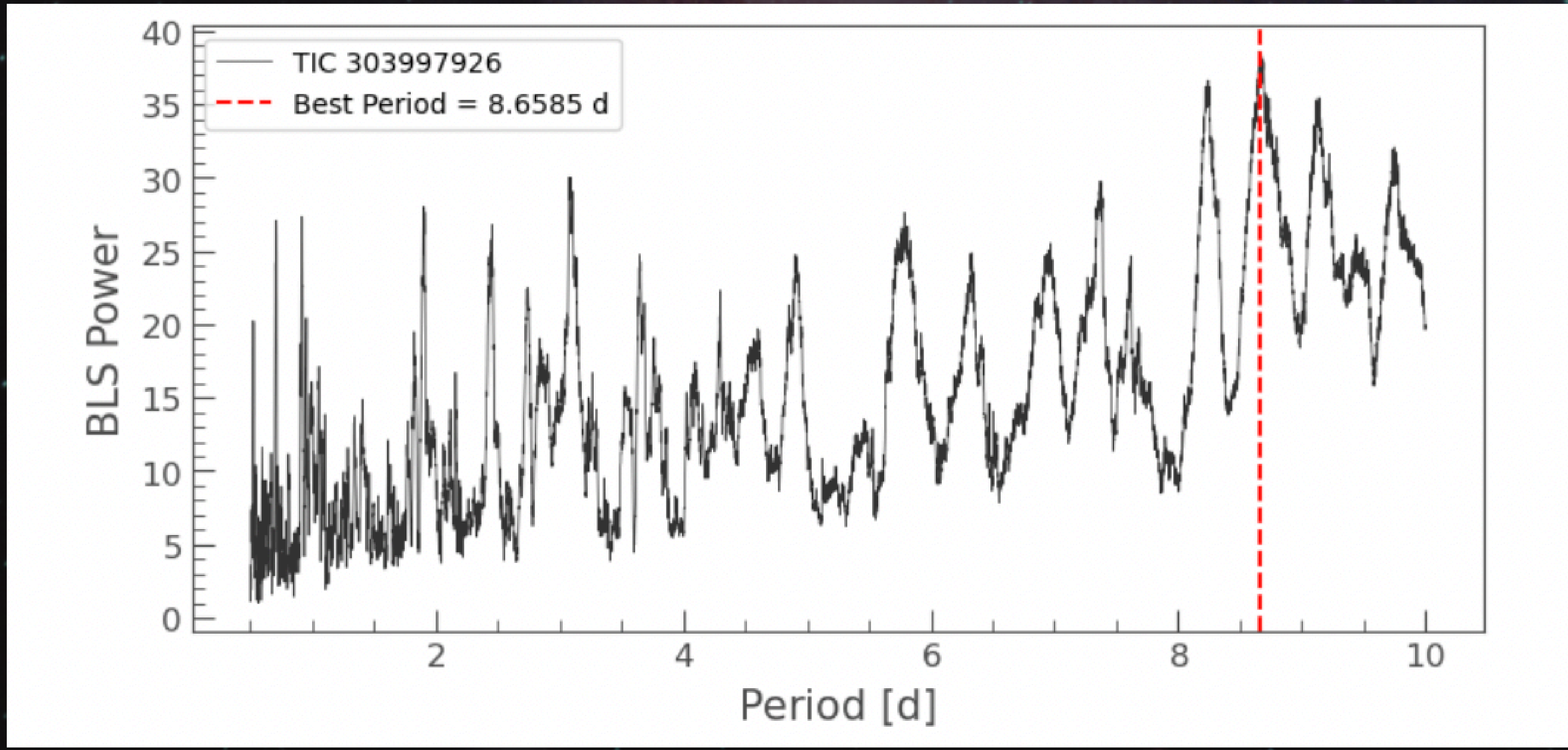
EXAMPLE


Light Curve for TIC 453818811
TIC 453818811
Flux
3660
3650
3640
3630
3620
3610
3600
3590
1600
1605
1610
1615
1620
1625
Time (BJD - 2457000)


BLS Periodogram
7.1057 d
BLS Power
0.000200
0.000175
0.000150
0.000125
0.000100
0.000075
0.000050
0.000025
0.000000
2
4
6
8
10
Period (days)

EXAMPLE


Light Curve for TIC 303997926
TIC 303997926
Flux
5300
5200
5100
5000
4900
4800
1655
1660
1665
1670
1675
1680
Time (BJD - 2457000)


TIC 303997926
Best Period = 8.6585 d
BLS Power
40
35
30
25
20
15
10
5
0
2
4
6
8
10
Period [d]

FINAL RESULT

256 likely candidates of the >12k light curves evaluated

32 flagged as extremely likely candidates

- ~2.1% candidate rate across TESS segments
- All candidates cross-checked with binary catalogs and transit shape metrics
- Suggests improved precision compared to prior pipelines
  - Area under precision-recall curve ≈ 98% as compared to Astronet-Triage-v2's 96.5%
- Further astrophysical vetting ongoing

WHAT THIS ADDS

- Advances on Astronet-Triage-v2 but incorporates temporal modeling with LSTMs for sequential pattern recognition
- Explicit handling of class imbalance using focal loss
- End-to-end hybrid architecture (CNN + LSTM) tailored specifically for transit photometry

FAILURE MODES

- Missed shallow transits due to normalization issues
- Some flagged false positives passed filters due to noisy periodicity
- May still underperform on low SNR or single-transit events

FUTURE DIRECTIONS

- Explore regularization techniques to reduce overfitting when training
- Experiment with sampling techniques or cost-sensitive learning to improve detection with imbalanced data
- More complex neural architectures (which require more resources to run) might work better for temporal analysis

- Integrate spectroscopic, stellar activity, and high-res imagery - multi modal capabilities to improve accuracy

MORE DETAILS + QUESTIONS

iPoster

Please contact me with suggestions/feedback if you have any!

Srishti Hazra
11th grade @ Edison Academy Magnet School
hazra.srishti@gmail.com