Finding Exoplanet Transits with Machine Learning

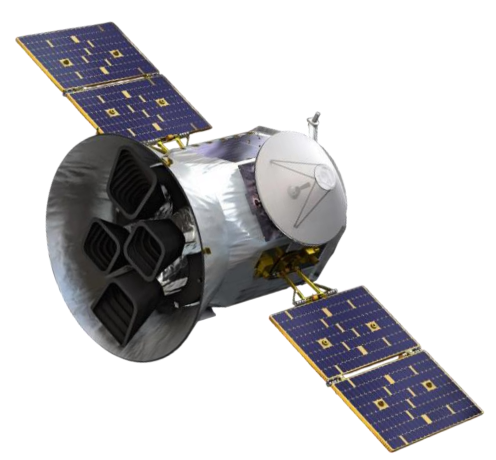
Allison Shaver[1], Brett Skinner[1], Anna Goodson[2], Diana Dragomir[1]

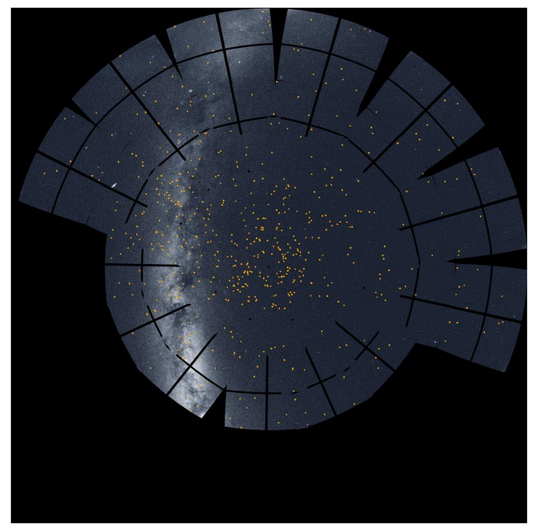
[1]University of New Mexico Department of Physics & Astronomy

[2]Texas A&M University Department of Physics & Astronomy

Transiting Exoplanet Survey Satellite

~ TESS surveys a hemisphere every 13 months
~ Scans the sky in chunks called "sectors"
~ Each sector lasts for one month, aligned with the lunar cycle
~ Light curves contains data gaps due to Sun/Moon brightness and data download

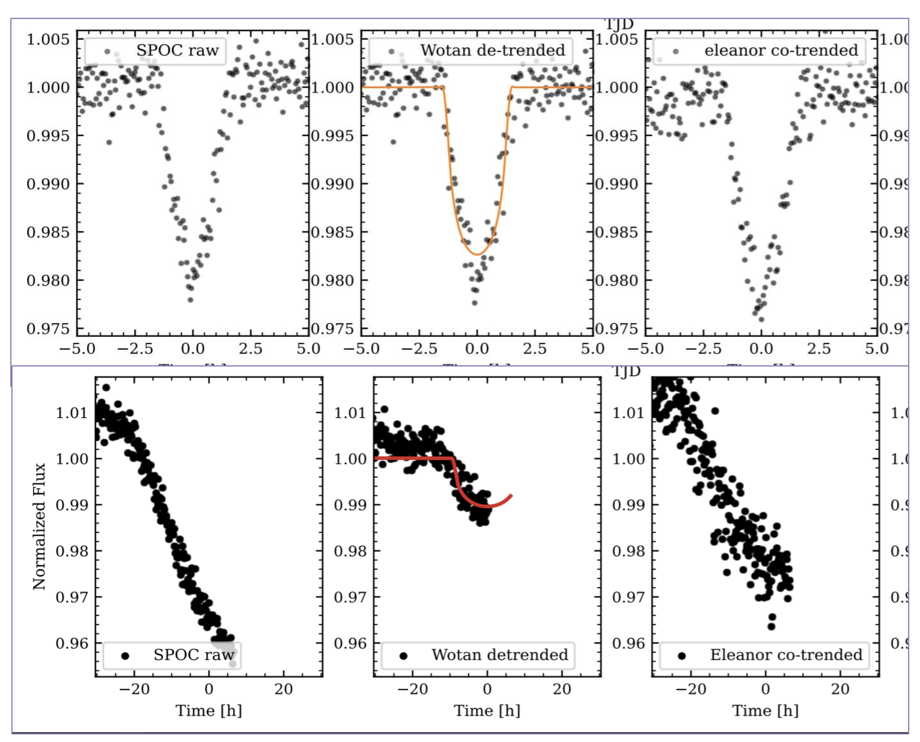




Deep-Transit

~ Machine Learning algorithm that utilized YOLOv3 for object detection
 ~ Originally trained on Kepler space telescope data, which observed continuously (no data gaps)
~ *DT* commonly flags gaps as transits in TESS data
 ~ Nearly 50% of the *DT* flagged objects are gaps!
 ~ Time consuming to spot actual transits out of "junk"

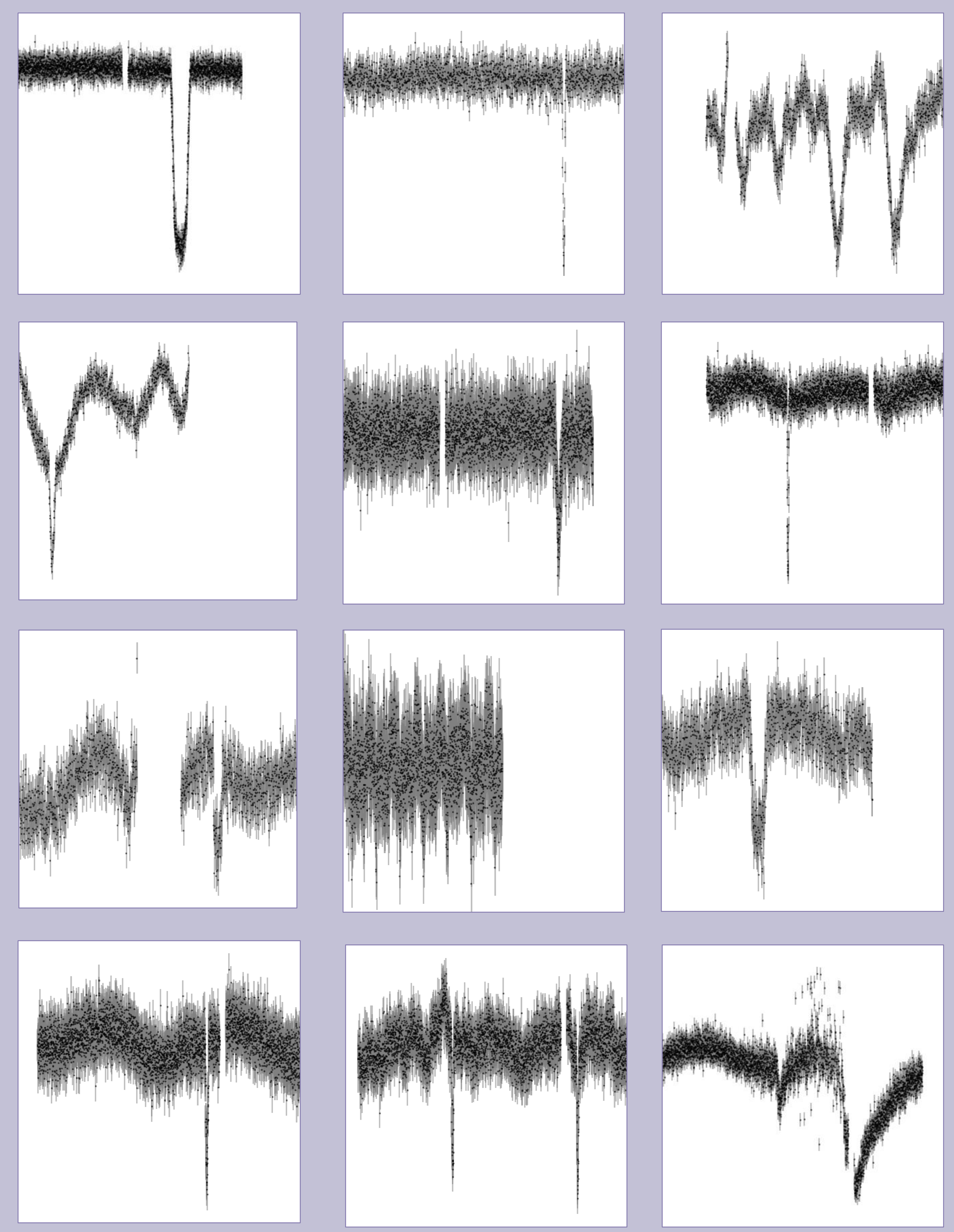
TESS observes ~100,000 light curves → DT flags ~ 1000 targets with "transits" → Citizen scientists manually identify ~ 250 transits → UNM team retains a couple dozen events



DT must be retrained on TESS data!!

Training Process

~ Deep Transit is trained on image files coupled with label files
 ~ Each image is 416 x 416 pixels
 ~ Label file consists of [x and y centers, width, height, signal-noise ratio]
~ Limitations
 ~ TESS images with low signal-noise ratio limits learning capabilities
~ DT model trained on 140,000 Kepler light curves
 ~ We have access to only 39 TESS sectors
~ Choices made to optimize training
 ~ Decreased to 16 sectors, which have been vetted, and extracted 2,691 unique TICs
 ~ Expanded the dataset from 2,691 to 10,756 through multiple image generation and image mirroring transfer learning



↑ Which identified "transits" are junk? ↑

Where We Are

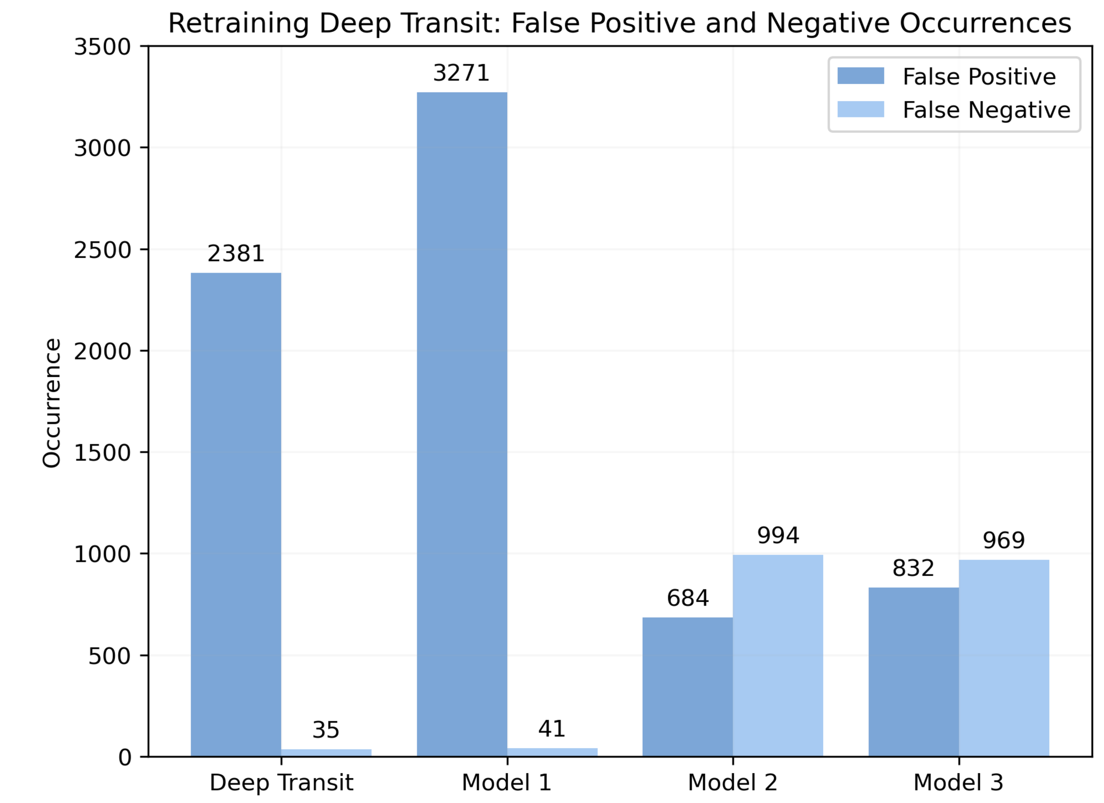
~ In Summer 2025, 10 models were trained
 ~ 15 sectors, 2,730 light curves
 ~ Adjusting batch and epoch sizes
 ~ Had 2 rounds of vetting to refine data

Model	Test Results (Number of Events)			
	0	1-6	7+	Max
DT	223	2374	134	32
1	75	3	2653	6648
2	56	12	2663	925
3	3	0	2728	9144
4	2731	0	0	0
5	2652	17	62	152
6	2352	361	18	25
7	206	10	2515	7110
8	0	0	2731	5137
9	2599	122	10	10
10	2731	0	0	0

~ Analysis led to the conclusion of a small sample size causing model overfitting, so the project continued into Fall 2025 with a new sector
~ To expand the dataset, 4 images were generated from each transit

Model	Precision	Recall
DT	0.607	0.978
1	0.566	0.967
2	0.568	0.976
3	0.570	0.999
4	-	0.000
5	0.650	0.033
6	0.605	0.148
7	0.550	0.893
8	0.570	1.000
9	0.564	0.048
10	-	0.000

~ Two images generated with different x-axis values
~ Both both images had a mirrored pair
~ Ran 3 models in increased training set



~ These models points to a decline in FPs compared to the *DT*, but at the cost of an increased in FNs
~ Running 3-6 more iterations with this larger training set will inform us of the optimal training parameters

Next Steps

~Run more models with current dataset and adjusting parameters
 ~ Keep batch size 32, 64, or 96
 ~ Increase epoch values (50-200)
~ Ways to increase machine learning training set:
 ~ Including newer sectors
 ~ Insert synthetic transits into "quiet" light curves
 ~ Randomized split and swap of transit events
 ~ Increase data set by order of magnitude (~100k)
~ Adjust dataset split for training method
 ~ Current split is 77/14/09 (percentage) for train, validate, and test
 ~ Aim for closer split of 85/10/05 as dataset increases

~ Use Center for Advanced Research Computing (CARC) for new training models, ideally allowing for quicker code running and larger epoch sizes.
 ~ May be able to parallelize code to optimize time, or train multiple unique models simultaneously
 ~ Optimal for personal storage on local devices

References & Acknowledgements

Kaiming Cui et. all, Identify Light-Curve Signals with Deep Learning Based Object Detection Algorithm. I. Transit Detection (2021)

TESS images courtesy of NASA

Special thanks to Brett Skinner for general assistance and mentorship, to Anna Goodson for helping kickstart the project, and to Dr. Diana Dragomir for mentorship and guidance.

This poster presentation was made possible through the generous UNM Physics and Astronomy Department through the Rayburn Scholarship.

Contact Me! →