



# Improving Exoplanet Detection with Machine Learning

Allison Shaver  
Physics Day  
April 25<sup>th</sup>, 2026

# Outline

**01**

**Objectives**

**02**

**Methodology**

**03**

**Training**

**04**

**Analysis**

**05**

**Conclusions**

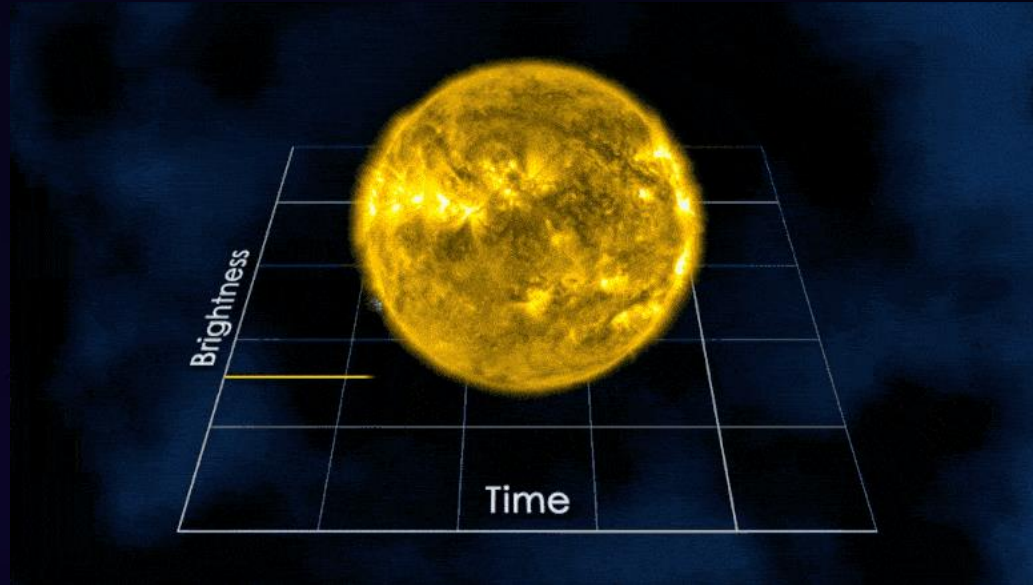
# 01 Objectives

Scientific Background, Terminology,  
Motivations, and Goals



# Exoplanets

What are they?  
How to find them?



Animation: From NASA



# Terminology

○  
**Kepler**



●  
**TESS**

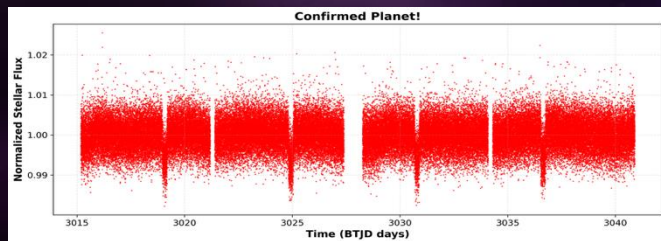
○  
**Long Period  
Exoplanets**

●  
**Machine  
Learning  
Models**

○  
***Deep Transit*  
(DT)**



# Motivation



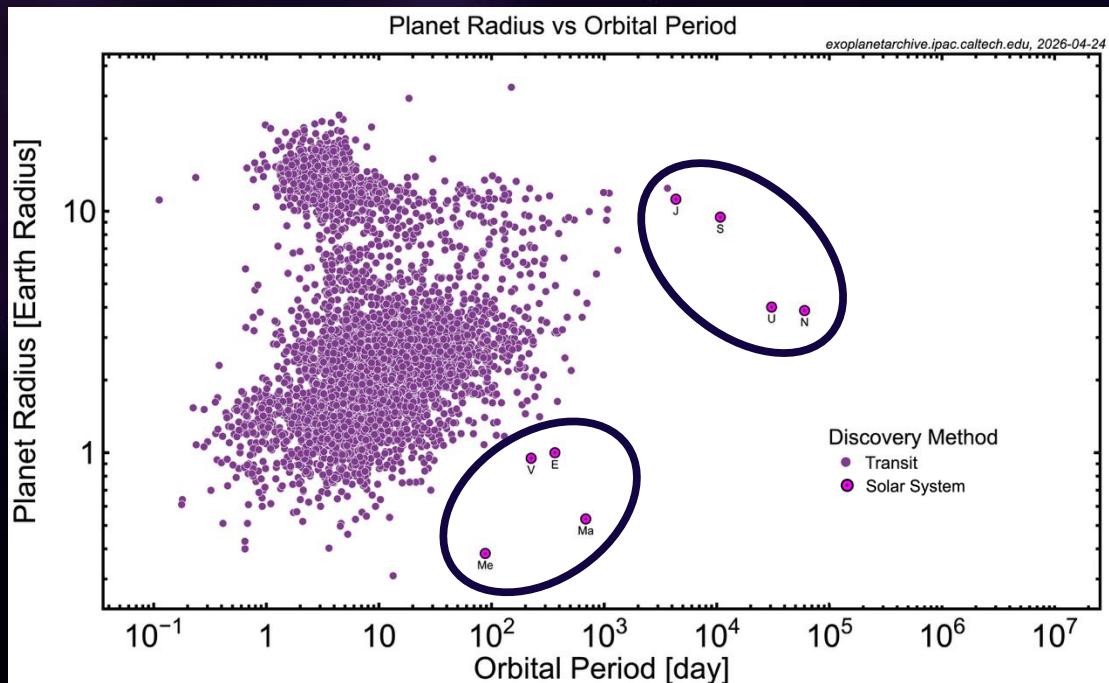
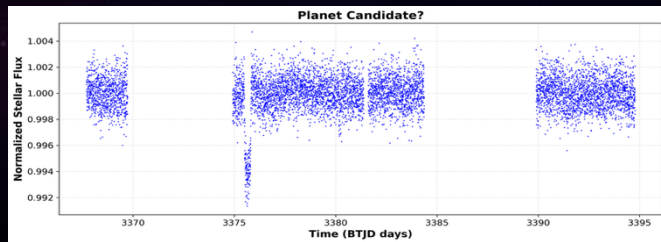
## Parameters

Period: 5.86 days

Transit Depth: ~ 1.5%

Stellar Radius:  $1.55 (R_{\text{sun}})$

Planet Radius:  $1.255 (R_{\text{Jupiter}})$



# 02

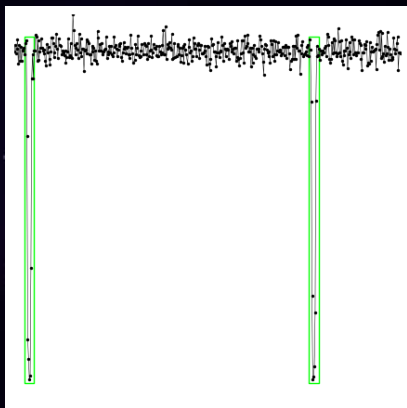
## Methodology

Functions of Machine Learning Models for Transits

# Deep Transit

## Origins

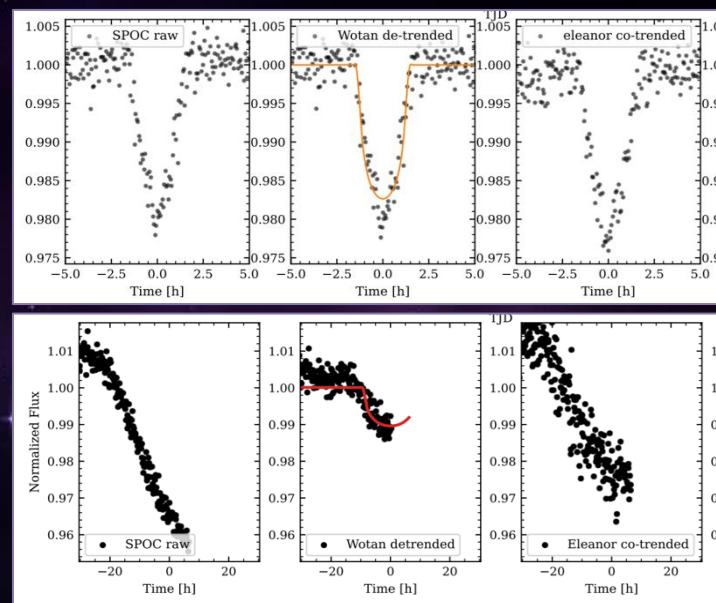
140,000 images, trained originally with Kepler, plus TESS additional data



Pictures from Cui et al. 2021

50% of the *DT* detected items are gaps in data, due to TESS data formation methods

## Limitations

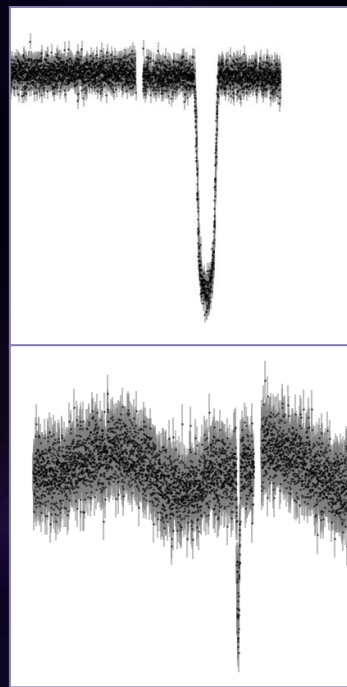
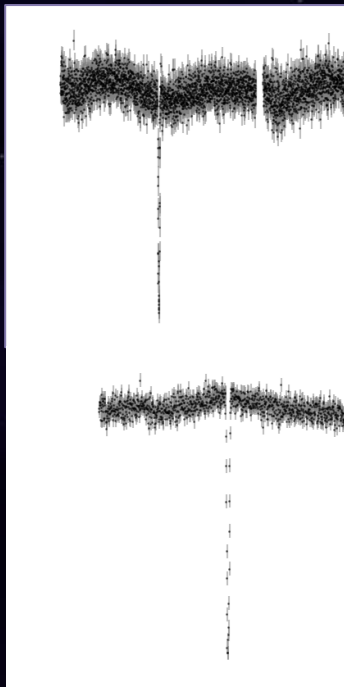




# Light Curves

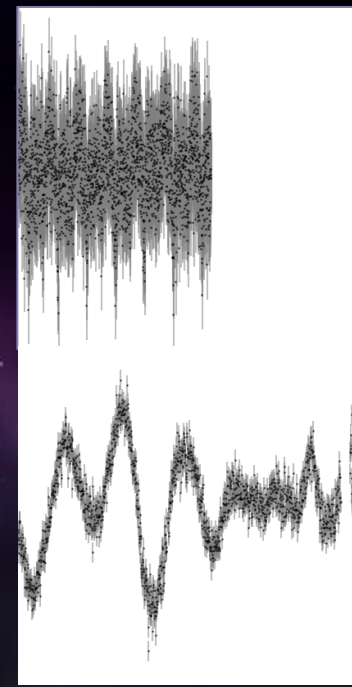
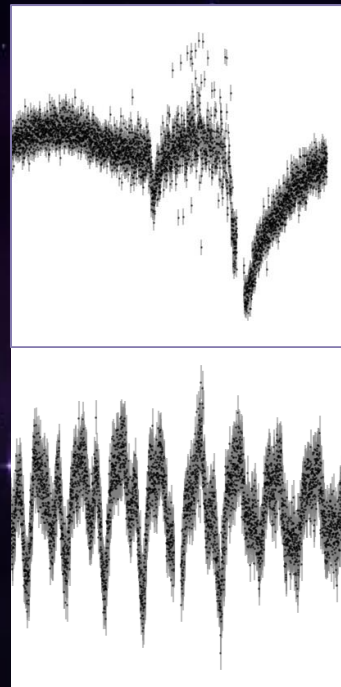
## Aim

Find single dip in light curves over the course of longer data strings



## Avoid

Avoid repeated transits, stellar variance, instrumentation errors



The background is a dark blue space scene with several bright stars and two large, glowing blue planets. The number '03' is centered in a white circle.

**03**

# Training

Training preparation, terminology,  
and parameters

# Preparation

Gather transits that have already been detected with *DT* and passed human inspection

Create images to *DT*'s image specifications and associated label for image

Generate 2 images for each transit w/ new placement and then mirror each.  
4 images per transit

Take generated data and split into train, validate, and test sets (77 %, 14%, and 9%)

Run multiple model training using transfer learning from *DT*'s base model



# Terminology Pt. 2

○  
**False Positives  
& Negatives**

●  
**Precision**

○  
**Recall**

●  
**Epoch**

○  
**Batch**



# Parameters

## Summer 2025 Models



**Images**

2730



**Targets**

2730

## Fall '25 & Spring '26 Models



**Images**

10756



**Targets**

2690

## Training Stats



**Batch**

32, 64, 96



**Epochs**

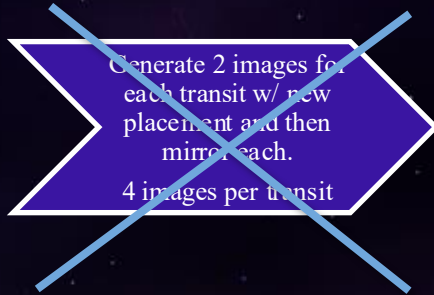
20-50 currently,  
increase to 100-300

The background of the slide is a deep blue space scene. It features two large, glowing blue spheres that resemble planets or moons, positioned on the right side. The background is filled with numerous small, bright stars of various colors, including blue, orange, and white. A thin, white circular outline is centered on the left side of the slide, framing the text.

# 04 Analysis

Current Progress and  
Conditions

# Summer Models



No mirroring images, first 10 attempts, smaller dataset (2730 targets & images)

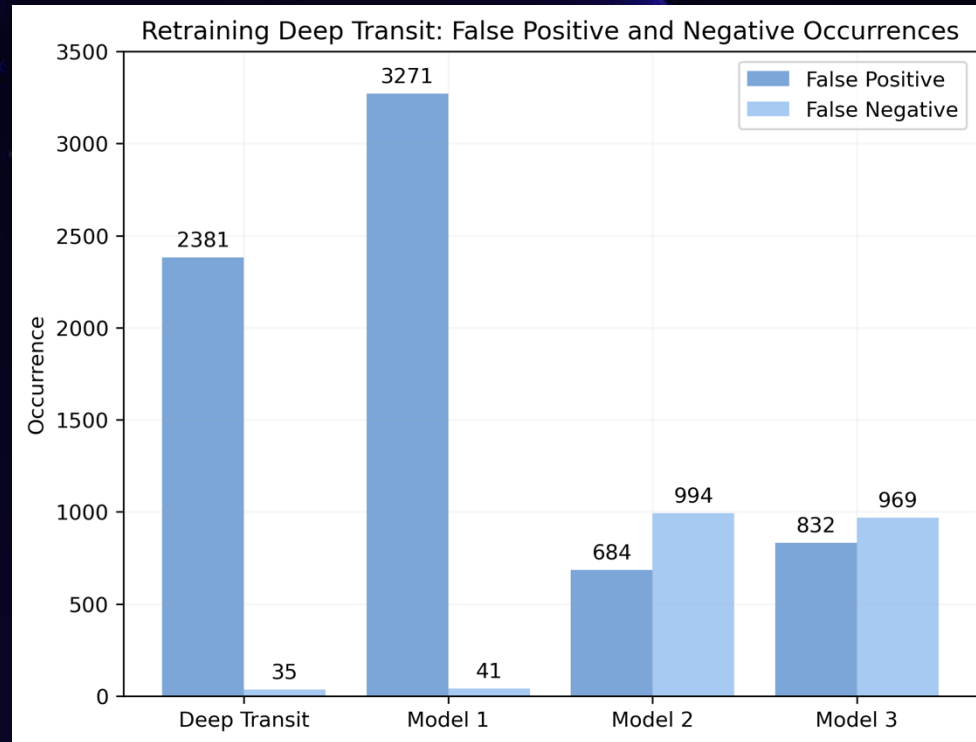
## Statistics

Low training dataset led to no significant change from *DT*

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

# Fall/Spring Models

Model	Precision	Recall
DT	0.680	0.993
1	0.636	0.993
2	0.659	0.571
3	0.627	0.591



# Fall/Spring Models Pt. 2



# 05 Conclusions

Continuation and Expansion  
of Project



# Next Steps

## Parallelize

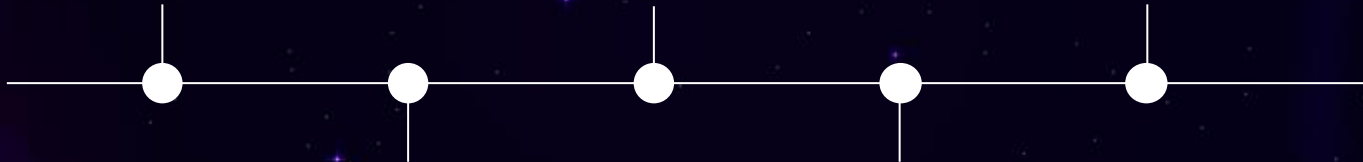
CARC for quicker model generation

## Analyze

Determine optimal parameters

## Implement

Set in use to detect exoplanets

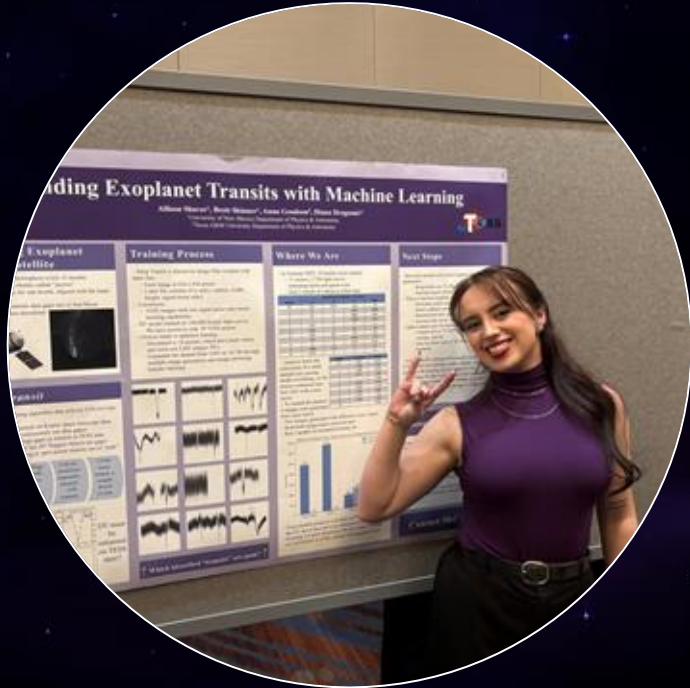


## Repeat

More models, varying parameters

## Final

Create final model



# Acknowledgements

Special thanks to Dr. Diana Dragomir, Brett Skinner, and Anna Goodson for helping with this project

Another thank you to the Rayburn Reaching Up Funds for providing me the opportunity to work on this project



