# Machine Learning for Stellar Occultation Detection in Ground-Based Observations:

## Identifying Occultations by Earth-Orbiting Objects

Peyton E. Cooper*

**The increasing density of objects in Earth's orbital regimes calls for advancements in Space Domain Awareness (SDA) to maintain the safety and efficiency of space operations. This study tackles the challenge of detecting objects orbiting Earth that are too dim for traditional sensors by examining ground based optical collects for indications of a stellar occultation by a satelliting body. With advancements in camera integration times, these rapid and infrequent occurrences around Earth's orbit can finally be captured. The challenge arises when extracting occultation events from other variable noise sources within a ground based optical collection, necessitating the use of machine learning (ML) algorithms,. To achieve this, synthetic data simulating noisy observations was generated and used to train and test a Long Short-Term Memory (LSTM) autoencoder model to identify occultations. The model achieved an accuracy of 99.6% in detecting occultation events, effectively separating noise from meaningful dips in the light curve. These results indicate that, combined with the appropriate tools, this approach can significantly enhance the detection of dim orbiting objects, offering a promising new method for improving Space Domain Awareness.**

## Phantom Fellowship

*The Department of the Air Force-Massachusetts Institute of Technology Artificial Intelligence Accelerator (DAF-MIT AI Accelerator) Phantom Fellowship is a rigorous five-month career-enhancing opportunity for high-performing Department of Defense professionals to assess, develop, and gain exposure to Artificial Intelligence (AI) and Machine Learning(ML) technologies. During the program, enlisted, officer, and government civilian personnel (Phantoms) embed directly with the DAF-MIT AI Accelerator, a cutting-edge organization with the mission of advancing AI/ML technologies for the DoD while also addressing broader societal and ethical issues related to AI/ML. During their Fellowship, Phantoms are required to apply AI/ML foundational knowledge to produce a short-form research paper that seeks to address an AI/ML problem related to their home unit and/or career field. The views expressed in this article are those of the author(s) and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government. The inclusion of external links and references does not imply any endorsement by the author(s), the publishing unit, the Department of the Air Force, the Department of Defense or any other department or agency of the U.S. Government. They are meant to provide an additional perspective or as a supplementary resource.*

## I. Introduction

The ever-increasing congestion of Earth's orbital regimes has pushed advancements in Space Domain Awareness (SDA) sensor technology, data fusion, and analysis methodologies. SDA plays a critical role in modern security by safeguarding space operations that provide satellite communications and GPS to millions of people worldwide. It provides situational awareness pivotal in monitoring, tracking, operating and understanding activities and objects in space by using various sensors such as ground-based radars, optical telescopes, infrared sensors, and satellites. To keep pace with the congested orbital environment around Earth, as depicted in Figure 1, it is imperative that SDA continue to evolve to accurately identify both artificial (e.g. satellite debris) and natural (e.g. micrometeroid) satelliting objects.

*Phantom Fellow, Department of the Air Force-Massachusetts Institute of Technology (DAF-MIT) Artificial Intelligence (AI) Accelerator

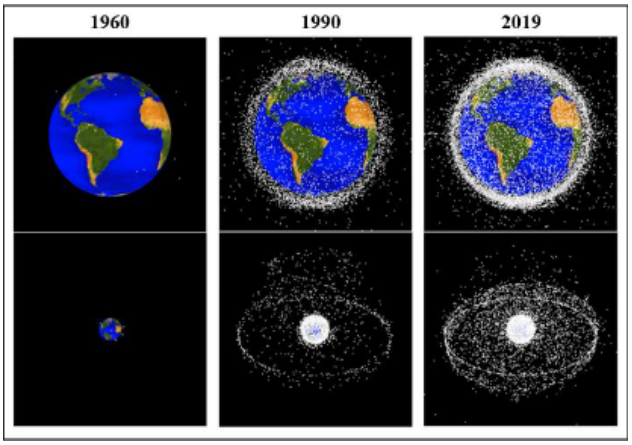


**Fig. 1 Visualization of orbital objects- 1960, 1990, 2019 (Credit: National Aeronautics and Space Administration (NASA) Orbital Debris Program Office)**

This work examined a niche and emerging method for detecting debris too dim for traditional sensors: detecting stellar occultations by passing satelliting objects. An occultation is an event that occurs when one object is hidden from the observer by another object that passes between them. A satelliting object (occulter) that passes between an observer (ground based camera) and a background star will cause a temporary reduction in observed brightness that can be identified as an occultation as shown in Figure 2. This method is typically used to discover exoplanets in distant star systems, where a planet passing in front of a known star is considered to be the occulter [1][2]. However, there is little published research on the probability and photometric effects on stellar occultations by objects in Earth's orbit.

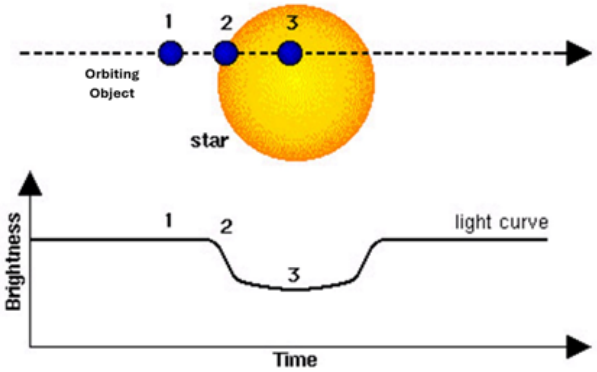


**Fig. 2 Light curve produced by stellar occultation. (Credit: European Space Agency, Space Science Division)**

This method's scope has been relatively unexplored because stellar occultations by objects in Earth's major orbital regimes happen extremely infrequently and at crossing-times far quicker than the average camera's integration time. However, the most recent leading study in this area authored by Groot demonstrates that semiconductor technology has advanced to the point that enables capturing stellar occultations by passing satellites from ground-based sources[3]. While ground-based imaging poses certain challenges (e.g. atmospheric noise, weather), it offers a unique vantage point from which employing this method has actionable SDA applications.

Building on the work by Groot, this study explores various machine learning techniques that can be used for picking out these stellar occultations from other various noise sources in observation data. In combining state-of-the-art imaging technology with machine learning algorithms for event detection, the study aims to improve understanding and identification accuracy of occultation phenomena around our planet.

## II. Preparatory Research

This section presents foundational calculations for stellar occultations by a Resident Space Object (RSO)*. Calculations include determining the probability that a ground based camera will capture a stellar occultation by a satelliting object at a single instance in time. The probability is influenced by the star density within the camera's field of view (FOV) and the RSO's cross-sectional area. With the exception of Equation 3, most of the work comes from Groot's research [3]. These equations contribute to data generation and identification of occultation events presented in later sections.

### A. Probability of Satellite Occultation

Satellite occultations are a rare occurrence. To begin, it was necessary to calculate the density of stars that are visible to the ground based optical camera. For the purposes of this study, parameters from a real-world collector was used in the calculations and is referenced in Subsection III.A.

$$\text{Star Density} = \frac{\text{\# Visible Bright Stars in FOV}}{\text{Total FOV Area}} \tag{1}$$

The RSO will produce an occultation disk with an angular diameter, ($\theta_s$), where $R_s$ is the cross-sectional radius of the RSO and $h$ is the orbital altitude. The occultation disk produced is also the instantaneous fractional area blocked by the satellite [3].

$$\theta_s = \text{atan}\left(\frac{2R_s}{h}\right) \tag{2}$$

At a single instance in time, the probability of observing an occultation by a satellite with a cross-sectional radius of R within the FOV is given by:

$$\text{Probability of Occultation} = \frac{\theta_s^2}{\text{FOV}} \times \text{Star Density} \tag{3}$$

### B. Photometric Effects of Satellite Occultation

When an RSO passes between an observer and a star, it blocks a portion of incoming light creating a dip in the light versus time curve. The effect on nighttime ground-based photometry occurs when occulted stars are blocked out for a duration, $\tau_s$. The impact on the resulting photometry depends on the ratio of the crossing time to the integration time of an exposure, denoted as $f = \frac{t_{\text{int}}}{\tau_s}$. When a satellite shadow passes through the field of view, it creates an occulting trail that obstructs part of the integrated light from stars along its path, with a width $\theta_s$. The fractional decrease in the light of any given star is equal to $\frac{1}{f}$ assuming a complete occultation for the crossing time, $\tau_s$. This crossing time is also dependent on the ratio $\tau_s \propto \frac{2R_s}{D}$, where $D$ represents the aperture diameter of the telescope [3].

## III. Data Generation

When identifying occultations of any type, the dip in the light curve that results from an object passing between the observer and the star is a key indicator. The data structure needed for identifying occultations through machine learning needed to include some measure of brightness versus time as well as a label, indicating if the measure was a result of an occultation.

Producing synthetic data became necessary for multiple reasons due to the scope of the project. Firstly, existing datasets primarily focused on occultation brightness versus time curves for exoplanet detection, lacking adequate comparability for objects significantly smaller orbiting Earth. Secondly, the majority of available data depicted observations from a space-based perspective rather than ground-based, crucial for modeling noise in this ground-focused imaging endeavor. Generating new data enabled the creation of large volumes of labeled noisy data, closely resembling real-life observations captured by the scenario camera.

*A Resident Space Object (RSO) refers to a natural or artificial object that orbits another body. For the purposes of this study, the RSO in question is orbiting Earth.

### A. Problem Setup

| Hardware and Scenario Setup for Data Generation | |
|---|---|
| **Telescope Model** | |
| Aperture Diameter | 1 m |
| Focal Ratio | f/2.5 |
| **Camera Model** | |
| Frame Rate | 5.6 ms |
| Crop | 128 x 128 |
| Binning Capabilities | 4x4 |
| FOV | 0.038 side deg |
| IFOV | 4.3 arcseconds |
| **Scenario Model** | |
| RSO | 5-meter sphere |
| Orbital Parameters | semi-major axis, a=35,786km<br>inclination, $i \neq 0$<br>eccentricity, $e \approx 0$ |
| Time Constraints | Nighttime collects |
| Star Brightness Threshold | $V_{mag}$<=11 |

**Fig. 3 Table showing the hardware and scenario parameters conceptualized for data generation.**

The scenario camera chosen was the Andor iXon 888 EMCCD with parameters listed above in Figure 3. This specific camera was chosen because it will be used to collect real-world data to pass through the ML model for future iterations of this project. The same logic was used in modeling the Planewave 1000 prime-focus telescope found in Figure 3.

The RSO is conceptualized as a 5-meter sphere, observed in eclipse to preclude sunlight reflection. Within the study's framework, the duration of the eclipse is disregarded. The RSO's orbital parameters listed in Figure 3 mirror the behavior of authentic objects either maintaining station or drifting gradually at Geostationary Equatorial Orbit (GEO). The study incorporates a 10% decrement in photon flux as a baseline assumption. This assumption originates from the camera's quantum efficiency curve (provided by the manufacturer). Collecting at a wavelength of 550 nm the camera achieves 90% quantum efficiency. The camera's FOV when pointing at GEO in terms of Right Ascension (RA) and Declination (DEC) can be visualized by the red dashed line in Figure 4.

### B. Simulating Noisy Observations

The data generation process began with querying the Gaia catalog using the astropy python package [4] within specified RA and DEC coordinates shown in Figure 4 at a given epoch time to identify stars with a visual magnitude brighter than 11 as defined in Section III.A. The RA and DEC coordinates were determined by the camera's physical location on Earth (Maui, HI) and pointing parameters listed in Figure 3. The Gaia catalog returned the baseline Visual Magnitude, RA, DEC, for each visible Star ID in the FOV at an instance in time. The visible stars for a singular point in time can be seen in Figure 4,

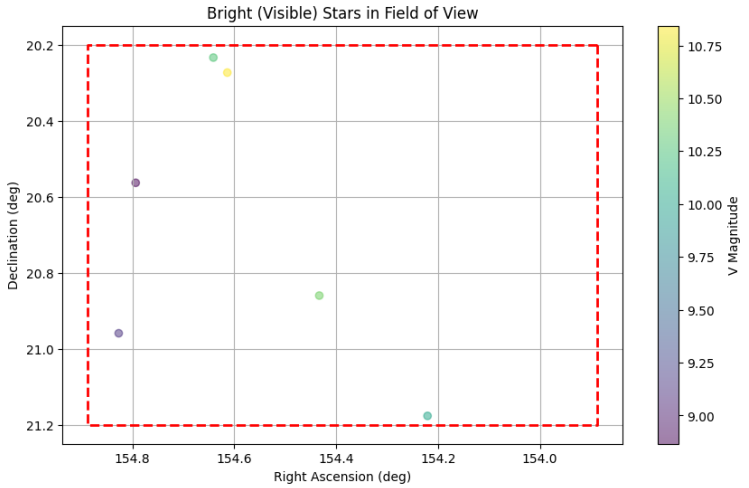


**Fig. 4 Gaia query of all stars within the FOV that are bright enough for the camera to detect (Vmag <=11). The red dashed line represents the camera's FOV.**

Once the visible stars were identified, variable noise was incorporated into the dataset by sampling from a Gaussian distribution of the uncertainty in the observed star's magnitude which was derived from the definition of magnitude differences in terms of flux ratios [5]. This distribution is visualized for one star within the FOV in Figure 6 and the equation for the uncertainty in the star's observed magnitude, $\Delta m$, can be seen below in Equation 4.

$$\Delta m \approx 2.5 \log\left(\frac{S+N}{S}\right) \tag{4}$$

For each observation with variable noise, the base incident photon count was computed, after which noise sources were introduced. Electronic noise and sky background noise were converted from electrons to photons using the quantum efficiency factor, assumed to be 90% efficient stemming from the description provided in Section III.A. The photon counts from these sources were then individually summed to obtain the total photon count with noise, visualized in Figure 5. This process ensured that the generated dataset more accurately represented the noise characteristics inherent in real observational data.

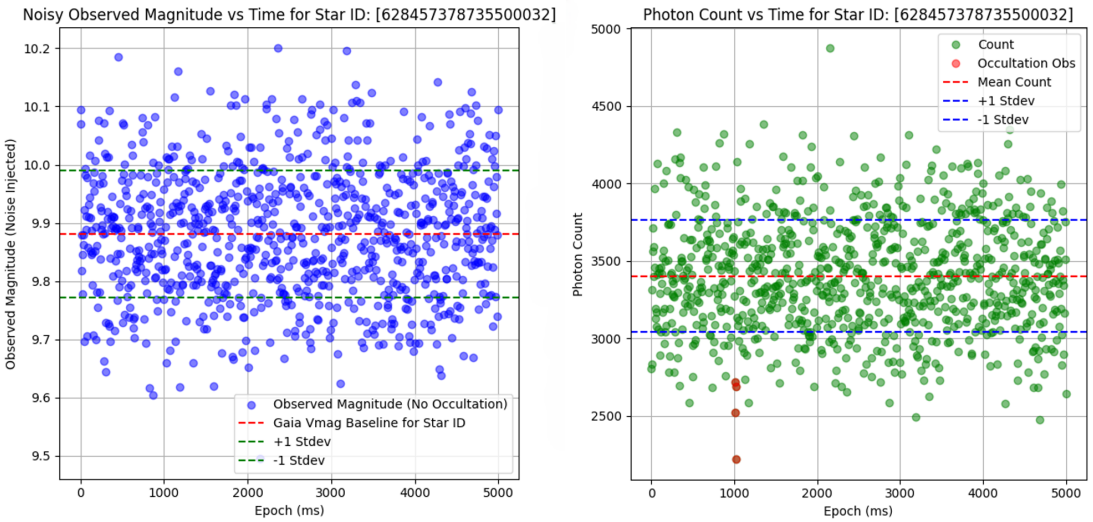


**Fig. 5 Left: Observed Visual Magnitude values for a singular star over a 5 second observation period. Observed Visual Magnitude is the recorded Visual Magnitude in the Gaia database with the noise model applied. The distribution of noise can be seen in Figure 6. There are no occultations injected in this Left image. Right: The photon count values calculated from the Observed Visual Magnitude values on the left with occultations injected (shown in red).**

Employing the derived equations in Section II.B to forecast the fractional decrease in brightness anticipated during occultation events, the dataset underwent a phase where occultations were randomly injected. An observation that was selected for occultation injection underwent a percentage decrease in the photon count. Each observation was labeled to distinguish between instances of occultations and nominal observations. This occultation injection process can be seen above in Figure 5 on the right and below in Figure 6.

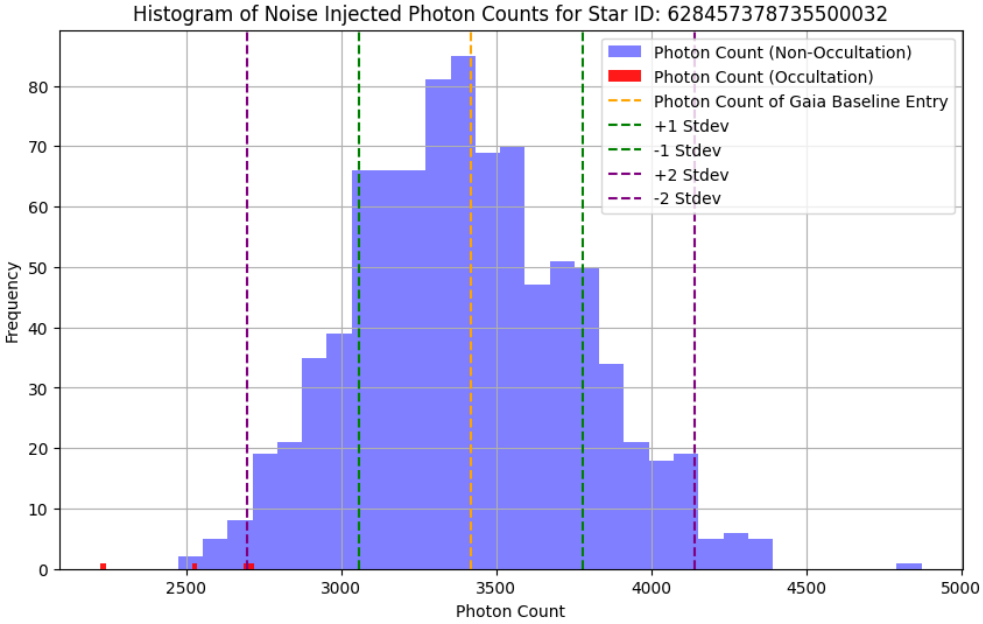


**Fig. 6 Visualization of the noise distribution for photon count readings. This star was randomly selected for an occultation injection at a series of time points, $f = \frac{t_{\text{int}}}{\tau_s}$, with decrease in count, $\frac{1}{f}$, as notated by the red dots. Total time represented is 5 seconds for one star.**

This process was systematically replicated across multiple dates in time to ensure a diverse and comprehensive dataset. This iteration process yielded an extensive dataset spanning over 5 distinct Julian days, within each day a minimum of 5 seconds and a maximum of 21 seconds of data was collected, at an interval of 5.6 ms increments. This spread produced a diverse set of 103,251 observation points in total, providing a robust foundation for training and refining machine learning models tailored for occultation detection.

## IV. Machine Learning for Detection of Stellar Occultations by Satelliting Objects

Various methodologies exist for detecting anomalies in time series data. Rule-based methods are effective for data with minimal noise, where specific constraints can be applied to identify outliers. However, when datasets contain scattered noise, as illustrated in Figure 5, rule-based methods become inadequate, necessitating the use of linear regression techniques to identify outliers.

When these linear methods prove insufficient in their their ability to accurately detect anomalous data, more complex nonlinear regression techniques, including neural networks, are employed. In this study, a Long Short-Term Memory (LSTM) autoencoder was chosen. LSTM networks, a type of recurrent neural network (RNN), are known for their robustness against input fluctuations [6]. The following sections will explore the advantages of both linear and nonlinear regression methods for detecting stellar occultations from satelliting objects, with a focus on the application of the LSTM autoencoder.

### A. Linear Model

The first and most straight-forward model trained was a linear regression model. The algorithm first identifies the ideal polynomial degree with which to train the linear model by analysing mean squared error and residuals. The model is trained using probabilistic inference of the posterior distribution to make predictions. More specifically, the linear model took the normalized observed photon count as input and outputs a predicted value for the following observation's photon count, from here the model was able to decide based on the residuals and mean squared error if there was an outlier. The linear model proved useful for noise reduction in the data. The linear model begins to break down with larger data sets (n>600) and under-fitting quickly becomes problematic.

Despite its convenience, the linear regression model inherently assumes a direct linear relationship between photon count and time. While this assumption may hold for ideal, low-noise simulated data, it often fails when applied to more complex and noisy real-world data [7].

### B. LSTM Autoencoder

The decision to employ an LSTM (Long Short-Term Memory) autoencoder for occultation detection stems from its proficiency in handling noisy time series data, rather than temporal dependencies. In the context of occultation detection, where anomalies are sporadic and embedded within noisy datasets, the LSTM autoencoder offers several advantages. Firstly, its resilience to noise and irregularities allows it to effectively filter out irrelevant fluctuations and focus on identifying meaningful patterns, such as occultation events. Additionally, the flexibility of LSTMs in processing sequences of variable lengths aligns well with the unpredictable nature of occultations, which can vary in duration and intensity. By training exclusively on normal data, the LSTM autoencoder can learn to reconstruct the typical patterns present in the data while minimizing reconstruction error, thus enabling it to discern anomalies such as occultations based on deviations from normal behavior. This unsupervised learning approach eliminates the need for labeled anomalous data, which may be scarce or difficult to obtain in practice.

#### *1. Data Conditioning, Data Split, and Training*

The data was first normalized, plotted as seen in Figure 7, and stored. Before training, all anomaly (occultation) counts were filtered out of the data set. As the task of the autoencoder was mapping normal observations, the data was shuffled out of sequential order for a more representative distribution.

The resulting data was split into three distinct subsets: the training set took 85% of the data, validation set took 10%, and the testing set took 5%. The model was trained, validated, and tested exclusively with normalized photon counts which did not contain occultations.


**Fig. 7 Visualization of all normalized "normal" (no occultation) and "anomalous" (only occultation) data points. The shaded areas around the red and blue distributions represent +/- 1 standard deviation. The area where the two shadows overlap is the area of uncertainty that the model is attempting to clarify in order to better separate occultations from noise.**

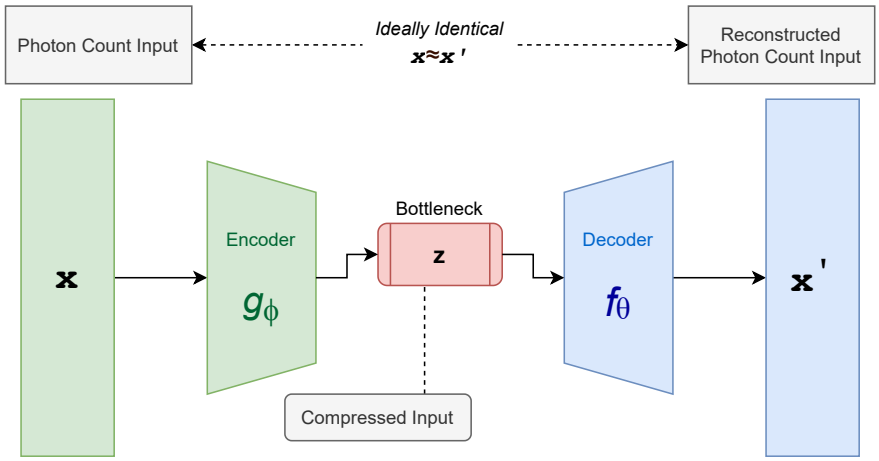


**Fig. 8 LSTM Autoencoder process. The encoder function, $g_\phi$ receives the normalized photon count input and translates it into latent low-dimensional code, $\mathbf{z} = g_\phi(\mathbf{x})$. The decoder function, $f_\theta$, recovers the data from the compressed code and reconstructs the input as $\mathbf{x}' = f_\theta(g_\phi(\mathbf{x}))$ [8].**

The encoder utilized two LSTM layers to compress the normalized photon data input. Likewise, the decoder contained two LSTM layers and an output layer which held the final reconstruction, visualized in Figure 8. The parameters $(\theta, \phi)$ were jointly optimized to produce a reconstructed data sample that closely matched the original input. This process to learn the identity function can be expressed as $\mathbf{x} \approx f_\theta(g_\phi(\mathbf{x}))$. To evaluate the difference between the original and reconstructed vectors, an L1 loss function was used. The L1 loss function quantified this difference by calculating the mean absolute error between the actual and predicted values. The training algorithm looped through

feeding the model with the training set and minimizing loss at each epoch by evaluating and recording the performance with the validation set. Training was conducted on 41,261 (of the 103,251) observation points and the model was saved for evaluation.

*2. Occultation Detection Results*

Once the model was trained, a threshold value was needed in order to establish a binary classification task. A normal observation would be identified when the reconstruction loss was below the threshold, and an occultation would be identified if the loss was higher than the threshold. The threshold was chosen by predicting the losses on the trained data set and identifying the point at which the upper 5% of losses resided. The results of the loss distributions can be visualized in Figure 9.

Using this threshold, and notating "positive" as an occultation and "negative" as a normal noisy observation, the model achieved a 99.6% accuracy.

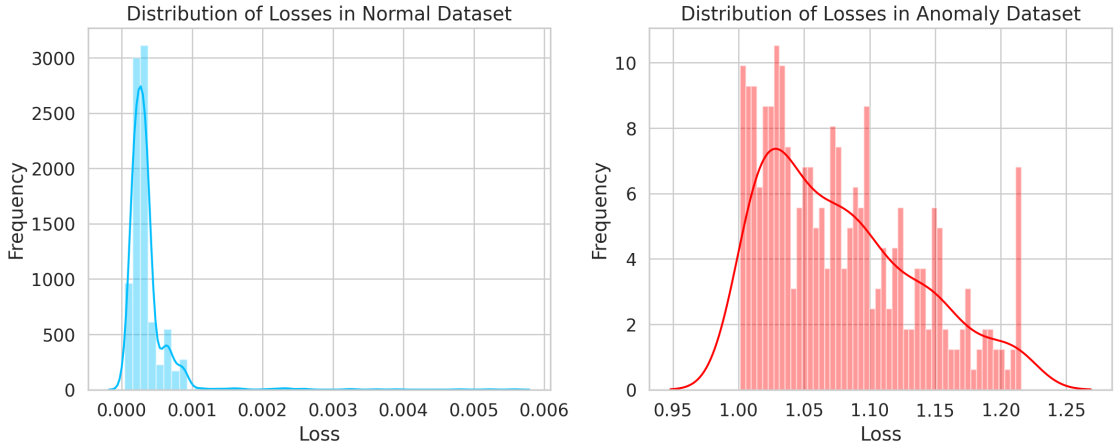


**Fig. 9 Left: Modeled distributed losses of normal observations (no stellar occultations) of normal distributions. Right: Modeled distributed losses of anomalous observations (occultations). Based on these distributions, occultations were able to be picked out of the noisy data.**

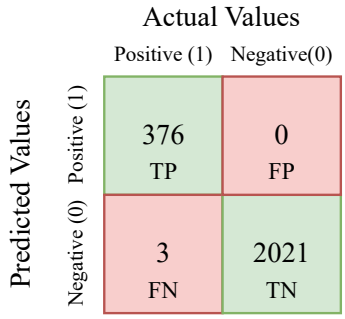


**Fig. 10 Confusion matrix for 2,400 test points in test set with a 5m RSO.**

## V. Discussion

The application of an LSTM autoencoder to detect stellar occultations by orbiting satellites has demonstrated promising results, achieving high accuracy in separating occultation events from noisy data. However, several limitations and considerations must be acknowledged to contextualize these findings and guide future work.

### A. Model Limitations

One notable limitation is related to the data generation process. The synthetic data used to train and evaluate the model was based on the assumption that all RSOs have a cross-sectional diameter of 4 meters or larger. However, many RSOs are smaller than this threshold. Additionally, the study was conducted solely for objects in GEO. To improve the model's accuracy and applicability, it is essential to generate synthetic data that includes smaller RSOs and to extend the study to encompass Medium Earth Orbit (MEO) and Low Earth orbit (LEO) objects. Immediate steps should focus on incorporating these smaller objects into the data generation process and re-evaluating the trained model's performance across GEO, MEO, and LEO conditions.

Additionally, while the LSTM autoencoder performed well on the simulated data, its performance on real-world data may vary. The synthetic data, despite being noisy, might not capture all the complexities and variabilities present in actual observational data. Factors such as atmospheric interference, varying star densities, and unexpected noise sources could affect the model's efficacy. Consequently, gathering real-world data and testing the model in a practical setting is crucial. The results from these tests will provide insights into the model's robustness and highlight areas needing refinement.

### B. Immediate Steps and Future Iterations

The immediate next steps involve collecting real-world observational data using the Andor iXon 888 EMCCD camera and Planewave 1000 prime-focus telescope as outlined in the study. This data will be used to test the current LSTM model, providing a benchmark for its real-world performance. Depending on the outcomes, the model may require retraining or fine-tuning to better handle the nuances of real-world observations. While the LSTM autoencoder is the main product here, a strong GEO observation generation algorithm was also developed. Validating the noise model within the data generation algorithm will allow the algorithm to be used and distributed as a tool for when real-world data is not an option.

Subsequent iterations could include a trade study on camera parameters such as integration time, collection wavelengths, frame rate, and sensitivity. Optimizing these parameters could enhance the model's detection accuracy and reliability.

### C. Broader Implications

While this is a small piece of the puzzle, the successful implementation of this approach has broader implications for improving Space Domain Awareness. Enhanced detection capabilities for dim RSOs can significantly improve collision avoidance measures, satellite tracking, and overall space traffic management. This, in turn, contributes to the safety and sustainability of space operations, which is increasingly critical given the growing congestion in Earth's orbital regimes.

## VI. Acknowledgements

This undertaking would have been entirely impossible without the help of my MIT Lincoln Laboratory technical advisor, Dr. Deborah Woods. Her experience in related works, research practices, and all things STEM has been invaluable. She went above and beyond what was asked of her, and I am honored to have her mentorship. Thank you, Deb!

Special thanks for Steven Vogl for spending hours with me early on thinking of research project ideas. It led to an amazing journey, and I am truly excited to keep digging down this rabbit hole. Finally, to my amazing husband and MIT AIA Phantom Cohort X partners, thank you for being an endless source of advice, support, and laughter.

## References

[1] Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. , Bean, J. L., Berta-Thompson, Z. K., Brown, T. M., Buchhave, L., Butler, N. R., Butler, R. P., Chaplin, W. J., Charbonneau, D., Christensen-Dalsgaard, J., Clampin, M., Deming, D., Doty, J., Lee, N. D., Dressing, C., Dunham, E. W., Endl, M., Fressin, F., Ge, J., Henning, T., Holman, M. J., Howard, A. W., Ida, S., Jenkins, J., Jernigan, G., Johnson, J. A., Kaltenegger, L., Kawai, N., Kjeldsen, H., Laughlin, G., Levine, A. M., Lin, D., Lissauer, J. J., MacQueen, P., Marcy, G., McCullough, P. R., Morton, T. D., Narita, N., Paegert, M., Palle, E., Pepe, F., Pepper, J., Quirrenbach, A., Rinehart, S. A., Sasselov, D., Sato, B., Seager, S., Sozzetti, A., Stassun, K. G., Sullivan, P., Szentgyorgyi, A., Torres, G., Udry, S., and Villasenor, J., "Transiting Exoplanet Survey Satellite (TESS)," *Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave*, Vol. 9143, SPIE, 2014, pp. 556–570. https://doi.org/10.1117/12.2063489,

URL https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9143/914320/Transiting-Exoplanet-Survey-Satellite-TESS/10.1117/12.2063489.full.

[2] Ofman, L., Averbuch, A., Shliselberg, A., Benaun, I., Segev, D., and Rissman, A., "Automated identification of transiting exoplanet candidates in NASA Transiting Exoplanets Survey Satellite (TESS) data with machine learning methods," *New Astronomy*, Vol. 91, 2022, p. 101693. https://doi.org/10.1016/j.newast.2021.101693, URL https://www.sciencedirect.com/science/article/pii/S1384107621001196.

[3] Groot, P. J., "Satellite Shadows Through Stellar Occultations," *Astronomy & Astrophysics*, Vol. 667, 2022, p. A45. https://doi.org/10.1051/0004-6361/202244496, URL https://www.aanda.org/articles/aa/abs/2022/11/aa44496-22/aa44496-22.html, publisher: EDP Sciences.

[4] "Gaia Archive," , ???? URL https://gea.esac.esa.int/archive/.

[5] Simmonetti, J., "Measuring the Signal-to-Noise Ratio S/N of the CCD Image of a Star or Nebula," , Jan. 2004. URL https://www1.phys.vt.edu/~jhs/phys3154/snr20040108.pdf.

[6] Brownlee, J., "A Gentle Introduction to LSTM Autoencoders," , Nov. 2018. URL https://machinelearningmastery.com/lstm-autoencoders/.

[7] de Beurs, Z. L., Vanderburg, A., Shallue, C. J., Dumusque, X., Cameron, A. C., Leet, C., Buchhave, L. A., Cosentino, R., Ghedina, A., Haywood, R. D., Langellier, N., Latham, D. W., López-Morales, M., Mayor, M., Micela, G., Milbourne, T. W., Mortier, A., Molinari, E., Pepe, F., Phillips, D. F., Pinamonti, M., Piotto, G., Rice, K., Sasselov, D., Sozzetti, A., Udry, S., and Watson, C. A., "Identifying Exoplanets with Deep Learning. IV. Removing Stellar Activity Signals from Radial Velocity Measurements Using Neural Networks," *The Astronomical Journal*, Vol. 164, 2022, p. 49. https://doi.org/10.3847/1538-3881/ac738e, URL https://ui.adsabs.harvard.edu/abs/2022AJ....164...49D, aDS Bibcode: 2022AJ....164...49D.

[8] Weng, L., "From Autoencoder to Beta-VAE | Lil'Log," , ???? URL https://lilianweng.github.io/posts/2018-08-12-vae/.