

Olivet Nazarene University

Digital Commons @ Olivet

Honors Program Projects

Honors Program

Spring 2022

Preprocessing of Astronomical Images from the NEOWISE Survey for Near-Earth Asteroid Detection

Rachel Meyer
rrmeyer22@gmail.com

Follow this and additional works at: https://digitalcommons.olivet.edu/honr_proj



Part of the [Physical Sciences and Mathematics Commons](#)

Recommended Citation

Meyer, Rachel, "Preprocessing of Astronomical Images from the NEOWISE Survey for Near-Earth Asteroid Detection" (2022). *Honors Program Projects*. 137.
https://digitalcommons.olivet.edu/honr_proj/137

This Thesis is brought to you for free and open access by the Honors Program at Digital Commons @ Olivet. It has been accepted for inclusion in Honors Program Projects by an authorized administrator of Digital Commons @ Olivet. For more information, please contact digitalcommons@olivet.edu.

Preprocessing of Astronomical Images from the NEOWISE Survey for Near-Earth Asteroid
Detection with Machine Learning

By

Rachel R. Meyer

Honors Scholarship Project

Submitted to the Faculty of

Olivet Nazarene University

for partial fulfillment of the requirements for

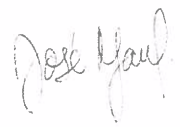
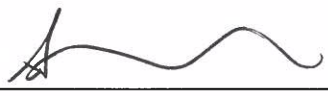

GRADUATION WITH UNIVERSITY HONORS

March, 2022

BACHELOR OF SCIENCE

in

Engineering, Electrical Concentration

<u>José Manjarrés</u> Scholarship Project Advisor (printed)	 Signature	<u>2/17/2022</u> Date
<u>Stephen Case</u> Honors Council Chair (printed)	 Signature	<u>4.27.22</u> Date
<u>Daniel Sharda</u> Honors Council Member (printed)	 Signature	<u>4/27/22</u> Date

Preprocessing of Astronomical Images from the NEOWISE Survey for Near-Earth Asteroid

Detection with Machine Learning

By

Rachel R. Meyer

Honors Scholarship Project

Submitted to the Faculty of

Olivet Nazarene University

for partial fulfillment of the requirements for

GRADUATION WITH UNIVERSITY HONORS

March, 2022

BACHELOR OF SCIENCE

in

Engineering, Electrical Concentration

José Manjarrés
Scholarship Project Advisor (printed)



Signature

2/17/2022
Date

Honors Council Chair (printed)

Signature

Date

Honors Council Member (printed)

Signature

Date

ACKNOWLEDGEMENTS

Financial support for this project was provided by the Craighton T. and Linda G. Hippenhammer Faculty Scholarship Grant, as well as the Honors Department of Olivet Nazarene University. The Pence-Boyce STEM Student Scholarship also provided the opportunity to work on this project full-time as an internship. Mentorship and guidance were provided by Dr. José Manjarrés, Assistant Professor in the Department of Engineering in the School of STEM. This project would not be complete without his knowledge and assistance. I would also like to acknowledge Dr. Stephen Case for his guidance and expertise in astronomy. Thank you to the other students and professors in my honors cohort, Cohort XII: The Imposters. I would also like to thank my roommates, Megan Drenth and Lauren Kee, for sticking with me through this process, and encouraging me the whole time. Lastly, I would like to thank Mani Martinez for the many nights he stayed up with me and supported me while I worked on this research. I couldn't have done it without you all.

TABLE OF CONTENTS

Acknowledgements.....	ii
List of Figures	iv
Abstract.....	1
Introduction	2
Asteroid Surveys.....	2
Use of Machine Learning.....	5
Methods.....	8
Data Collection	8
Initial Data Analysis	9
Preprocessing Pipeline	12
Numerical Source Data Classification.....	14
Results.....	15
Discussion	18
References	24

LIST OF FIGURES

Figure 1. Right ascension vs declination for each source	10
Figure 2. Absolute magnitude vs visual magnitude for each source	11
Figure 3. Wavelength 1 vs wavelength 2 for each source	11
Figure 4. Preprocessing pipeline	13
Figure 5. Logistic regression confusion matrix.....	16
Figure 6. Support vector machine confusion matrix	17
Figure 7. Naïve Bayes confusion matrix.....	17
Figure 8. Random forest confusion matrix	18
Figure 9. Collected images from the NEOWISE survey of known asteroids	21
Figure 10. An additional image of a larger known asteroid.....	21
Figure 11. A NEOWISE image example of a non-asteroid.....	22

ABSTRACT

Asteroid detection is a common field in astronomy for planetary defense, requiring observations from survey telescopes to detect and classify different objects. The amount of data collected each night is continually increasing as new and better-designed telescopes begin collecting information each year. This amount of data is quickly becoming unmanageable, and researchers are looking for ways to better process this data. The most feasible current solution is to implement computer algorithms to automatically detect these sources and then use machine learning to create a more efficient and accurate method of classification. Implementation of such methods has previously focused on larger asteroids, but this leads to less accuracy in classifying smaller and slower-moving asteroids. We use the source data from the Near-Earth Object Widefield Infrared Survey Explorer (NEOWISE) survey telescope to create a preprocessing pipeline allowing for better training and testing data for machine learning algorithms. We were able to create a pipeline that finds sources on an image within 0.02 degrees as well as a collection of data on the known sources. Next, we used several machine learning classifying models, including a logistic regression classifier, a support vector machine (SVM), a naïve Bayes classifier, and a random forest. Finally, we present and discuss these results as a confusion matrix for each model, describing the ability to classify a source as an asteroid or not. This was done with only the numerical data collected. In the future, to create a better classifier, we would use this data along with classified images to develop a system that could predict the possibility of a detected source within a NEOWISE image being an asteroid or not.

Keywords: machine learning, asteroid detection, astronomy, near-earth object

INTRODUCTION

Improving asteroid detection has become increasingly important, as near-Earth asteroids (NEAs) have proven dangerous even at small sizes and velocities. For example, the Chelyabinsk meteor that exploded over a town in Russia that caused structural damage as well as many injuries, was only 20 meters or 66 feet in diameter. All information collected about these asteroids is beneficial either for impact prevention or for potential asteroid mining, which is becoming a desirable option for gathering natural resources for use on Earth. There is an increase in data each year used to detect NEAs and potentially hazardous asteroids (PHAs), and the incoming images and the additional data from survey telescopes are quickly becoming unmanageable to process via human analysis [1], [2]. In 1998 the Very Large Telescope (VLT) produced 10GB of data, while in 2019 the Large Synoptic Survey Telescope (LSST) produced around 20TB or 20,000GB of data per night. Recent studies have investigated the implementation of machine learning to decipher images collected from these telescopes to show that there are more accurate and faster ways to detect NEAs. Previous studies have demonstrated the effectiveness of machine learning for the detection and classification of NEAs in survey data [3], [4], [5]. The present research project first explores the astronomy behind asteroids, including their magnitude and other astronomical units such as right ascension and declination, before discussing the application of machine learning and the different models and algorithms used to classify asteroids.

Asteroid Surveys

The primary type of data in previous asteroid detection research has been image surveys, created either with the intent to observe all types of transient objects or to observe asteroids specifically. In such studies, the survey telescope images the same part of the night sky

multiple times per night with the intention of observing any transients moving over time [1], [9], [10], [13]. Astronomers and researchers often processed the raw data in different ways to fit better with the objective of the research. For example, Duev et al. converted survey images into grayscale because color was not necessary for the detection of the bright pixels and ultimately simplified the data and therefore the processing [2]. In multiple approaches, researchers subtracted known sources such as stars, planets, or satellites from images [6], [7], [8], and [13]. The researchers then utilized these subtracted or differenced images, focusing solely on the transients or unknowns that remained after the knowns were subtracted out. Another example of data preparation is the use of random horizontal and vertical image flipping rotations to avoid bias in the results [3]. The purpose of flipping the orientation of the images allowed for a better image set because there were source image examples from a variety of different perspectives that could then be used for processing. The most common sources of data that were used for these surveys included the Zwicky Transient Facility survey [2], [7] and the images collected from the Linear Synoptic Survey Telescope [9], [10]. Larger amounts of data are available for use for researchers as well as the general public, but, as made clear in previous studies, this raw data must be tweaked and manipulated with preprocessing steps in order to best serve the objective of the research and ultimate goal of the researcher.

The Near-Earth Object Widefield Infrared Survey Explorer (NEOWISE) was created to find NEAs and provides much of the sky pictured in two infrared wavelength bands (3.4 and 4.6 μm) as well as data on multiple types of sources. The two different infrared bands capture different features based on the wavelengths given off from each source, allowing for more features to be observed on each source. Each image taken with NEOWISE has a specific scan identification as well as a frame number, allowing each image to have a unique name or

identification number. Some of the data provided for each image include the location of the image in right ascension and declination, absolute magnitude and visual magnitude of the objects in the image, and the difference in intensity of the infrared bands for each image object. Right ascension and declination are absolute units used to locate objects in the sky. Right ascension can be measured in hours, minutes, and seconds (0 to 24 hours) as well as in degrees (0 to 360 degrees), while declination is measured only in degrees (-90 to 90 degrees). Absolute magnitude is the brightness of a celestial object as seen from a set distance and gives information about the true luminosity of the object, whereas visual magnitude is the brightness of the object as viewed from Earth. The smaller the magnitude the brighter the object. Most of the stars in our night sky have visual magnitudes from 1 to 6 while our Sun has a magnitude of -26.

Previous asteroid-detection work has focused on faster-moving asteroids, often to the detriment of detecting slower-moving asteroids [6], [11], [12]. In survey images, faster asteroids show up as trails or streaks; the faster the asteroid is moving, the longer the streak. A common detection method is recognizing these streaks, but this method performs poorly on slower-moving asteroids that may appear as only a couple of pixels and can look very similar to other stars. A better way to detect slower-moving asteroids is to recognize moving pixels between sets of images from the same portion of the night sky allowing a small amount of time to elapse between each image. The NEOWISE survey dataset contains considerable overlap between images, allowing for tracklets of images (or string of images with the moving object) to be created. These tracklets allow for better detection of slower asteroids, which appear as “moving pixels” when the images in each tracklet are compared.

Detection of most asteroids has historically relied on human interaction in terms of viewing the images and picking out each image that included streaks or pixels moving over time. This method is quite time-consuming and requires multiple human reviews for each image. Speeding up this process is beneficial in the long run because it would allow for same night follow-up, allowing further observation of the object of interest before it moves out of view and is unable to be seen. Multiple researchers point to machine learning as the solution to quicken and improve the asteroid-detection process [3], [4], and [5].

Use of Machine Learning

Researchers have previously used algorithms such as random forest ([1], [7]), k-Nearest Neighbors ([1], [8], [9]), and support vector machines ([10]) for the classification of detected sources. These methods have sped up the process of classification as compared to previous systems that relied heavily on human interaction.

One example of machine learning that could be fruitfully applied to asteroid detection is logistic regression. Logistic regression is a binary classifier that uses a logistic function (a common sigmoid curve) to model a dependent variable. The output of logistic regression model is the classification of a source as an asteroid or not (true or false). The goal is to determine the boundary that divides classes of data based on fixed values of the model parameters.

Another machine learning algorithm that could be applied to asteroid detection is a support vector machine. To train a support vector machine, a user must use datasets that are marked as belonging to two distinct categories. The model can then predict the possibility of a new data point (in this case, an object in an image) belonging to one class or another (e.g., “asteroid or not”) based on where it falls on a graph created using the features of both classes. The goal of a support vector machine is to find a hyperplane, or an area of overlap, within a

space that distinctly classifies the source points with the maximum margin, or the maximum distance between both points. The dimension of the hyperplane depends on the number of features. A support vector machine kernel transforms the data and finds the process, beyond a linear hyperplane, that best separates the data based on the user-defined classes. This becomes very useful when dealing with a non-linear classification problem.

The naïve Bayes classifier can be used for source classification by applying the Bayes Theorem, which describes the probability of an event based on the prior knowledge of the conditions that are related the event (in this case being an asteroid), with the assumption of conditional independence between every pair of features given the value of the class variable. For example, the features for an asteroid might include the location, brightness, or size of the asteroid. The Gaussian naïve Bayes algorithm assumes the likelihood of the features to be Gaussian, which is a function that is visualized as a symmetric bell curve.

Finally, decision trees offer another aspect of machine learning that could usefully be applied to asteroid detection. Decisions trees use a tree-like model of decisions in order to reach a conclusion about the classification of a source. A common example is the classification of iris flowers. First, the tree looks at the sepal petal length and sets a threshold that splits the possible values into two different categories, and the next decisions are based on the sepal width, petal length, petal width until every picture is classified into 1 of 3 different categories. The model should be able to predict the class of a source by creating simple decision rules assumed from the data features. A random forest is a model of machine learning that constructs a multitude of decision trees for the classification of the data points.

In each of these models, results are displayed using confusion matrices. In order to convey the number of correct and incorrect predictions, the results are broken into classes: true

positives (sources that are true and classified as true), true negatives (sources that are false and classified as false), false positives (sources that are false but classified as true) and false negatives (sources that are true but classified as false). Each of these categories are depicted in the confusion matrix. These can then be used to calculate other important aspects of the classifier, such as accuracy, precision, and recall.

A machine learning model that has proven especially efficient in the field of image classification is Convolutional Neural Networks (CNNs). CNNs are a type of machine learning architecture that implement convolution kernels, or filters, on the input features of the data to create feature maps. A CNN model needs training data to establish training weights and testing data to check its performance. Each image input passes through convolutional layers with filters or kernels. The process involves both convolution layers and pooling layers, which reduce the dimensions of the data by combining outputs at one layer into a single neuron, which is just one part of the larger neural layer, in the next layer, and finally a fully connected layer that applies the sigmoid function, an s-type curve activation function used to add non-linearity, to classify a source with a binary value of 1 or 0. For this research we are concerned with binary classification (asteroid or not an asteroid) and smaller images, so this would lead to fewer layers overall in a CNN.

CNNs can speed up processing by providing more accurate predictions than human-based classification by eliminating the need for a person to check anomalies and asteroid candidates in each image. Previous research has shown that a CNN-based approach is more accurate at classifying streaks and other abnormal pixels as asteroids [5], [6], [10], and [11]. The CNN architecture can perform better than more simple machine learning algorithms, such as the

ones previously mentioned, with image classification because of its hierarchical model that works on building a complex network.

Given the limitations of the current asteroid-detection approaches, this research attempted to develop a pipeline that could more accurately detect slow-moving asteroids. To detect the slower and dimmer asteroids, the images from the NEOWISE survey from multiple years were processed using machine learning to become more efficient at detecting slower and dimmer asteroids. This included a focus on the images, as well as the metadata from the database on each individual object such as the visible magnitude, absolute magnitude, and the magnitude of the object in the first and second wavelength bands. Our work builds on existing efforts in implementing deep learning to detect transients and possible asteroid candidates in asteroid surveys and improving the ability to recognize irregular moving asteroids, by exploring the different machine learning models outlined here with data collected from the NEOWISE asteroid survey.

METHODS

Data Collection:

Data was collected from the NASA/IPAC Infrared Science Archive (IRSA) database, specifically the Near-Earth Object Wide-Infrared Explorer (NEOWISE) survey. To retrieve the images and other source data, we used MySQL programming to perform database inquiries. This allowed us to collect images that contained different types of source. The images collected were single-frame images from a range of -50 to 50 degrees in declination and the full range of right ascension. This allowed full coverage of the celestial equator around the entire celestial sphere.

We kept the range of declination smaller to reduce the number of images needed, and we noticed that there were less asteroids in the areas that we did not observe.

Within these images, we collected data on 550,264 sources, which included asteroids, comets, planets, and planetary satellites. Data on each source included the object's location (in right ascension and declination) and its absolute and visual magnitude. In addition, data included the object's magnitude in infrared bands 1 and 2, the difference between these band values of each object, the image identification, and the x and y coordinates of the source within its image. To convert astronomical coordinates into pixels, we first converted the right ascension and declination values that were originally in degrees into arcseconds. These values were then converted to pixels using the NEOWISE pixel scale of 2.75 arcseconds per pixel. To save the images and data as workable files, we saved them as NumPy arrays and saved the data separately as CSV files. Both file names were named as the unique source identification and frame number of that image to keep them coordinated. We then applied a logarithmic filter to the image to improve visualization, scaling the values down to readable values and leaving the background and sources on the image.

Initial Data Analysis

To better understand the collected data, we began by analyzing the sources separately by their specified labels. For this dataset, the labels were asteroid (A), comet (C), planet (P), and planetary satellite (S). To analyze the data, we calculate the minimum, median, and maximum for each different measurement collected for the source. Figures 1 to 3 show some of the different features plotted for both asteroids and non-asteroids. Figure 1 shows the general location of the objects, indicating that many of the non-asteroids lie within the region of the asteroids in the solar system, along the ecliptic in the night sky. Figure 2 plots the absolute and

visual magnitude of each source, and there is more of a definitive difference in this graph.

Asteroids are shown to have higher absolute magnitudes, which indicates that they are usually dimmer than the non-asteroids. Finally, Figure 3 shows the difference between the magnitude of the objects in the two different wavelengths of infrared, indicating that the properties of the non-asteroids still lie within the area of the plotted asteroid data points.

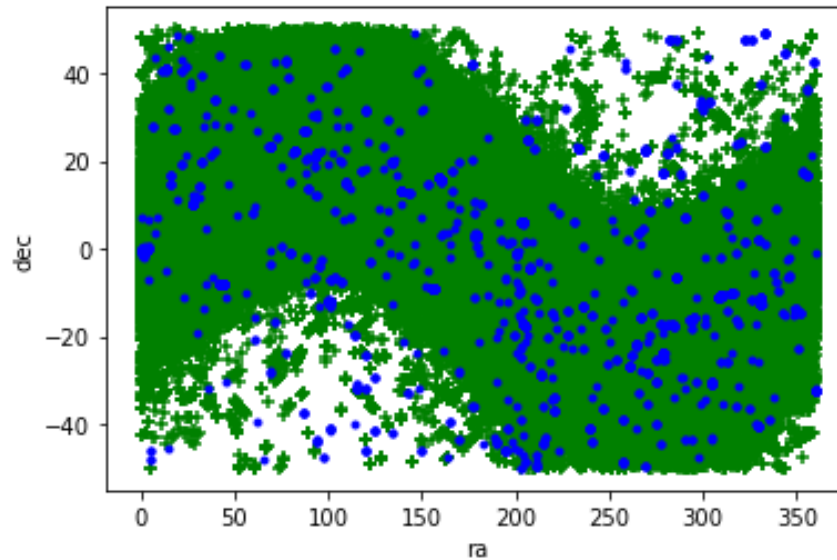


Figure 1. Right ascension vs declination for each source. The blue dots represent the non-asteroids, and the green dots represent the known asteroids. Both axes are in degrees showing right ascension from 0 to 360 degrees and declination from -50 to 50 degrees.

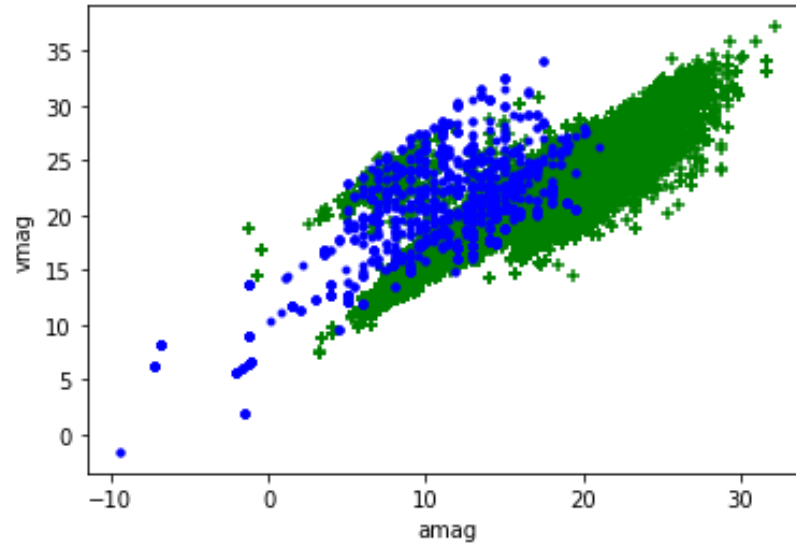


Figure 2. Absolute magnitude vs visual magnitude for each source. Blue dots represent non-asteroids, whereas green represent asteroids. Asteroids seem to have a higher absolute magnitude on average, but visual magnitude seems to be distributed evenly for both classes. This shows that asteroids are dimmer than other types of sources.

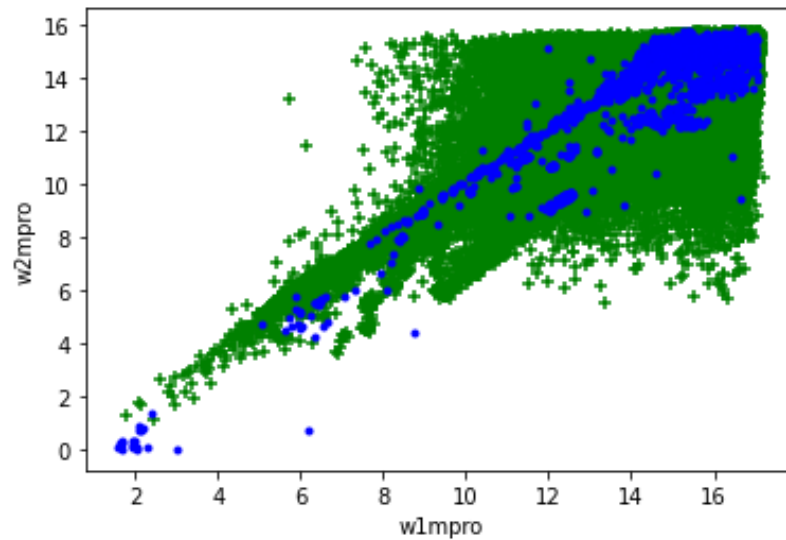


Figure 3. Wavelength 1 vs wavelength 2 for each source. Blue dots represent non-asteroids, whereas green represent asteroids. Most of the non-asteroids lie within a smaller region within the area of the asteroids.

Preprocessing Pipeline

Our image preprocessing pipeline started with the input of a 1016x1016 pixels grayscale image collected from the database with final goal of detecting the point sources directly on the image and comparing them with known sources collected from the IRSA database stored in the CSV file. If we are able to create a pipeline that can accurately detect the sources that are already in the IRSA database, we then can rely on that processor to find new data points that can be used to test other machine learning classifiers.

The first step of the pipeline was data conditioning. The data contained negative numbers representing broken pixels, so we limited the arrays to contain only positive numbers to allow for image visualization. The second portion of the conditioning was limiting the data to values of 0 to 255, the standard values used in image processing. We accomplished this by converting all the images to 8-bit integers (whole numbers ranging from 0 to 255). Next, we fed the images into a source detection algorithm called DAOSTarfinder [14]. This algorithm can detect stars and can also be used to detect other point-like sources. The sources can be detected by first calculating the median value of pixels, which is then assigned as the value of the background. Points that lie outside of the full width half maximum (FWHM) of this value are then flagged as sources. The last step of the pipeline was to add a threshold. We applied a threshold to pixels with values over five times the standard deviation. The output of this algorithm was a table with the x and y coordinates of each detected source on the image. At this point in the processing pipeline, our algorithm should have picked out all the point sources in the images. These could include stars, comets, asteroids, planets, and planetary satellites.

Finally, to compare the actual locations of the sources and the sources detected from the algorithm outlined above, we used the coordinates stored in each individual CSV file and

compared that with the closest point possible from the source table generated. This comparison was done by calculating the smallest Euclidian distance (the length of the line segment between the two points). The final output was an image with all the sources circled in blue and the sources of interest (asteroids, comets, planets, planetary satellites) circled in red, as well as the value of the Euclidian distance of the original point to the detected point provided. Figure 4 is a visual representation of this pipeline.

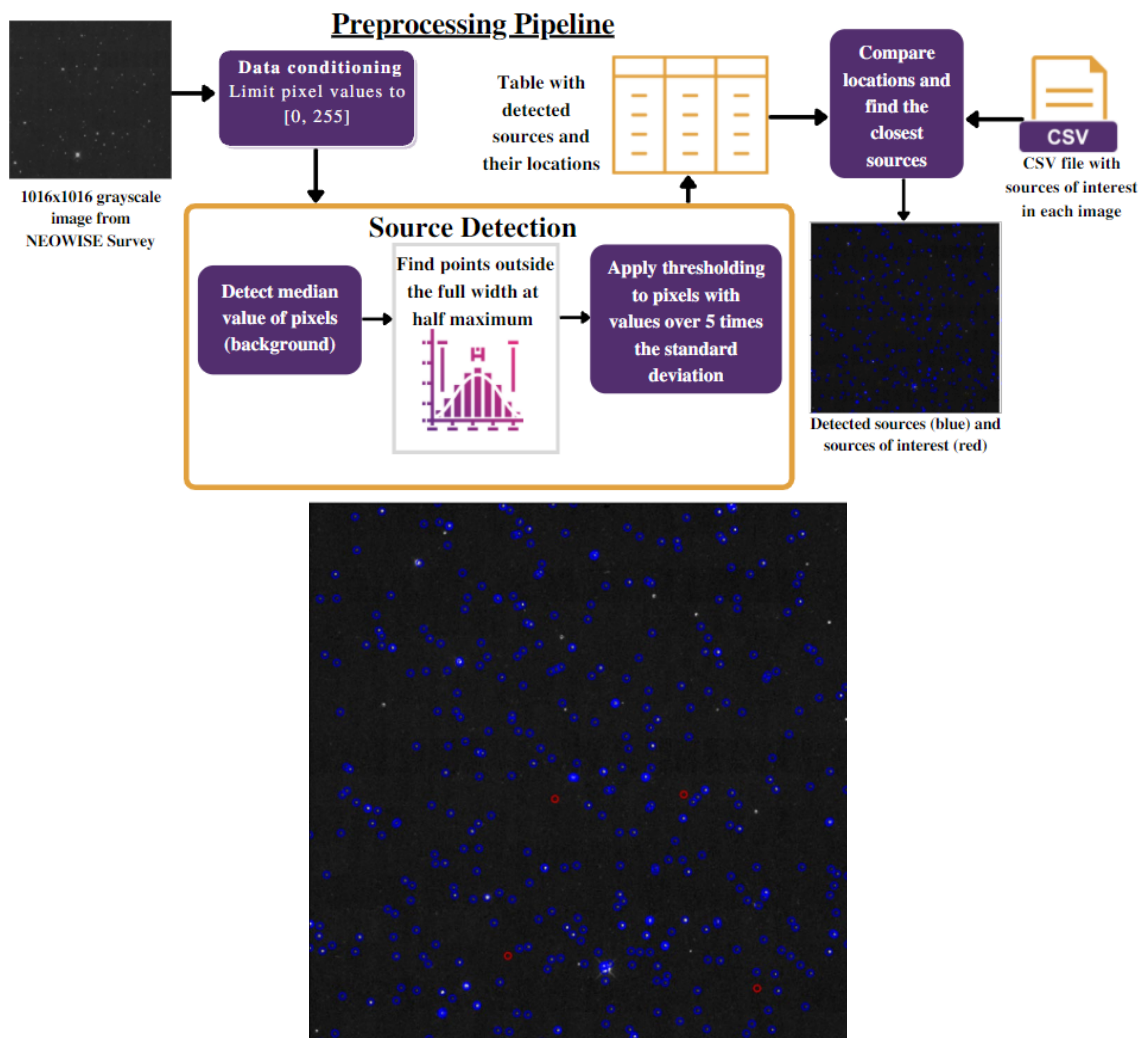


Figure 4. Preprocessing pipeline. This figure shows the preprocessing pipeline developed to prepare images for machine learning detection, including the inputs and outputs of the pipeline, as well as an example output images with the blue detected sources and red sources of interest.

Numerical Source Data Classification

The next step in our research was to test our data with the various types of machine learning algorithms discussed above to determine if any of these models were viable systems to classify our numerical data on each source. The first model we tested was logistic regression. To begin, we imported all of the known data for each source into a Pandas data frame. The imported CSV file included information on all 550,264 sources. To split the data into the training and testing sets we used the `train_test_split` function from `sklearn.model_selection` and set the test size to `.2` of the original data frame. So, the testing data is 20% of randomly selected data from the original dataset, while the rest of the data makes up the training data. We want the training data to be a majority of the available data because more data means more training which should lead to a more efficient classifier. The features included the 'ra', 'dec', 'amag', 'vmag', 'w1mpro', 'w1sgmp', 'w2mpro', 'w2sgmp', 'band_difference' columns, whereas the target only included the label column labeled 't'.

The collected NEOWISE data is heavily biased towards asteroids. There were 438,960 asteroids and only 1,251 non-asteroids in this training set. (This bias and its implication for the results are explained more in-depth in the discussion section of this thesis.) Finally, we trained and tested the data with the `LogisticRegression` model from Scikit Learn. The output is a confusion matrix described in the results section. We then followed a similar process for the SVM, naïve Bayes, and random forest classifiers. For each classifier we generated a training and testing dataset and then used the respective model from Scikit Learn to generate a confusion matrix.

The next step in our system was to use a convolutional neural network to classify the images. We anticipated using the probability from the CNN as well as the best machine learning

classifier in order to create a precise source classification system, but we were ultimately unable to perform the CNN on the images for reasons explained in the results section.

RESULTS

The preprocessing pipeline contained 548,700 asteroids, 1,331 comets, 35 planets, and 198 planetary satellites in the images used. The average distance estimated between the object source location and the actual location on the image was 26.76 pixels, or 73.6 arcseconds. The standard deviation was 29.62 pixels or 81.4 arcseconds. This indicates that our process was accurate and can continue to be used. Along with the estimated distances of the sources of interest we also gathered and analyzed information on all of the sources, summarized in Table 1.

	Right Ascension	Declination	Visual Mag.	Absolute Mag.	Band 1/Band 2 Difference
Asteroids					
Minimum	0.002	-49.99	-1.200	7.456	-7.593
Median	183.9	-2.080	15.13	20.13	0.911
Maximum	360.0	49.99	32.10	37.20	10.63
Comets					
Minimum	0.471	-49.83	2.000	9.579	-3.143
Median	192.6	-2.048	11.83	21.10	0.857
Maximum	359.6	49.00	21.00	34.07	7.208
Planets					
Minimum	0.626	-12.41	-9.400	-1.613	0.240
Median	168.9	-1.590	-6.045	5.720	1.009
Maximum	330.8	19.70	-1.52	8.151	2.010
Planetary Satellites					
Minimum	0.101	-12.44	-2.000	5.637	-0.712
Median	91.19	-0.473	6.498	15.83	0.788
Maximum	359.6	19.63	16.90	26.20	5.409

Table 1. Source Data Statistics. This table shows the different data associated with each source type. The minimum, medium, and maximum are reported for each feature.

Most machine learning models returned a confusion matrix. The total number of sources labeled true or false changes slightly for each model. The matrices for the various machine learning models are given in the figures that follow. These matrices show a lot of correctly labeled true cases (asteroids). However, even with the relatively high number of true positives there was still a considerable number of sources incorrectly labeled within the context of our data, explained in the discussion.

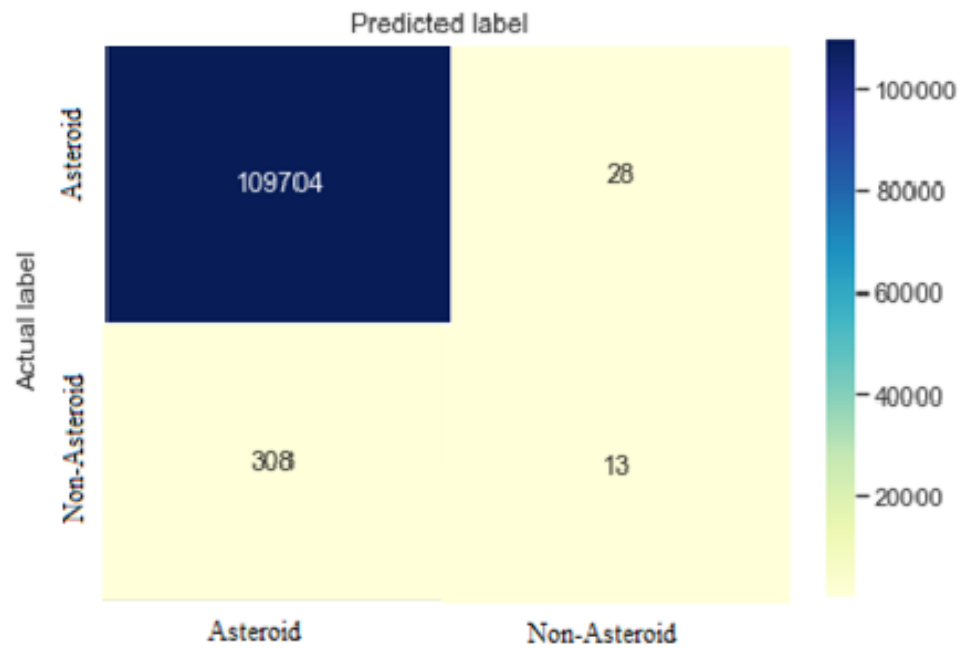


Figure 5. Logistic regression confusion matrix

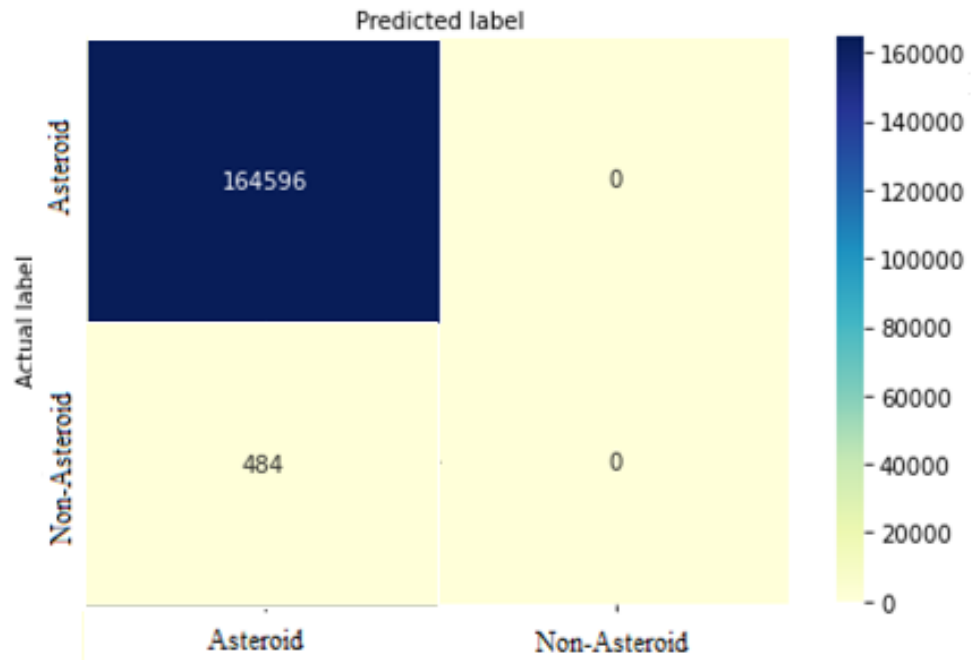


Figure 6. Support vector machine confusion matrix

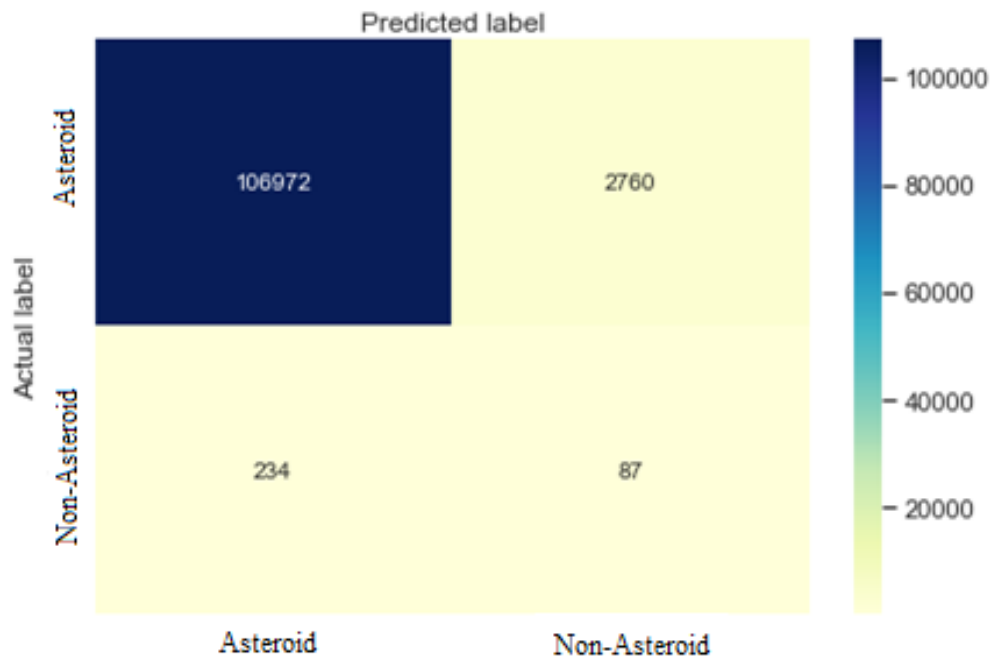


Figure 7. Naïve Bayes confusion matrix

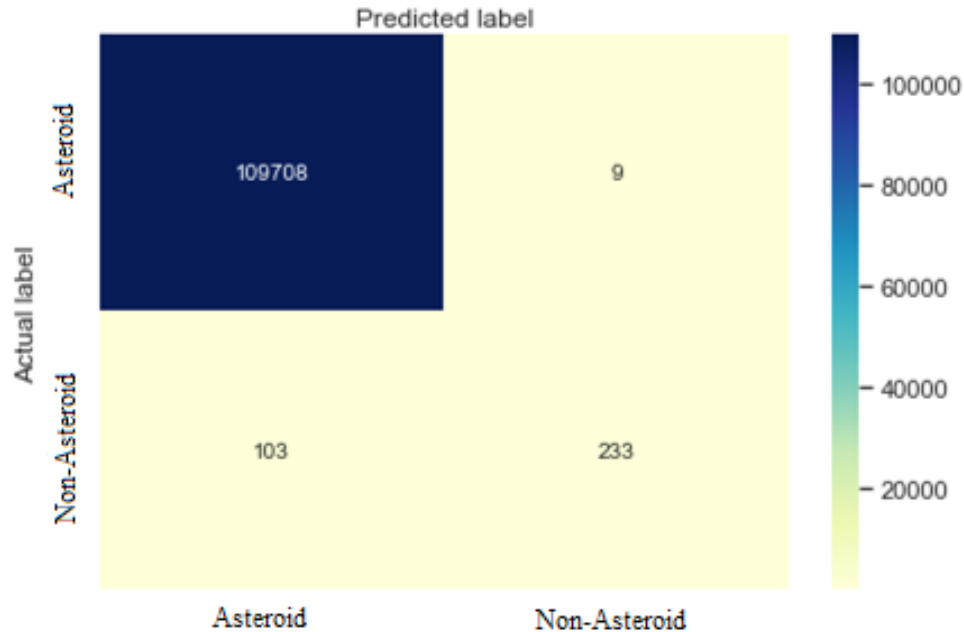


Figure 8. Random forest confusion matrix

DISCUSSION

Our analysis showed that the preprocessing pipeline utilized in this research was useful at finding sources, as the mean average of the distance between the real source coordinate and the detected source coordinate was around 73.6 arcseconds (around .02 degrees in the sky) or 26.76 pixels. Since we know the pipeline is able to detect the asteroids with some accuracy, we can be confident in using it to train future machine learning algorithms to detect other asteroids. However, there is still some work needed to ensure that more of the sources of interest are detected in the images. One possibility is making sure there are no anomalies, such as diffraction spikes or other lensing errors in the images, and no mistakes when collecting the coordinates of the sources from the database.

Another factor to note is that the NEOWISE survey data has a large bias towards asteroids, as there is a considerably larger amount of data collected for that source as compared

to the other types of objects. We can also use the data statistics previously outlined in Table 1 to determine certain parameters for common asteroids that should be used in training for machine learning and what the threshold might be for these parameters. For example, looking at the data collected in Table 1, there are some clear differences between asteroids and planets and planetary satellites, but the data is similar for asteroids and comets. This implies comets will be hardest to distinguished from asteroids in the images, as well as analyzing at the data.

We can see from the results of the different classification algorithms that there are a lot of true positives or true negatives, showing that each classifier was able to identify a majority of the asteroids. However, even though each confusion matrix showed a large percentage of asteroids identified, this does not necessarily mean these algorithms were successful. Because the training set has around 439,000 asteroids and only 1,300 non-asteroids, any algorithm had a good chance of labeling a source an asteroid simply due to the large amount of them. Because of this, we are more interested in the other boxes of the confusion matrix. There are non-asteroids on the order of 1000 instead of 100,000, so even a small number of non-asteroids incorrectly identified is an issue. In every model above there are a couple hundred, if not more, incorrectly identified. Adding the probability of the classified sources from a CNN used on the images would be able to lower the incorrectly classified sources, but there were issues on implementing this system.

For a CNN image classifier, a clear training set of images is essential. The detected source should be in the center of the image, and there should be no other objects pictured. In our data, after creating an image with the collected source coordinates at the center, the source of interest was not in the center. This meant there was some loss in precision converting the center of the object to degrees and then pixels on the image. At the moment we are unsure

where this loss occurs, since the coordinates of the object and the coordinates of the center of the image are pulled directly from the IRSA database. We converted both coordinates into the same unit and then found the difference of the coordinates of the source from the center and then created an image by drawing a 90-by-90-pixel box around the supposed location of the object. Without being able to center the image on the intended object, we cannot assure that only the object will appear in the image because we cannot locate where the object is exactly without looking at the image and zooming in. There are also a lot of background sources that would need to be removed if in the future we are to use a CNN with the images. Therefore, we cannot ensure an accurate system using a CNN, due to the error in our images. Some of these images are pictured in figure 9, there is a brighter object in each image, but it is not at the center.

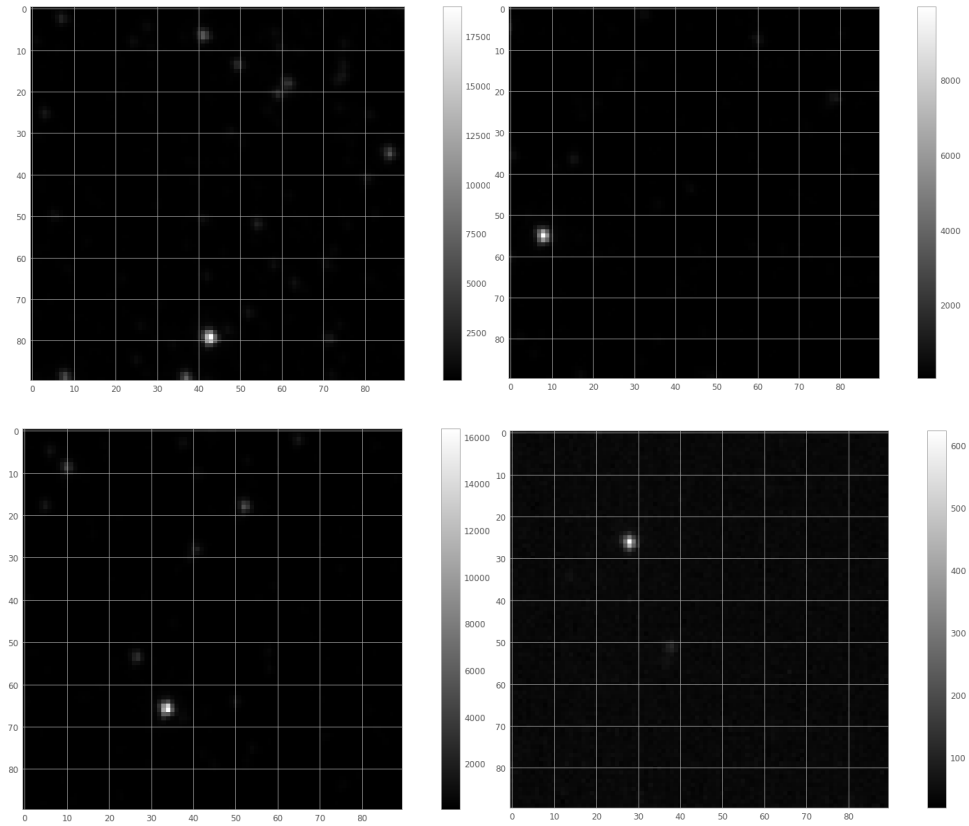


Figure 9. Collected images from the NEOWISE survey of known asteroids. Each image starts as a FITS file and is then depicted using matplotlib and astropy.visualization.

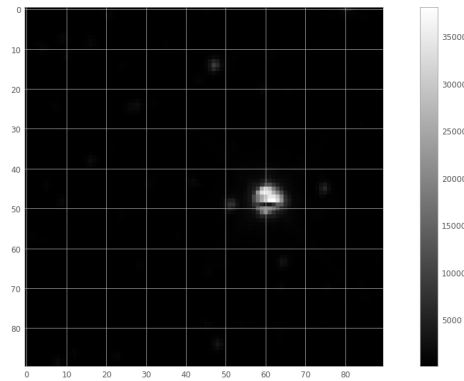


Figure 10. An additional image of a larger known asteroid. This image of a larger asteroid shows some of the details that can be captured within these images.

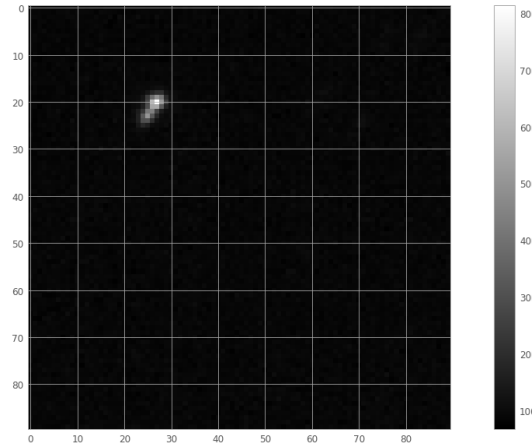


Figure 11. A NEOWISE image example of a non-asteroid. This image shows how a non-asteroid might look as compared to the previous images of asteroids.

To continue this research, we would need to figure out how to calibrate these images to work well with the CNN which would allow us to create a more in-depth classification system along with the numerical data analysis. We would first start by checking the coordinates from the original database, or we could use the detected centers used from before. But this brings up the issue of which is more accurate, the provided coordinates or the detected sources. The end goal of this research would have been to use the images to classify an asteroid in figure 9 and 10 and a non-asteroid which is shown in Figure 11, where you can see there is a difference between that image and the previous ones. Future researchers could combine the probability of the correct class from the image CNN classifier with one of the other machine learning algorithms used on the numerical data to create a more precise classifier that would be useful on other NEOWISE source and image data. Based on our collected results from each of the machine learning classifiers, that there is not one that outperforms the others significantly and that can confidently be chosen to be part of the combined system. For future development of this research there should be more investigation into these different classifiers as well as other

options not explored in this research. This combined system would then be able to detect the smaller and slower-moving asteroids with the help of the preprocessing steps as well.

REFERENCES

- [1] Ruprecht, J., "Asteroid Detection Results Using the Space Surveillance Telescope",
in *Advanced Maui Optical and Space Surveillance Technologies Conference*, 2015.
- [2] D. A. Duvvuri et al., "DeepStreaks: Identifying Fast-Moving Objects in the Zwicky Transient Facility Data with Deep Learning," *Mon. Not. R. Astron. Soc.*, vol. 486, no. 3, pp. 4158–4165, 2019, doi: 10.1093/mnras/stz1096.
- [3] A.C. Rabeendran and L. Denneau, "A Two-Stage Deep Learning Detection Classifier for the Atlas Asteroid Survey", *Publ. Astron. Soc. Pacific*, vol. 133, no. 1021, 2021, doi: 10.1088/1538-3873/abc900.
- [4] C. J. Fluke and C. Jacobs, "Surveying the Reach and Maturity of Machine Learning and Artificial Intelligence in Astronomy," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 10, no. 2, pp. 1–40, 2020, doi: 10.1002/widm.1349.
- [5] J. Kremer, K. Stensbo-Smidt, F. Gieseke, K. S. Pedersen, and C. Igel, "Big Universe, Big Data: Machine Learning and Image Analysis for Astronomy," *IEEE Intell. Syst.*, vol. 32, no. 2, pp. 16–22, 2017, doi: 10.1109/MIS.2017.40.
- [6] J. R. Masiero, A. K. Mainzer, J. M. Bauer, R. M. Cutri, T. Grav, E. Kramer, J. Pittichová, and E. L. Wright, "Asteroid Diameters and Albedos from NEOWISE Reactivation Mission Years Six and Seven," *The Planetary Science Journal*, vol. 2, no. 4, p. 162, 2021.
- [7] A. Mahabal et al., "Machine Learning for the Zwicky Transient Facility," *Publ. Astron. Soc. Pacific*, vol. 131, no. 997, p. 38002, 2019, doi: 10.1088/1538-3873/aaf3fa.
- [8] E. A. Smirnov and A. B. Markov, "Identification of Asteroids Trapped Inside Three-Body Mean Motion Resonances: A Machine-Learning Approach," *Mon. Not. R. Astron. Soc.*, vol. 469, no. 2, pp. 2024–2031, 2017, doi: 10.1093/mnras/stx999.

- [9] J. D. Ruprecht, H. E. M. Vigh, J. Varey, and M. E. Cornell, "SST Asteroid Search Performance 2014-2017," *IEEE Aerosp. Conf. Proc.*, vol. 2018-March, pp. 1–8, 2018, doi: 10.1109/AERO.2018.8396388.
- [10] R. L. Jones et al., "The Large Synoptic Survey Telescope as a Near-Earth Object Discovery Machine," *Icarus*, vol. 303, pp. 181–202, 2018, doi: 10.1016/j.icarus.2017.11.033.
- [11] M. Lieu, L. Conversi, B. Altieri, and B. Carry, "Detecting Solar System Objects with Convolutional Neural Networks," *Mon. Not. R. Astron. Soc.*, vol. 485, no. 4, pp. 5831–5842, 2019, doi: 10.1093/mnras/stz761.
- [12] P. Vereš, R. Jedicke, L. Denneau, R. Wainscoat, M. J. Holman, and H.-W. Lin, "Improved Asteroid Astrometry and Photometry with Trail Fitting," *Publ. Astron. Soc. Pacific*, vol. 124, no. 921, pp. 1197–1207, 2012, doi: 10.1086/668616.
- [13] A. Waszczak et al., "Small Near-Earth Asteroids in the Palomar Transient Factory Survey: A Real-Time Streak-Detection System," *Publ. Astron. Soc. Pacific*, vol. 129, no. 973, 2017, doi: 10.1088/1538-3873/129/973/034402
- [14] Stetson, P. B., "DAOPHOT: A Computer Program for Crowded-Field Stellar Photometry", *Publ. Astron. Soc. Pacific*, vol. 99, p. 191, 1987. doi:10.1086/131977.