# THE LARGEST CATALOG OF COSMIC EXPLOSIONS

HARLAN T. PHILLIPS[1,2], DR. ANNA Y. Q. HO[1], DR. DANIEL PERLEY[3]

1. CORNELL UNIVERISTY 2. UNIVERSITY OF CALIFORNIA, BERKELEY 3. LIVERPOOL JOHN MOORES UNIVERSITY

## ZWICKY TRANSIENT FACILITY

**Figure 1:** Palomar Observatory, California Institute of Technology

Optical transients are brief astronomical phenomena, such as supernovae. Our understanding of stellar deaths and other phenomena was historically limited, but technology and optimized telescope surveys have greatly expanded our knowledge. Robotics and automated detection have revolutionized the field. Facilities such as the Zwicky Transient Facility (ZTF) now survey the northern sky in two days. In 2010, over 3000 optical transients were discovered; today, the biggest catalogs have thousands of objects ($10^3$). This project will produce the highest magnitude ($10^4$) of objects in one database, setting the tone for catalogs.

## ZTF'S CATALOG

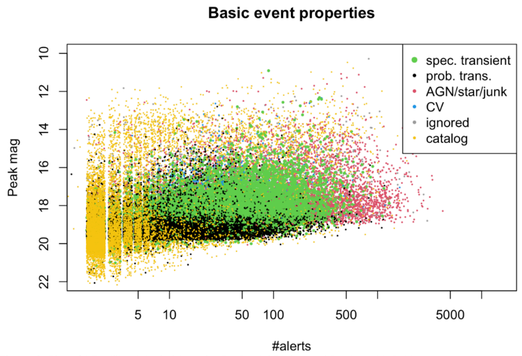



**Figure 2:** Plot of 1.5 million initial objects by Dr. Daniel Perley and Dr. Anna Y. Q. Ho

- Started with over 1.5 million ZTF objects
- The initial steps involved filtering objects to exclude those far from the Milky Way, noise (fake detections), fainter than close objects, and those that became brighter.
- The primary filtering steps further refined the dataset by focusing on color and brightness characteristics. During this phase, we removed known asteroids, stars, active galactic nuclei (AGNs), and any previously discovered sources.
- After these steps, we successfully filtered down to over 250,000 transients.

## ENGINEERING A PIPELINE FOR CLASSIFYING OVER 250,000 OBJECTS

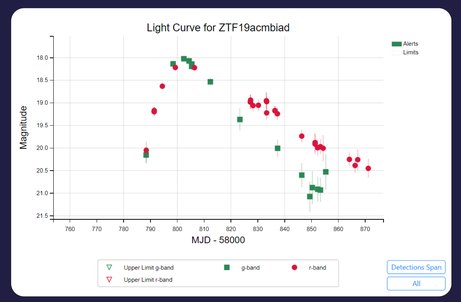



**Figure 3:** Supernova Ia light curve generated from the classification tool. Displays the brightness over time in different color-bands.

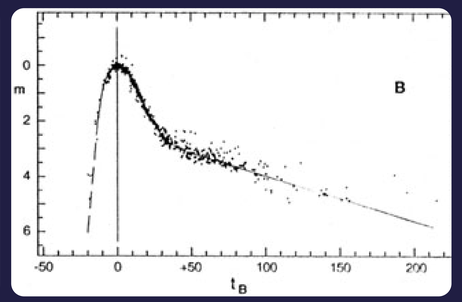



**Figure 4:** Supernova Ia Light Curve (Cadonau, R. 1987. Ph.D thesis. Univ. Basel)

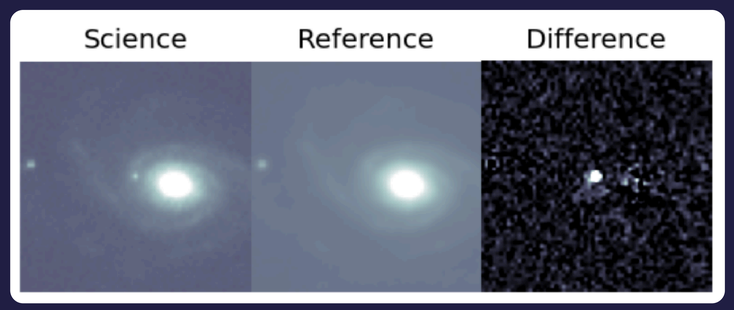



**Figure 5:** Science image is recent photo generated by ZTF, reference is a photo taken much earlier. The difference is the subtraction of pixels between images. If there is a bright pixel in the center, means something new appeared.

## OVERVIEW & FEATURES

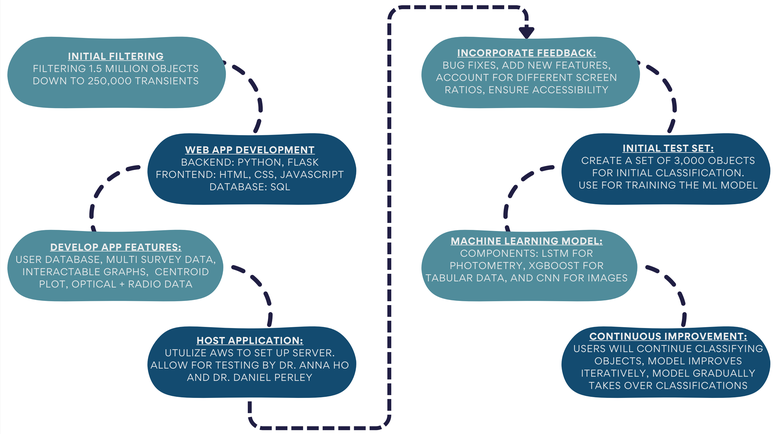
This classification tool was developed for the scientific community to classify objects in the Zwicky Transient Facility (ZTF) catalog. It will be utilized by over 100 researchers and astronomers. The application incorporates unique features and data sources to enhance the classification process. It includes multiple plots and integrates data from various telescopes, ensuring the most recent and comprehensive information is available. Both optical and radio data are included, along with interactive graphs to facilitate deeper analysis and understanding.

## FUTURE PLANS

Current results of the machine learning model are based on only 400 samples of very similar supernova classes, making distinctions nearly impossible. The plan is to provide accessibility of the tool for various scientific groups, with over 100 users. After classifying a subset of 3,000 objects, this data will train the machine learning model, which will adapt continuously based on new classifications. Once the model achieves around 95% accuracy, automated classifications will be implemented.

This will provide access to hundreds of thousands of transients, enabling extensive statistical analysis and new discoveries, ultimately leading to a broader understanding of our universe.

## METHODOLOGY




## CURRENT ML MODEL

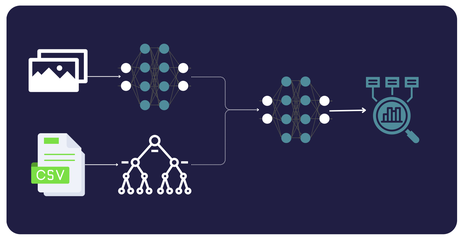



**Figure 6: Current Architecture of Machine Learning Model**
Figure 6 illustrates the machine learning model architecture, which integrates tabular, photometric, and image data to predict classifications. The process begins with tabular and photometric data being inputted into an XGBoost model, known for its effectiveness in handling structured data. Concurrently, image data is processed by a convolutional neural network (CNN), which excels at analyzing visual information. The outputs from the XGBoost model and the CNN are then combined and fed into another CNN, which processes this integrated data to generate final predictions. This multi-stage approach leverages the strengths of both data types, ensuring accurate and reliable classifications.

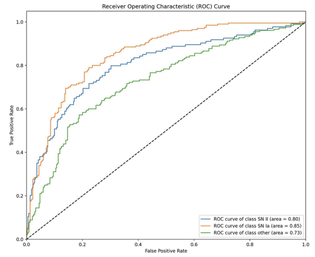



**Figure 7: ROC Curve**
The Receiver Operating Characteristic (ROC) curve demonstrates the model's performance across various threshold settings, illustrating the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) for each class. A higher area under the curve (AUC) signifies better model performance, indicating its ability to distinguish between SN Ia, SN II, and Other objects.

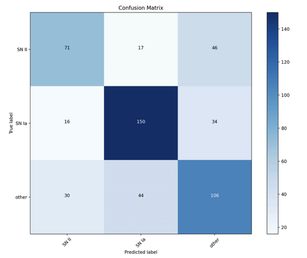



**Figure 8: Confusion Matrix**
The confusion matrix provides a detailed comparison of the model's predictions versus actual classifications. It shows the counts of true positives, true negatives, false positives, and false negatives for each class. This matrix highlights the model's accuracy and misclassification rates, identifying areas of strength and potential improvement.

**ACKNOWLEDGEMENTS:** Harlan Phillips participated in the summer 2024 REU program at the Cornell Center for Astrophysics and Planetary Science at Cornell University. This material is based upon work supported by the National Science Foundation under award NST/AST- 2244064. We would like to acknowledge Dr. Daniel Perley and Dr. Yujing Qin for their collaborative efforts and constructive feedback on the project.
Image attributions: NASA, Arnaud Girault (Unsplash), Palomar Observatory (California Institute of Technology)