# Classifications of unknown transients using ParSNIP

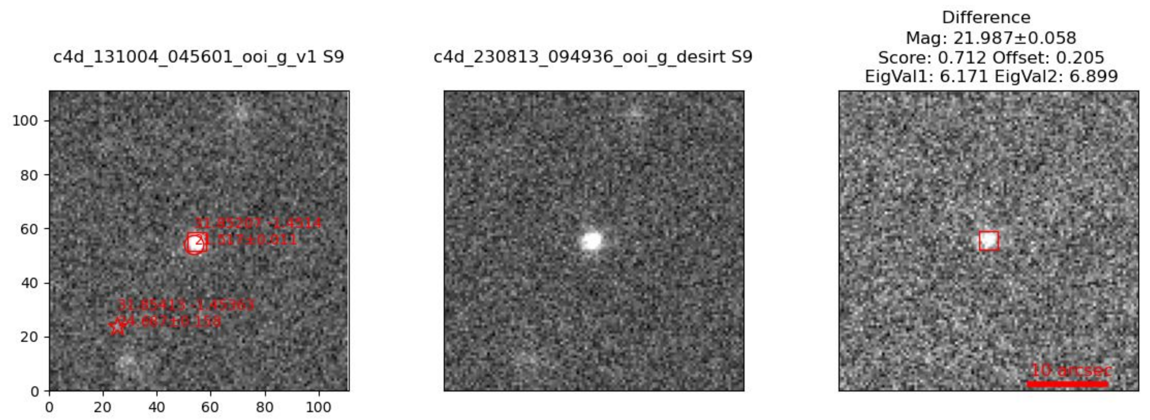
Jack Collard, Yaamini Jois, Nathan Olson, Orestis Sabetai, Xavier Silva, Mathew Thomas, Mentor: Emilie Cote

University of California, Berkeley    Undergraduate Lab at Berkeley, Physics & Astronomy Division

## BACKGROUND

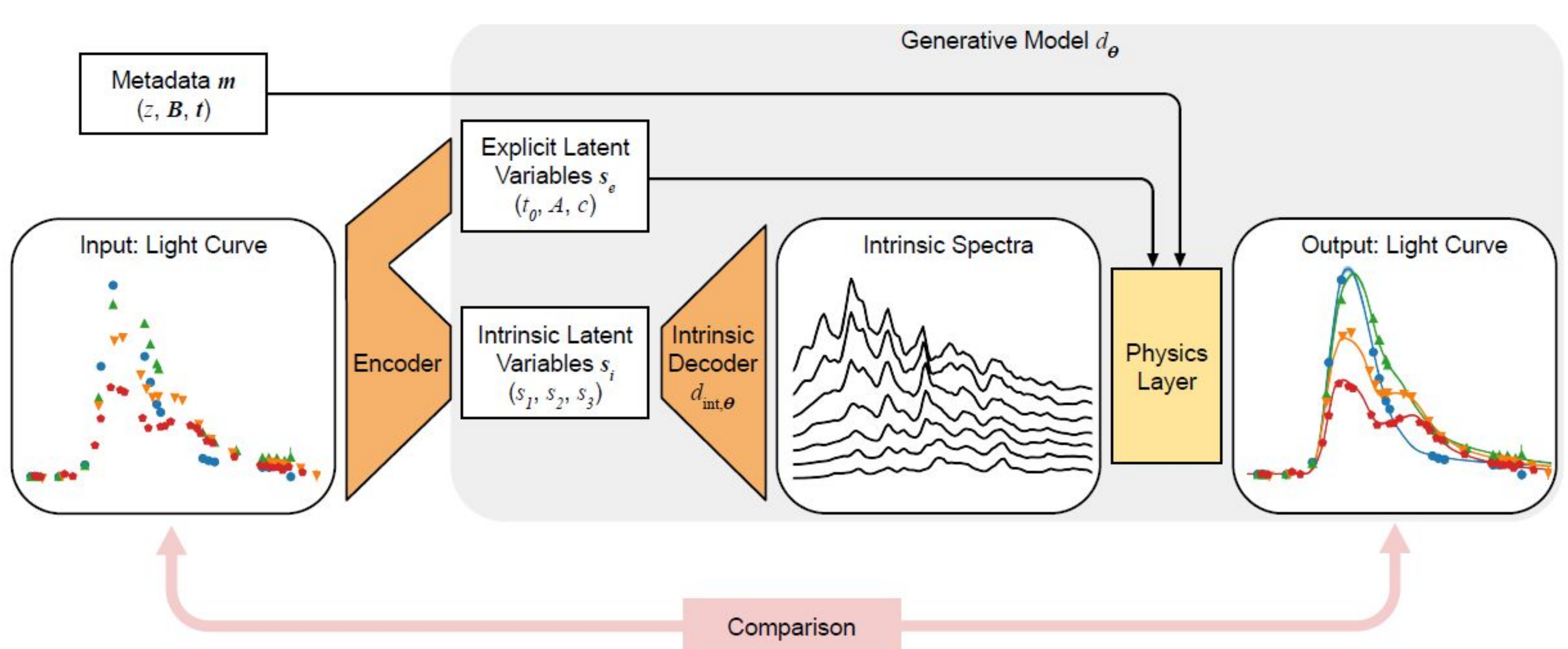
The term “transients” encompasses a variety of galactic and extragalactic events, as they are astronomical phenomena with durations ranging from a fraction of a second to years. The classification of transients requires sorting through lots of complex and unlabeled data, leaving room for human error. Traditionally, classification is done through tedious spectroscopic observations, but this isn’t feasible with how many transients we can observe now. Dimmer transients are also much harder to classify by hand. ParSNIP (Parametrization of Supernova Intrinsic Properties), is a generative neural network that generates models that simplify the classification of transients and increase classification accuracy without the need for labeled data. The received data can change in value depending on how it was taken even for identical transient types, and eliminating the influence of external factors can be done efficiently with ParSNIP. In this way, dimmer transients can also be classified without traditional spectroscopy. ParSNIP trains itself using the simulated datasets we feed it to improve its accuracy for predictions.

**One of over 50 observations for Example 1 light curve, SNII object**




## METHODS – PARSNIP

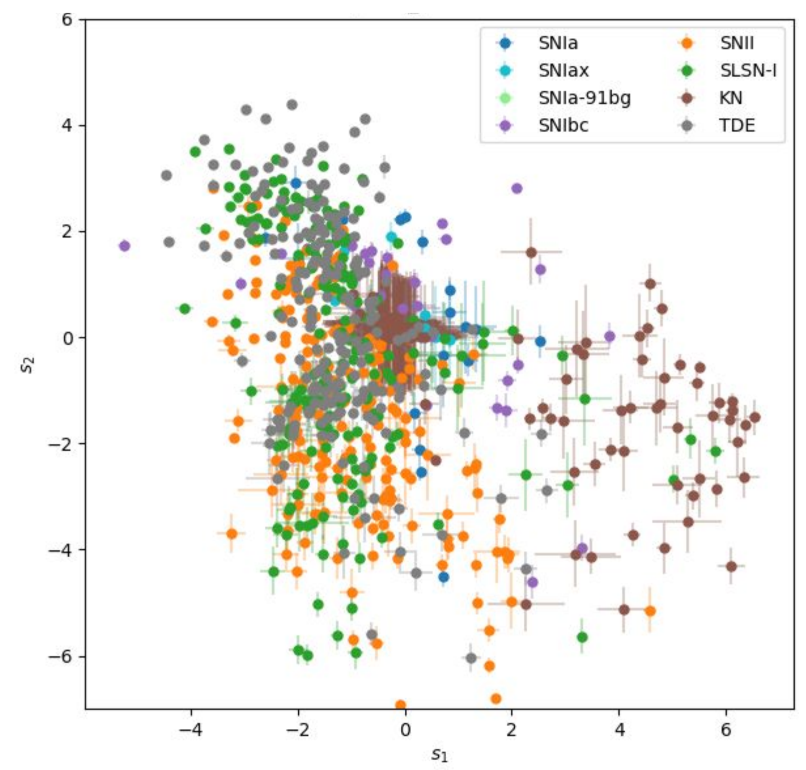
ParSNIP uses a variational autoencoder, so it takes the data we input, compresses it, and places it in a latent vector space composed of latent variables. The network groups similar characteristics in inputted data together by physically placing the latent variables - such as different traits of the transient data - close together in the latent space. ParSNIP especially has an additional layer that allows it to filter out the unphysical - external factor - traits of the data - which is why it doesn’t need labeled data. Traits such as: bandpasses, redshifts, and how the telescope took the data. Our job is to try to produce a better grouping of the latent variables than seen in previously produced latent spaces because that would translate to similar characteristics of a transient being accurately predicted/classified for future datasets. ParSNIP produces generative models of transients from datasets of unlabeled light curves. The generative neural network is trained on labeled data where each transient is already correctly classified beforehand, so that it can generate predictions for the unlabeled data. The data surveys that we use generate a vast amount of data, from the measurements of light intensity at different wavelengths (photometric data) to spectral information (spectroscopic data). The surveys where photometric classifiers have been applied have a large amount of data and they are expected to detect a large number of transients.




**Boone 2021**

## DATA

We obtained data from the DECam joint with LSST and DESI in order to make predictions and classifications with ParSNIP. Raw data had to be molded into lcdata format for project use, forming meta tables with object IDs and uniform observation bands (r, g, z). Over the course of the project we worked with multiple datasets, some including and some excluding transients with no spectroscopic redshift measurement.

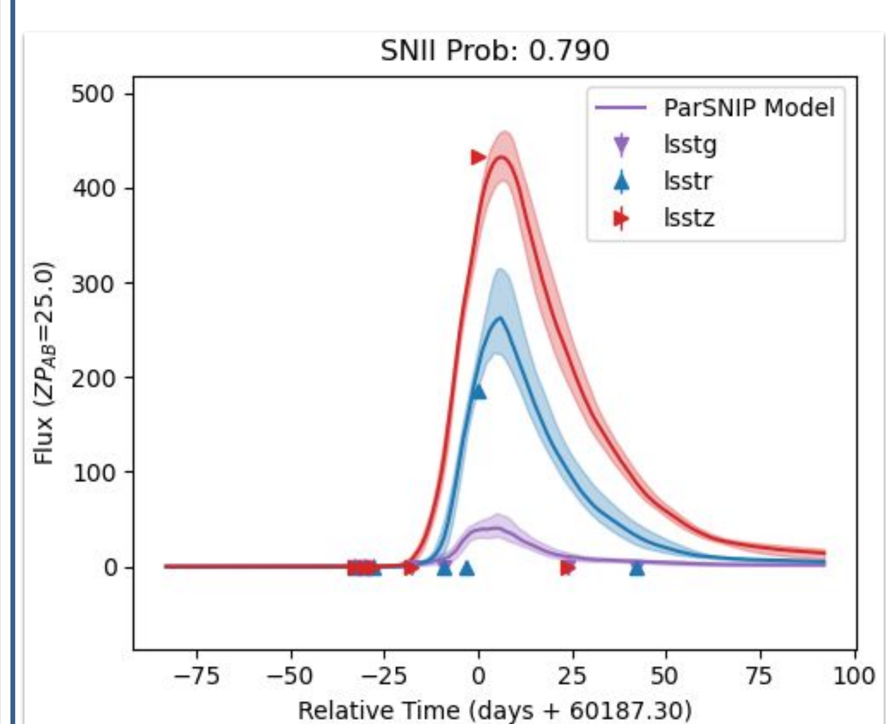



**PLAsTiCC Latent Space**

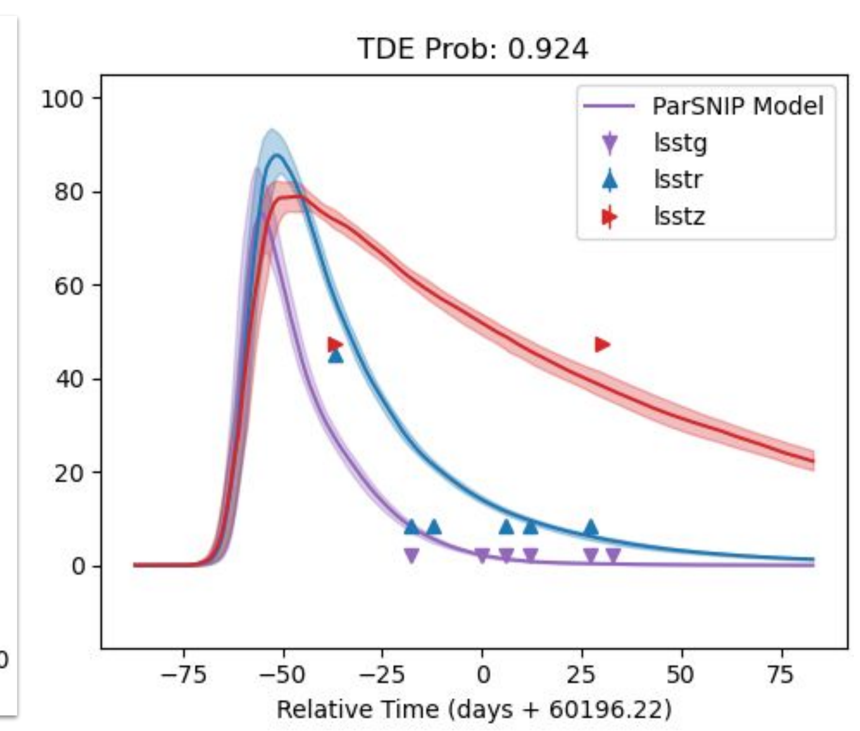
ParSNIP was trained on labeled data from two sources: first, from simulated light curve data done by PLAsTiCC (Photometric LSST Astronomical Time-Series Classification Challenge) to mimic data from the LSST at Vera C. Rubin Observatory, and second, from real data taken by Pan-STARRS1 (PS1). Thus we had two separate models with which to make predictions and classifications.

## RESULTS

The results we obtained are an outcome of using a built-in ParSNIP function that takes a data file and a model and uses the data points to plot a lightcurve that best fits the data it was given. This lightcurve, labeled ‘model’, then corresponds to the characteristic lightcurve of a type of transient, and hence the transient from the data can be classified. Below, figure (insert number) shows an example of an acceptable lightcurve, where the model fits the data points well, and the error (shown by the faded color) is minimal and fits closely to the model line. Figure (insert number) shows a lightcurve that displays a limitation in the data we have: often, the flux stays consistent over time, raising questions about whether the observed “transient” is a transient at all. We speculate that it could be background noise. Following this limitation, the model produces a curve that tries to fit a horizontal line, increasing its chances of being inaccurate, despite a high TDE probability.




**Example 1**




**Example 2**

## RESULTS

We used four datasets each with a varying number of transients and redshifts to produce multiple stats tables. The following stats tables contained the least possible error values and came from a dataset that contained 446 transients of both photometric and spectroscopic redshifts. Each classification is associated with a certain possibility of accuracy, along with the number of objects that are predicted to have at least an 85% possibility of being a certain transient. This analysis is also how we can catch potential errors, as certain transients should not, theoretically, be as present as ParSNIP may predict it to be.

| Type | Total | Objects with prob > 0.85 | Highest Prob | Lowest Prob | Mean Prob | Median Prob | Max Data pts | Min Data pts | Ave Data pts |
|---|---|---|---|---|---|---|---|---|---|
| str7 | int64 | int64 | float64 | float64 | float64 | float64 | int64 | int64 | int64 |
| KN | 56 | 17 | 0.969 | 0.35 | 0.724 | 0.758 | 23 | 2 | 10 |
| SLSN-I | 259 | 178 | 0.998 | 0.327 | 0.841 | 0.925 | 40 | 2 | 16 |
| SNII | 37 | 2 | 0.96 | 0.273 | 0.518 | 0.503 | 39 | 10 | 16 |
| SNIa | 9 | 1 | 0.92 | 0.324 | 0.648 | 0.732 | 37 | 4 | 17 |
| SNIa-91 | 1 | 0 | 0.26 | 0.26 | 0.26 | 0.26 | 13 | 13 | 13 |
| SNIax | 2 | 0 | 0.629 | 0.359 | 0.494 | 0.494 | 22 | 18 | 20 |
| SNIbc | 20 | 4 | 0.917 | 0.269 | 0.637 | 0.625 | 32 | 8 | 16 |
| TDE | 62 | 32 | 0.994 | 0.26 | 0.765 | 0.858 | 36 | 4 | 17 |

**PS1 Results**

| Type | Total | Objects with prob > 0.85 | Highest Prob | Lowest Prob | Mean Prob | Median Prob | Max Data pts | Min Data pts | Ave Data pts |
|---|---|---|---|---|---|---|---|---|---|
| str7 | int64 | int64 | float64 | float64 | float64 | float64 | int64 | int64 | int64 |
| KN | 110 | 19 | 0.998 | 0.271 | 0.702 | 0.718 | 40 | 2 | 13 |
| SLSN-I | 130 | 64 | 0.999 | 0.354 | 0.788 | 0.842 | 40 | 2 | 17 |
| SNII | 96 | 34 | 0.986 | 0.206 | 0.73 | 0.751 | 39 | 6 | 17 |
| SNIa | 14 | 3 | 0.955 | 0.403 | 0.626 | 0.552 | 20 | 4 | 15 |
| SNIa-91 | 18 | 6 | 0.978 | 0.299 | 0.668 | 0.614 | 23 | 11 | 16 |
| SNIax | 11 | 1 | 0.892 | 0.489 | 0.688 | 0.738 | 20 | 6 | 14 |
| SNIbc | 19 | 4 | 0.978 | 0.282 | 0.634 | 0.643 | 20 | 9 | 16 |
| TDE | 48 | 25 | 0.998 | 0.354 | 0.802 | 0.858 | 37 | 4 | 17 |

**PLAsTiCC Results**

## CONCLUSION

Over the course of an academic year, we accomplished the broader goals of successfully running ParSNIP to produce light curves, stats tables, and latent spaces as an extension and hopeful improvement to the older version of the project. However, we were unable to achieve the high accuracy we aimed for within our predictions. This occurred as a result of our group and resources not being able to distinguish why our data showed magnitudes that did not match up with the predicted light curves that ParSNIP generated. There is also the possibility that ParSNIP itself was not properly trained on the datasets that we were drawing from (Dark Energy Camera (DECam) Legacy Survey). Most of our efforts this year have gone into working with the data to get it into the required format for ParSNIP to use, but due to our lack of time to run enough code to properly train our model and figure out the individual bands within each dataset, we were not able to properly identify why the data points were plotted so inconsistently. Since there is room for future improvement and continuation of this project, we received access to more datasets in the final week of our presentation as well as proposals for side projects to utilize the datasets.

## ACKNOWLEDGEMENTS

We would like to thank ULAB Physics and Astronomy for supporting our project: Directors Rav Kaur and Saahit Mogan, Lab Manager Jordan Duan, and Faculty Advisor Professor Dan Kasen. We would also like to thank the Legacy Survey Project and the collaboration between the Dark Energy Camera and DESI for sourcing all the datasets we used throughout this project.

## REFERENCES

1. Kyle Boone (2021). “ParSNIP: Generative Models of Transient Light Curves with Physics-enabled Deep Learning”
2. Young Supernova Experiment(2022). “Experiment Data Release 1 (YSE DR1): Light Curves and Photometric Classification of 1975 Supernovae