# Using Machine Learning to Identify Astrophysical Transients in the DESI Survey


Vashisth Tiwari[1], Amanda Wasserman[1], Segev BenZvi[1], DESI Collaboration[2]

Email: vtiwari2@u.rochester.edu

[1]Department of Physics and Astronomy, University of Rochester, Rochester, NY, USA

[2]http://desi.lbl.gov/


## About DESI

During the next five years, the Dark Energy Spectroscopic Instrument (DESI) will carry out a massive redshift survey of 35 million galaxies and quasars, mapping the large scale structure of the universe out to a redshift of 3. During the survey, we expect that many of these objects will contain bright transients such as supernovae (SN), tidal disruption events, and compact binaries that contaminate the spectra of the host galaxies.

## Identifying Supernovae

The identification of transients is important not only to ensure correct estimates of the host redshifts, but also for providing an opportunity to obtain "serendipitous" spectra of the transients themselves. Spectroscopic classification is the "gold standard" in categorization of transients, making these discoveries invaluable when combined with data from large photometric surveys. We have developed ML tools to identify and classify transients in galaxy spectra measured with DESI.

## Analysis

### Traditional Supernoava Classification

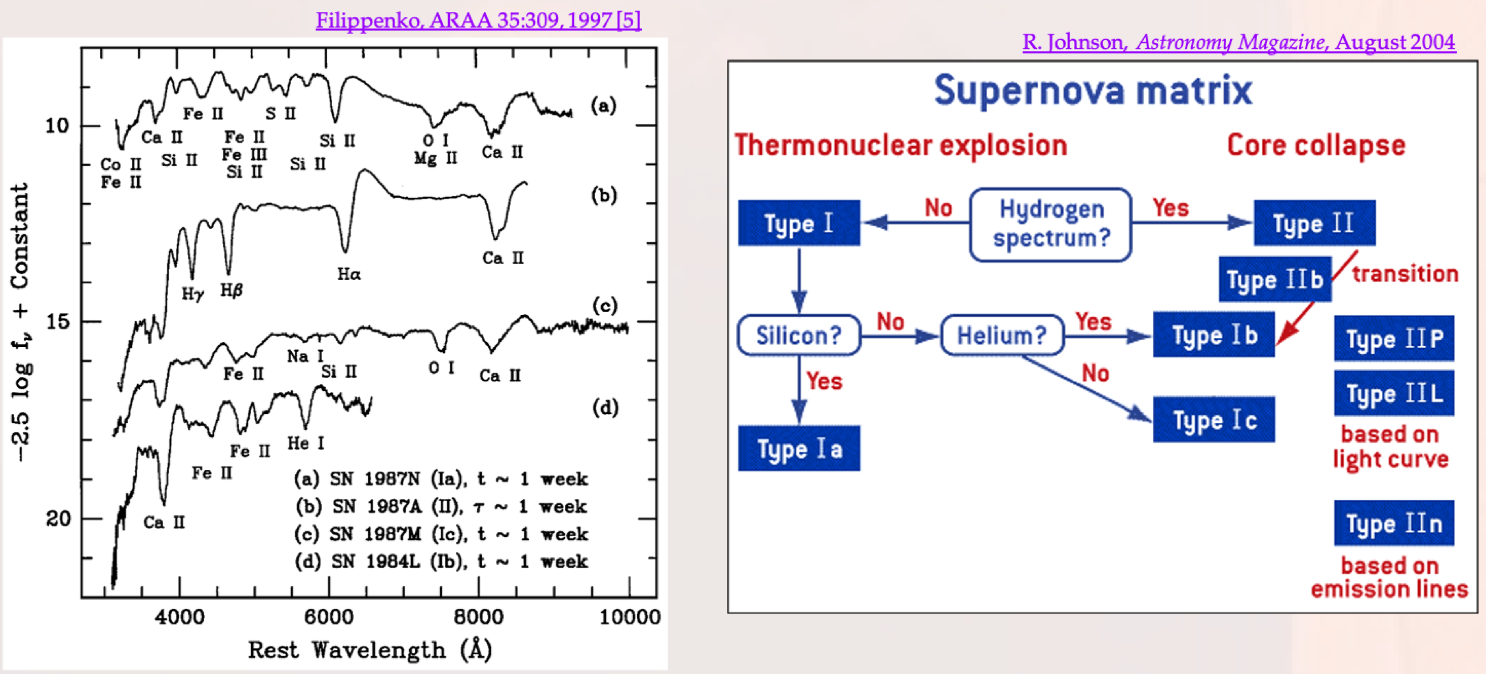


Figure 1: Traditional classification based on specific spectral features [1, 2].

### Simulation of data

- We simulated DESI Bright Galaxy Survey hosts and various types of supernovae.
- Uses bright galaxy mocks, BGS exposure, and observing conditions.

### Pre-processing:

- Weighted re-binning of simulations decimates data from 6000 to 150 points.
- Clip negative flux values which come from subtracting the background.
- Normalize spectrum between [0,1] to condition data and to reduce training time.
- De-redshift to view spectra in rest frame.

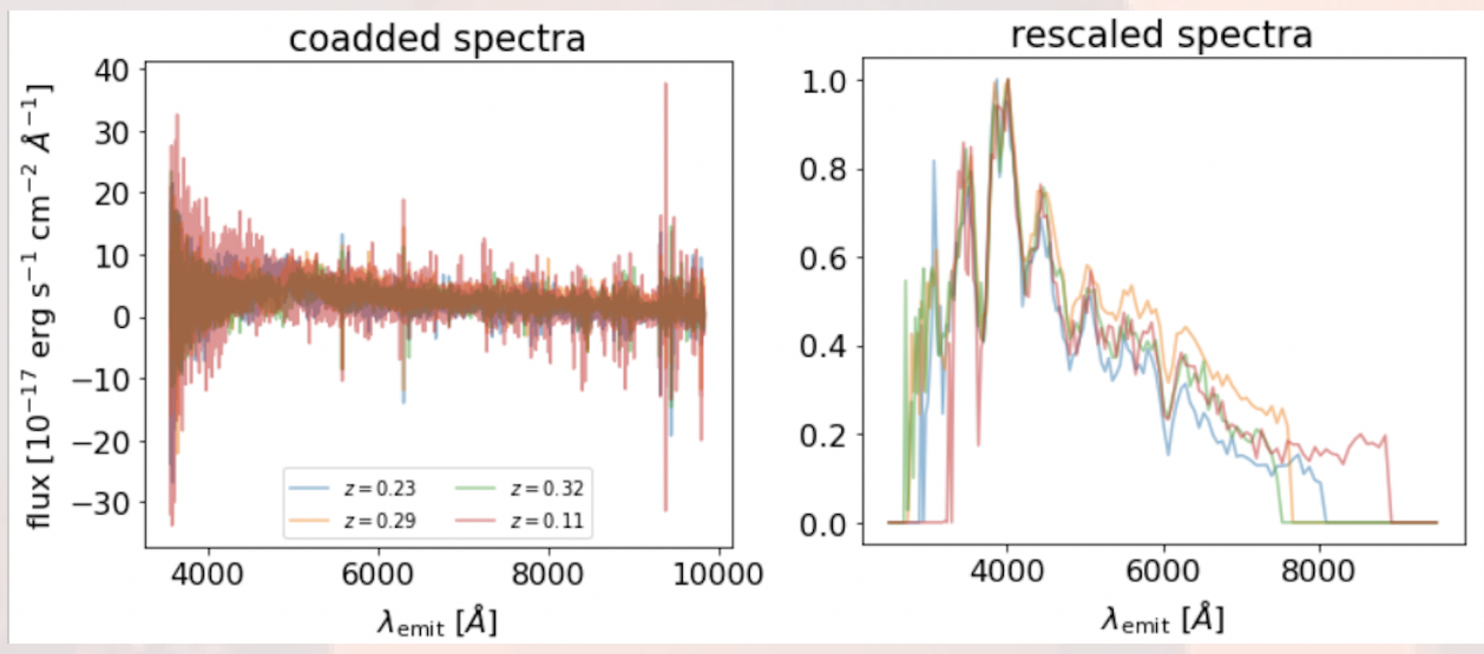


Figure 2: Spectrum before (grey) and after (red) preprocessing

### Classifier: Convolutional Neural Network (CNN)

- CNN consists of 4 convolutional layers, 1 dense layer, 1 output.
- Trained on 35,000, tested on 12,500, validated on 12,500 simulated spectra.
- 150 features in input layer, output 7 different classes (6 types of SN, hosts).

## Acknowledgments


We acknowledge support from the DOE Office of High Energy Physics.


## Results

### Accuracy

- For full sample, classifier is approximately 75% accurate.
- For hard cut, where classifier claims 99% probability of spectra being a transient, classifier is above 99% accurate

### Confusion Matrices

- The diagonal the better

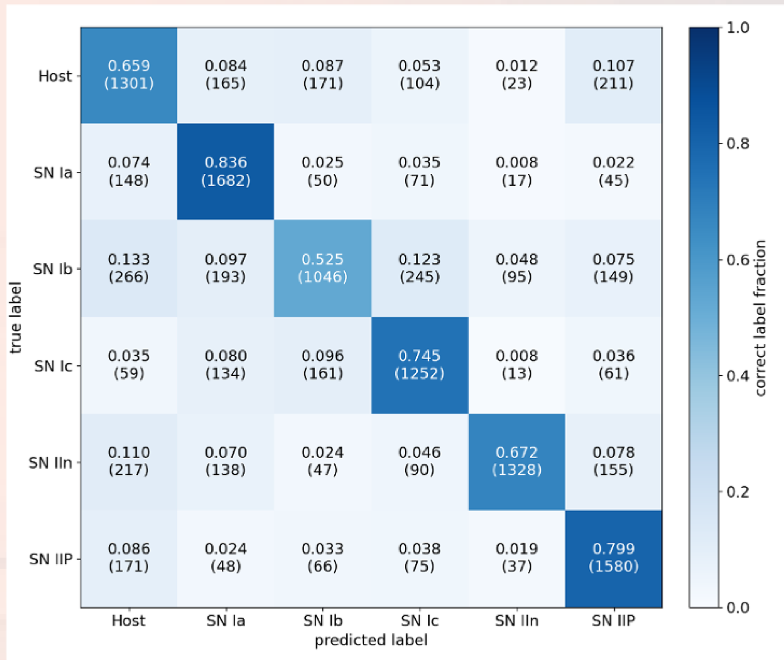
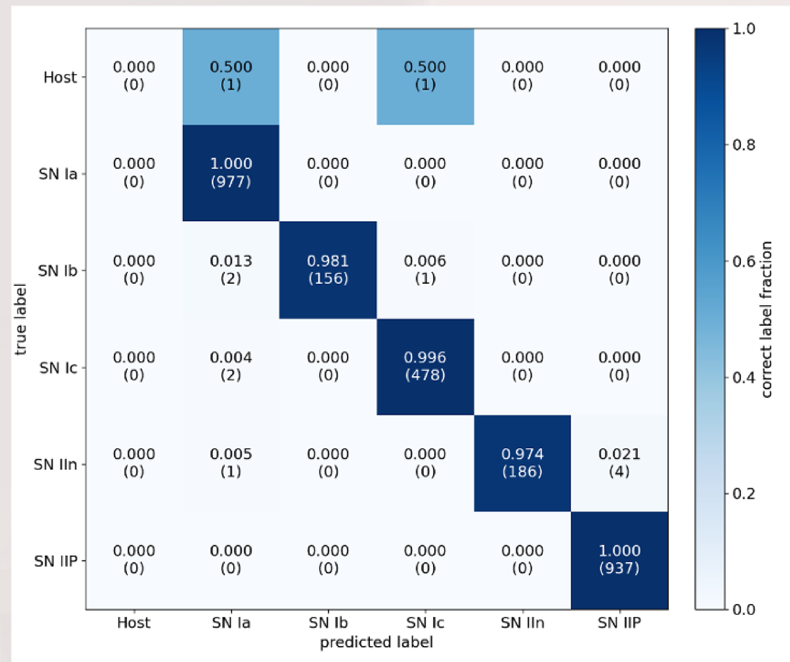


Figure 3: Confusion matrices for all spectra (left) and after making cuts (right)

### ROC Curves

- The ROC curve shows that it is straightforward to obtain a sample of Type Ia SNe with recall > 85% and false positive rate <1%

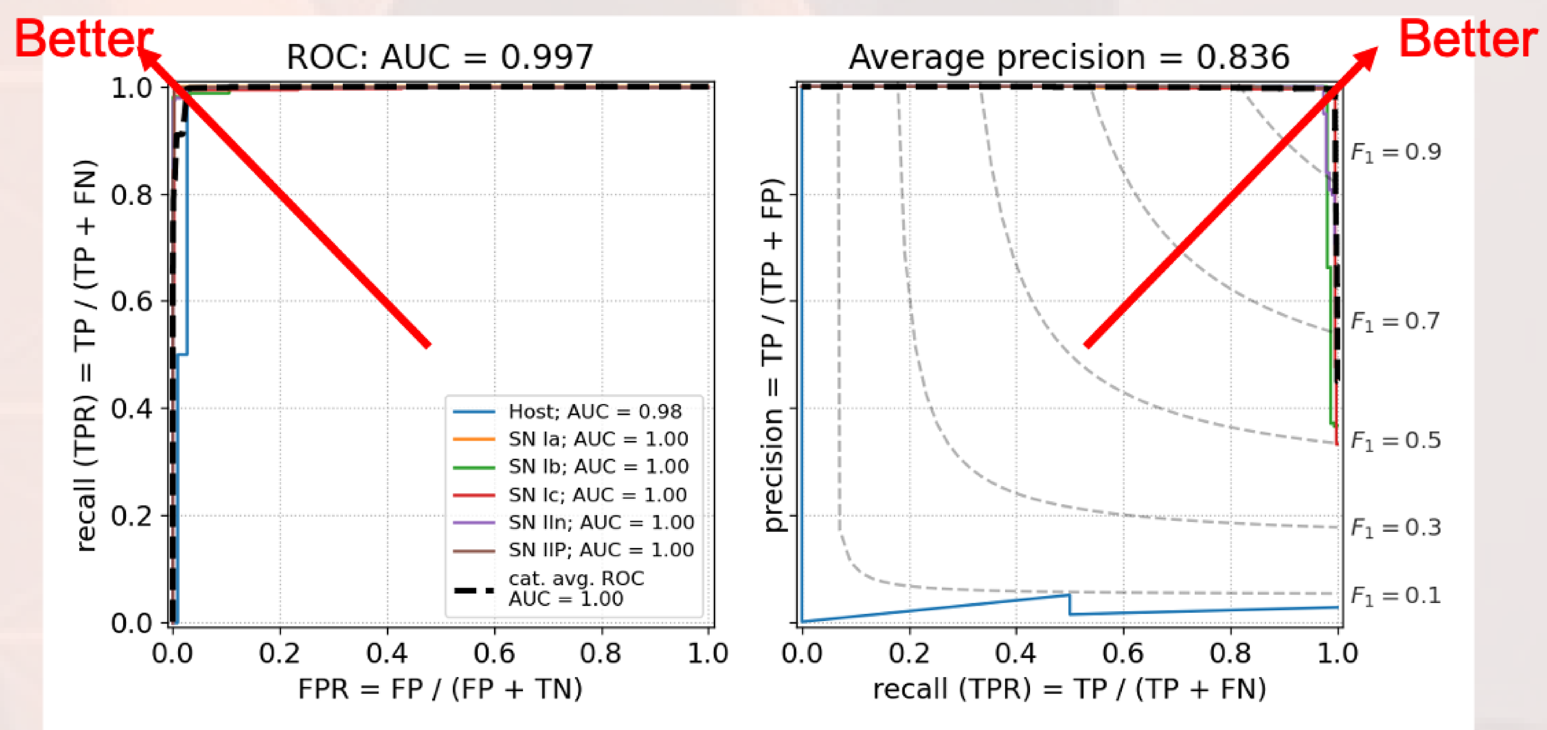


Figure 4: The ROC and PR curves for our CNN

## Conclusions

- High accuracy for simulated transients which is further increased by hard cuts on probable transients.
- Making the sample more pure decreases the false positive rate and stop most hosts from being falsely classified as transients.

## Future Directions

- Optimizing the network to run on real data.
- Understand the features that network is training on using SCORE-CAM
- Image classification instead of classifying 1-D fluxes.

## References


[1] Alexei V. Filippenko. "Optical Spectra of Supernovae". In: *Ann. Rev. Astron. Astrophys.* 35.1 (1997), pp. 309–355.

[2] Rick Johnson. "Know your supernovae". In: *Astronomy Magazine* (2004).

[3] Divyanshu Gandhi. *ML-CNN Classification of Transients (IA, IIP) and Hosts.* 2019.

[4] Ryan Rubenzahl. *Identifying Type Ia Supernovae in Extragalactic Spectra.* 2018.

[5] DESI Collaboration et al. *The DESI Experiment Part I: Science,Targeting, and Survey Design.* 2016. arXiv: 1611.00036 [astro-ph.IM].

[6] DESI Collaboration et al. *The DESI Experiment Part II: Instrument Design.* 2016. arXiv: 1611.00037 [astro-ph.IM].



U.S. Department of Energy · Heising-Simons Foundation · Science & Technology Facilities Council · Gordon and Betty Moore Foundation