SuperRAENN: A Semisupervised Supernova Photometric Classification Pipeline Trained on Pan-STARRS1 Medium-Deep Survey Supernovae

V. Ashley Villar[1,2,19], Griffin Hosseinzadeh[1], Edo Berger[1], Michelle Ntampaka[1,3], David O. Jones[4], Peter Challis[1], Ryan Chornock[5], Maria R. Drout[6,7], Ryan J. Foley[4], Robert P. Kirshner[1,8], Ragnhild Lunnan[9], Raffaella Margutti[5], Dan Milisavljevic[10], Nathan Sanders[11], Yen-Chen Pan[12], Armin Rest[13,14], Daniel M. Scolnic[15], Eugene Magnier[16], Nigel Metcalfe[17], Richard Wainscoat[16], and Christopher Waters[18]

[1] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138-1516, USA; vav2110@columbia.edu
[2] Department of Astronomy, Columbia University, New York, NY 10027-6601, USA
[3] Harvard Data Science Initiative, Harvard University, Cambridge, MA 02138, USA
[4] Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064-1077, USA
[5] Center for Interdisciplinary Exploration and Research in Astrophysics and Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3112, USA
[6] David A. Dunlap Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4 Canada
[7] Observatories of the Carnegie Institute for Science, 813 Santa Barbara Street, Pasadena, CA 91101-1232, USA
[8] Gordon and Betty Moore Foundation, 1661 Page Mill Road, Palo Alto, CA 94304-1209, USA
[9] Oskar Klein Centre, Department of Astronomy, Stockholm University, Albanova University Centre, SE-106 91 Stockholm, Sweden
[10] Department of Physics and Astronomy, Purdue University, 525 Northwestern Avenue, West Lafayette, IN 47907-2036, USA
[11] Berkman Klein Center, Harvard University, Cambridge, MA 02138, USA
[12] Graduate Institute of Astronomy, National Central University, 300 Jhongda Road, Zhongli, Taoyuan, 32001, Taiwan
[13] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218-2410, USA
[14] Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA
[15] Department of Physics, Duke University, Campus Box 90305, Durham, NC 27708, USA
[16] Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822-1839, USA
[17] Department of Physics, Durham University, South Road, Durham, DH1 3LE, UK
[18] Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08540-7219, USA

Abstract

Automated classification of supernovae (SNe) based on optical photometric light-curve information is essential in the upcoming era of wide-field time domain surveys, such as the Legacy Survey of Space and Time (LSST) conducted by the Rubin Observatory. Photometric classification can enable real-time identification of interesting events for extended multiwavelength follow-up, as well as archival population studies. Here we present the complete sample of 5243 "SN-like" light curves (in $g_{P1}r_{P1}i_{P1}z_{P1}$) from the Pan-STARRS1 Medium-Deep Survey (PS1-MDS). The PS1-MDS is similar to the planned LSST Wide-Fast-Deep survey in terms of cadence, filters, and depth, making this a useful training set for the community. Using this data set, we train a novel semisupervised machine learning algorithm to photometrically classify 2315 new SN-like light curves with host galaxy spectroscopic redshifts. Our algorithm consists of an RF supervised classification step and a novel unsupervised step in which we introduce a recurrent autoencoder neural network (RAENN). Our final pipeline, dubbed SuperRAENN, has an accuracy of 87% across five SN classes (Type Ia, Ibc, II, IIn, SLSN-I) and macro-averaged purity and completeness of 66% and 69%, respectively. We find the highest accuracy rates for SNe Ia and SLSNe and the lowest for SNe Ibc. Our complete spectroscopically and photometrically classified samples break down into 62.0% Type Ia (1839 objects), 19.8% Type II (553 objects), 4.8% Type IIn (136 objects), 11.7% Type Ibc (291 objects), and 1.6% Type I SLSNe (54 objects).

Unified Astronomy Thesaurus concepts: Supernovae (1668); Astrostatistics (1882); Light curve classification (1954)

Supporting material: machine-readable tables

1. Introduction

Time-domain astrophysics has entered a new era of large photometric data sets thanks to ongoing and upcoming wide-field surveys, including Pan-STARRS (PS; Kaiser et al. 2010), the Asteroid Terrestrial-impact Last Alert System (Jedicke et al. 2012), the All-Sky Automated Survey for SuperNovae (Shappee et al. 2014), the Zwicky Transient Facility (ZTF; Kulkarni 2018), the Legacy Survey of Space and Time (LSST; Ivezic & LSST Science Collaboration 2010), and the Roman Space Telescope (Spergel et al. 2015). LSST, to be conducted by the Vera C. Rubin Observatory between 2023 and 2033, is expected to

discover roughly 1 million supernovae (SNe) per year, an increase of more than two orders of magnitude compared to the current rate.

Historically, SNe and other optical transients have been classified primarily based on their optical spectra. Class labels are largely phenomenological, dependent on the presence of various elements in the photospheric-phase spectra (see, e.g., Filippenko 1997 for a review). SNe, for example, have historically been classified as Type I (equivalent to today's Type Ia) or Type II based on the absence or presence of strong hydrogen Balmer lines, respectively. As the number of events increased, further classes were created to account for the increased diversity (e.g., Uomoto & Kirshner 1985). Type Ib and Type Ic designations were created to

---

[19] Simons Junior Fellow.

indicate the presence and absence of helium, respectively. Today, semiautomated software such as GELATO (Harutyunyan et al. 2008), SNID (Blondin & Tonry 2007), and Superfit (Howell et al. 2005) are used to match SN spectra to a library of previously classified events to determine the spectroscopic class. More recently, Muthukrishna et al. (2019b) utilized a convolutional neural network to classify SN spectra.

However, spectroscopic follow-up remains an expensive endeavor, taking up to an hour on 8 m class telescopes to classify a single object given the depth wide-field surveys can now achieve. As a result, only ∼10% of the ∼10^4 transients currently discovered each year are spectroscopically classified.[20] Spectroscopic follow-up is not expected to significantly increase when the LSST commences, meaning that only ∼0.1% of events will be spectroscopically classified.

Given the growing rate of discovery and limited spectroscopic resources, classification of transients based on their photometric light curves is becoming essential. Luckily, the phenomenological labels often correspond to unique underlying processes that are also encoded in the light-curve behavior. For example, while SNe Ia are spectroscopically classified by strong Si II absorption and lack of hydrogen, these features distinctly originate from the thermonuclear detonations or deflagrations of carbon–oxygen white dwarfs, which also lead to specific light-curve evolution (Hillebrandt & Niemeyer 2000). Generally, unique progenitor system and explosion mechanisms likely lead to other observable features, some of which are captured in broadband optical light curves. Said features allow transients to be classified into their traditional subclasses (based on spectroscopy and photometry) using *only* their broadband, optical light curves.

There is a growing literature on light-curve classifiers that rely on data-driven and machine learning algorithms. Most studies use *supervised* learning, in which the training set consists of SNe with known classes (e.g., Lochner et al. 2016; Charnock & Moss 2017; Boone 2019; Villar et al. 2019; Möller & de Boissière 2020). However, SN classification can benefit from *semisupervised* methods, in which the training set contains both labeled and unlabeled SNe. The unlabeled set is used to better understand low-dimensional structure in the SN data set to improve classification. Richards et al. (2012), for example, created a diffusion map (a nonlinear dimensionality reduction technique) based on light-curve similarities in shape and color using unlabeled data from the Supernova Photometric Classification Challenge (SPCC; Kessler et al. 2010). They use the diffusion map to extract 120 nonlinear SN features from each labeled SN, which are then used to train a random forest (RF) classifier. More recently, Pasquet-Itam & Pasquet (2018) introduced the PELICAN classifier, also trained on synthetic SPCC data. PELICAN uses a convolutional autoencoder to encode nonlinear SN features and a set of fully connected neural network layers to classify the full set of simulated SPCC light curves as Ia or non-Ia SNe.

Here we introduce a new semisupervised classification method for SNe, which utilizes a recurrent autoencoder neural network (RAENN). This method is uniquely trained on real (rather than simulated) data from the Pan-STARRS1 Medium-Deep Survey (PS1-MDS) and is optimized for general SN classification (as opposed to Ia vs. non-Ia classification). Our method has been trained on a combination of 557 spectroscopically classified SNe and 2328 additional SN-like events.

We then use RAENN and hand-selected features with an RF to classify the PS1-MDS sample of 2315 previously unclassified SN-like transients with host galaxy spectroscopic redshifts. We publish the full set of light curves and associated labels for community use. We present an open-source code listed on the Python Package Index as SuperRAENN (Villar 2020). A companion paper, Hosseinzadeh et al. (2020, hereafter H20), presents and compares photometric classifications of the same data set using an independent classification method (following the supervised methods of our previous work in Villar et al. 2019).

The paper is organized as follows. In Section 2, we review the PS1-MDS and associated sample of SN-like transients. In Section 3 we introduce the RAENN architecture and training procedure. We present the classification results and discuss implications in Section 4 and Section 5, respectively. We conclude in Section 6. Throughout this paper, we assume a flat ΛCDM cosmology with $\Omega_M = 0.307$ and $H_0 = 67.8$ km s^{-1} Mpc^{-1} (Ade et al. 2014).

2. The PS1-MDS Supernova Sample

PS1 is a wide-field survey telescope located near the summit of Haleakala, Maui, with a 1.8 m diameter primary mirror and a 1.4 gigapixel camera (GPC1; Kaiser et al. 2010). PS1-MDS, one of several PS1 surveys (Chambers et al. 2016), was conducted between 2009 and 2014 July. It consisted of 10 single-pointing fields, each of approximately 7.1 deg^2, with a pixel scale of 0″.25. The survey was conducted in five broadband filters (Stubbs et al. 2010; Tonry et al. 2012) with a nominal cadence of 3 days per filter in four filters ($g_{P1}r_{P1}i_{P1}z_{P1}$) and a 5σ limiting magnitude of ≈23.3 per visit. In practice, Scolnic et al. (2018) find a cadence of roughly 6–7 days per filter. In general, PS1-MDS observed a field in g_{P1} and r_{P1} on the same night, followed by i_{P1} and then z_{P1} on subsequent nights. PS1-MDS also included observations in the y_{P1} band, primarily near full moon and with a shallower 5σ limiting magnitude of ≈21.7. Due to the poor cadence and shallow depth, we do not use the y_{P1} data here.

During its 5 yr survey, PS1-MDS discovered 5243 SN-like objects, defined as events with at least three observations with a signal-to-noise ratio (S/N) > 4 in any filter and no previous detection within the survey difference images (Jones et al. 2018, 2019). We obtain data for these events via the PS Data Processing System (Magnier et al. 2020a, 2020b; Waters et al. 2020). The photometric pipeline is based on photpipe (Rest et al. 2005, 2014) with improvements made in Scolnic et al. (2018). Images and templates, used for image subtraction, are resampled and aligned to match a "skycell" in the PS1 sky tessellation; we use the difference images to build the light curves used in this analysis. Image zero-points are determined by comparing point-spread function (PSF) photometry of stars to PS1 stellar catalogs (Chambers et al. 2016). PS1 templates are convolved to match nightly images and then subtracted using HOTPANTS (Becker 2015). For each event, a flux-weighted centroid is calculated and forced PSF photometry is performed at the centroid. Finally, a nightly zero-point is applied.

Of the 5243 SN-like objects, 4090 host galaxies were targeted through a concerted observational effort. To identify the most likely host galaxy for each SN, we used the galaxy size and orientation-weighted *R*-parameter from Sullivan et al. (2006), as outlined in Jones et al. (2017). The majority (3321 objects) were observed using the Hectospec multifiber instrument on MMT

---

[20] Based on data from the public Open Supernova Catalog (Guillochon et al. 2017) and the Transient Name Server.

(Fabricant et al. 2005; Mink et al. 2007). Additional host redshifts were obtained with the Anglo-Australian Telescope (290 objects), the WIYN telescope (217 objects), and the Apache Point Observatory 3.5 m telescope (APO; 5 objects). Host galaxies selected for follow-up were largely unbiased in terms of transient properties (e.g., we did not prioritize SNe based on luminosity, color, visibility in PS1-MDS template image, or amount of additional follow-up). Additional host redshifts were obtained from archival survey data: 2dFGRS (Colless et al. 2003), 6dFGS (Jones et al. 2009), DEEP2 (Newman et al. 2013), the Sloan Digital Sky Survey (SDSS; Smee et al. 2013), VIPERS (Scodeggio et al. 2018), VIMOS (Le Fèvre et al. 2005), WiggleZ (Blake et al. 2008), and zCOSMOS (Lilly et al. 2009).

We use the RVSAO package (Kurtz & Mink 1998) to determine the spectroscopic redshifts through cross-correlation with galaxy templates. We use the standard RVSAO galaxy templates (including spiral and elliptical galaxies and quasars), as well as galaxy templates provided by SDSS (Adelman-McCarthy et al. 2007).[21] We quantify the quality of the template matches using the Tonry & Davis (1979) cross-correlation parameter, R_{CC}. Following Jones et al. (2017), we remove host galaxies with $R_{CC} < 4$, ensuring that the vast majority ($\approx$98%) of the remaining galaxies have accurate redshift measurements. This cut removes 1084 SNe with redshift measurements.

To ensure that our final set of redshift measurements is robust, we identify a subset of spectra to be manually inspected and validated. We do *not* inspect any redshifts from external catalogs. Of redshifts that we initially estimate using RVSAO, we do *not* inspect spectra if the median redshift estimate across templates is equal to both the most likely redshift and the mode of the template matches *and* more than two templates match this redshift estimate. We visually inspected $\sim$600 redshift spectra to ensure that our final redshift estimates are as accurate as possible. In total, 2487 redshifts (of 3056 redshift estimates with $R_{CC} \geqslant 4$) match the most likely redshift provided by RVSAO. Of the remaining hosts, we remove 393 redshift estimates that we could not validate manually. A total of 145 redshifts ($\sim$4%) that were measured manually do not match the median or mode of the RVSAO redshift estimates. The galaxy spectra and further details are presented in H20.

We additionally remove events with $z < 0.005$, which are unlikely to be SNe given the peak absolute magnitudes (e.g., Chornock et al. 2010b). We visually inspect the light curves that have quasar-like hosts (based on template matching) or that overlap with the host galaxy's center. We remove events that are clearly variable over multiple seasons and lack a transient spectrum. Our final sample includes 2885 transients with redshifts measurements (from the hosts or transients themselves), including spectroscopically identified SNe.

2.1. Spectroscopic versus Photometric SN Sample

Approximately 10% of the PS1-MDS transients were spectroscopically observed in real time throughout the survey, without a specific selection function (although brighter objects were more likely to be targeted). For this work, we limit our spectroscopic sample (557 objects) to five potential classes:

1. Type I SLSNe (17 objects), which are thought to arise from the birth of highly magnetized neutron stars (Quimby et al. 2007; Chomiuk et al. 2011; Nicholl et al. 2017).
2. SNe II (93 objects; including Type IIP and Type IIL SNe[22]), which arise from red supergiant progenitors.
3. SNe IIn (24 objects), powered by the interaction the SN ejecta with preexisting circumstellar material (e.g., Smith et al. 2014).
4. SNe Ia (404 objects), which are the thermonuclear explosions of white dwarfs.
5. SNe Ibc (19 objects), which arise from the core collapse of massive stars that have lost their hydrogen (Ib) and helium (Ic) envelopes. Due to the small sample size, we consider SNe Ib and SNe Ic as a single class.

The SLSN and SN Ia light-curve samples have been previously published in Lunnan et al. (2018) and Jones et al. (2017), respectively. Model fits to the Type II light curves were presented in Sanders et al. (2015). For four objects, the transient spectra yield a reliable redshift but an ambiguous classification. A fifth object, PSc130816, has previously been identified as both an SN IIP/L (Sanders et al. 2015) and an SN IIn (Drout 2016). We do not include these five objects in our spectroscopic sample. An additional 15 objects are spectroscopically identified but do not fall in one of our five classes, including two tidal disruption events (TDEs), a lensed Type Ia, a Type Ibn, a Type Iax, and 10 fast-evolving luminous transients (FELTs). All except the TDEs are included in our photometric sample for training purposes, but not included in our spectroscopic sample. These objects are discussed in more detail in Section 5.1.

Our photometric sample contains 2315 objects with host galaxy spectroscopic redshifts, which are independent of the 557 SNe that are spectroscopically classified, as well as 13 SNe that do not fall in one of the five spectroscopic classes, discussed in Section 5.1. We refer to the union of the photometric and spectroscopic samples (the full set of 2885 events) as the "complete" photometric data set. We summarize the PS1-MDS SN-like objects, their associated hosts, and redshift information in Table 1. We also specify which SNe are used in the supervised/unsupervised portions of our classification algorithm.

Our spectroscopic data set is brighter than our complete photometric data set. As shown in Figure 1, the spectroscopic sample has a median peak r-band magnitude of $\sim$−21 mag, about 1 mag brighter than the photometric sample. We directly compare the redshift distributions in Figure 2. The spectroscopic sample peaks at a slightly lower redshift compared to the photometric data set ($z \approx 0.27$ vs. $z \approx 0.35$), with a tail extending to $z \approx 1.0$. The lack of confident high-redshift measurements is likely due to the key spectroscopic lines shifting out of the optical and due to the peak absolute magnitudes of most SNe falling below our limiting magnitude. The mismatch between the spectroscopic and photometric samples may translate to biases in our classification pipeline, which we explore in more detail in Section 5. The complete *griz* light curves of our sample are available on Zenodo (Villar et al. 2020).

---

[21] http://classic.sdss.org/dr5/algorithms/spectemplates/

[22] Type IIP and Type IL are thought to arise from the same progenitor population. See, e.g., Sanders et al. (2015).

Table 1
SN Properties

Object Name	PSc ID	IAU Name	CBET	SN R.A. (deg)	SN Decl. (deg)	Milky Way $E(B-V)$	SN Type	Redshift[a]	Host R.A. (deg)	Host Decl. (deg)	Num. Points[c]	Telescope[b]/ Source	Unsup.[d]	Sup.[e]
PS1-0909006	PS0909006	SN 2009ks	Rest et al. (2009)	333.9503	1.1848	0.0426	SN Ia	0.284	⋯	⋯	19	Gem-N	Y	Y
PS1-0909010	PS0909010	SN 2009kt	Rest et al. (2009)	37.1182	−4.0789	0.0256	SN Ia	0.27	⋯	⋯	20	Gem-N	Y	Y
PS1-0910012	PS0910012	SN 2009ku	Rest et al. (2009)	52.4718	−28.0867	0.0073	SN Iax	0.079	⋯	⋯	40	Gem-S, NOT, Magellan, Gem-S	Y	N
PS1-0910016	PS0910016	SN 2009kx	Rest et al. (2009)	35.3073	−3.91	0.0219	SN Ia	0.23	⋯	⋯	10	Gem-N	Y	Y
PS1-0910017	PS0910017	SN 2009kv	Rest et al. (2009)	35.2775	−5.0233	0.0221	SN Ia	0.32	⋯	⋯	15	Gem-N	Y	Y
PS1-0910018	PS0910018	SN 2009kz	Rest et al. (2009)	35.667	−4.0273	0.0242	SN Ia	0.265	⋯	⋯	11	Gem-N	Y	Y
PS1-0910020	PS0910020	SN 2009kw	Rest et al. (2009)	54.5975	−28.2533	0.013	SN Ia	0.242	⋯	⋯	13	Gem-S	Y	Y
PS1-0910021	PS0910021	SN 2009ky	Rest et al. (2009)	53.62	−27.9084	0.0081	SN Ia	0.256	⋯	⋯	20	Gem-S	Y	Y
PS1-10a	PSc000001	⋯	⋯	52.4531	−29.075	0.009	SN II	0.071	52.4536	−29.0744	13	2df	Y	Y
PS1-10aa	PSc010046	⋯	⋯	162.9188	58.1822	0.0115	⋯	0.039	162.9188	58.1822	21	WIYN	N	N
PS1-10aaa	PSc070003	⋯	⋯	333.0415	0.7193	0.0479	⋯	0.1376	333.0414	0.7194	63	NED, SDSS	Y	N
PS1-10aab	PSc070004	⋯	⋯	335.0863	0.524	0.0641	⋯	⋯	335.088	0.5244	31	⋯	N	N
PS1-10aac	PSc070039	⋯	⋯	334.6541	0.6184	0.0587	⋯	0.3933	334.6538	0.6186	28	MMT	N	N
PS1-10aad	PSc070048	⋯	⋯	241.0171	54.1988	0.0088	⋯	0.423	241.017	54.1988	15	MMT	Y	N

Notes.
[a] Redshift estimate from either the transient spectra or host galaxy spectra.
[b] Telescope used to acquire galaxy or SN spectra.
[c] Number of $>5\sigma$ data points in light curve.
[d] Included in unsupervised training set. These objects have reliable host redshift estimates.
[e] Included in supervised training set.

References. Dressler & Gunn (1992), Im et al. (2001), Szokoly et al. (2004), Le Fèvre et al. (2005), Norris et al. (2006), Cannon et al. (2006), Tajer et al. (2007), Garcet et al. (2007), Lilly et al. (2007), Bronder et al. (2008), Ross et al. (2008), Lamareille et al. (2009), Trump et al. (2009), Finkelstein et al. (2009), Scarlata et al. (2009), Owen & Morrison (2009), Rest et al. (2009), Balestra et al. (2010), Cowie et al. (2010), Valenti et al. (2010), Young et al. (2010), Zheng et al. (2010),Challis et al. (2010), Chornock et al. (2010a), Stalin et al. (2010), Drinkwater et al. (2010), Hewett & Wild (2010), Rovilos et al. (2011), Chornock et al. (2012), Fraser et al. (2012), Cappellaro et al. (2012a, 2012b), Smith et al. (2012), Ferrante et al. (2013), Drake et al. (2013), Valenti et al. (2013), Takats et al. (2013), Lunnan et al. (2014), Drout et al. (2014), Karhunen et al. (2014), Wen & Han (2015).

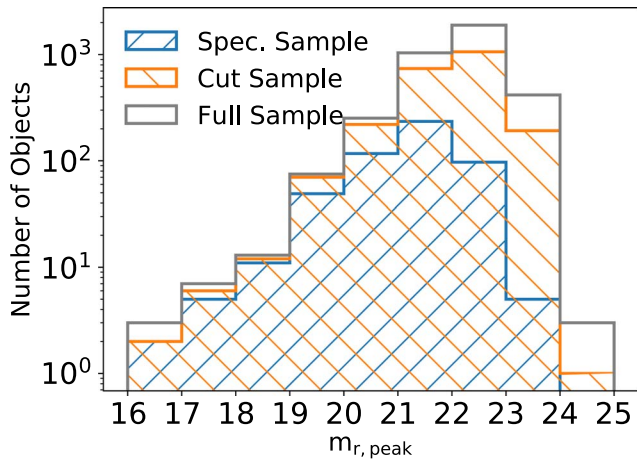(This table is available in its entirety in machine-readable form.)

Figure 1. Peak apparent *r*-band magnitude of the full SN-like data set (gray), objects used in our unsupervised method (orange), and the spectroscopic sample (blue). The spectroscopic data set is roughly 1 mag brighter than the full data set.
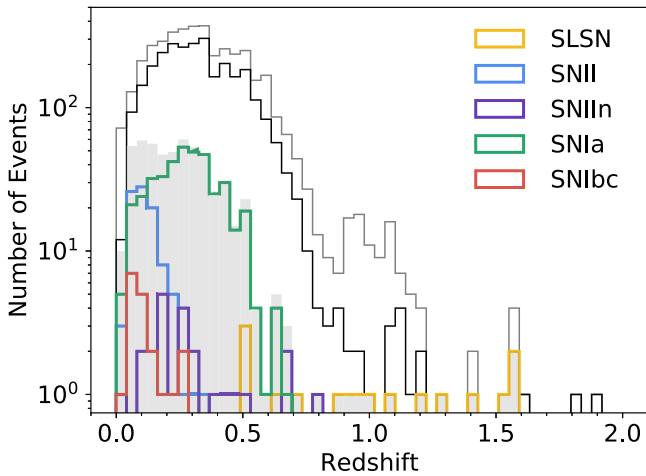

Figure 2. Histogram of the redshifts of the full SN-like data set (gray line; 4055 objects), the subset of host redshift measurements for objects used in our unsupervised learning algorithm (black line; 2885 objects), and the subset with spectroscopic classification (colored lines; 557 objects). The shaded gray region represents the summed, spectroscopically classified objects. The full sample and spectroscopic distribution peak at $z \approx 0.25$, although the spectroscopic sample has an additional peak near $z \approx 0.1$. At $z \gtrsim 0.75$, the spectroscopic sample is limited to SLSNe.

We explore the overall data quality of our sample in Figure 3, finding that the majority of events have ~20 data points across all filters with S/Ns of $\gtrsim 3$. Given a typical SN duration of a month and our typical cadence of a few days, we expect the majority of (but not all) SNe to have fairly complete light curves.

3. A Semisupervised Classification Pipeline

About 10% of our SN sample is spectroscopically classified. Traditional supervised classification methods are strictly limited to this subset of our data, as they require labeled SN examples. However, information about SN subtypes exists as substructure in the unlabeled data set as well. For example, SN classes may be clustered in duration and luminosity (e.g., Kasliwal 2012; Villar et al. 2017). Because we would like to leverage the information in both the labeled and unlabeled subsets of the training set, we use a recurrent RAENN paired
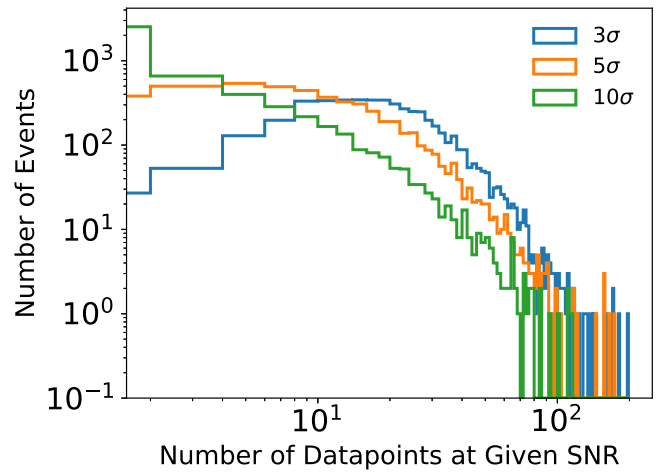
Figure 3. Histogram of the number of SN light curves with N data points with S/N of $\geqslant 3$ (blue), $\geqslant 5$ (orange), and $\geqslant 10$ (green) from the complete sample of SN-like objects (5243 events). Most events have $\approx$10–20 3σ data points, with only a handful having >100 points.

with an RF classifier for a semisupervised classification approach. In this section, we describe the complete algorithm and training process.

Our pipeline is composed of three steps: (1) a preprocessing and interpolation step using Gaussian processes (GPs), (2) an unsupervised step in which we train a RAENN on the complete photometric set (labeled and unlabeled), and (3) a supervised step in which we train an RF on the spectroscopically labeled set of SNe. The complete pipeline, dubbed `SuperRAENN` (Villar 2020), is available via GitHub.[23]

3.1. Preprocessing with Gaussian Processes

We generate and preprocess absolute magnitude light curves before extracting features. We correct each light curve for Milky Way reddening using the extinction map of Schlafly & Finkbeiner (2011). We estimate and normalize the absolute magnitude using the measured host redshift:

$$M_{\mathrm{norm}} = m - 5\log_{10}(d_{\mathrm{L}}/10\mathrm{pc})$$
$$+ 2.5\log_{10}(1 + z) - m_{\mathrm{lim}} - A_\lambda, \qquad (1)$$

where m_{lim} is a chosen limiting magnitude, which we take to be $m_{\mathrm{lim}} = 25$. This value is dimmer than the 5σ limiting magnitude of PS1-MDS. We choose a dimmer magnitude to ensure that even marginal detections will be included in the light curve. We perform the renormalization so that the GP mean will be zero (i.e., the light curve will be zero when no data are available). Finally, we correct all light curves for time dilation based on the measured redshifts. We do not attempt to make a wavelength-dependent *k*-correction to the rest-frame data given the complicated, diverse, and time-evolving spectral energy distributions (SEDs) of the various SN types.

We do not correct the SN light curves for host galaxy reddening. The intrinsic reddening of SNe adds an additional scatter in our feature space. Correcting for host galaxy reddening would require estimating both the color excess and dust law, which is not possible given our current data set.

The PS1-MDS light curves are irregularly sampled across the four filters (see Section 2 for the PS1 observing strategy).
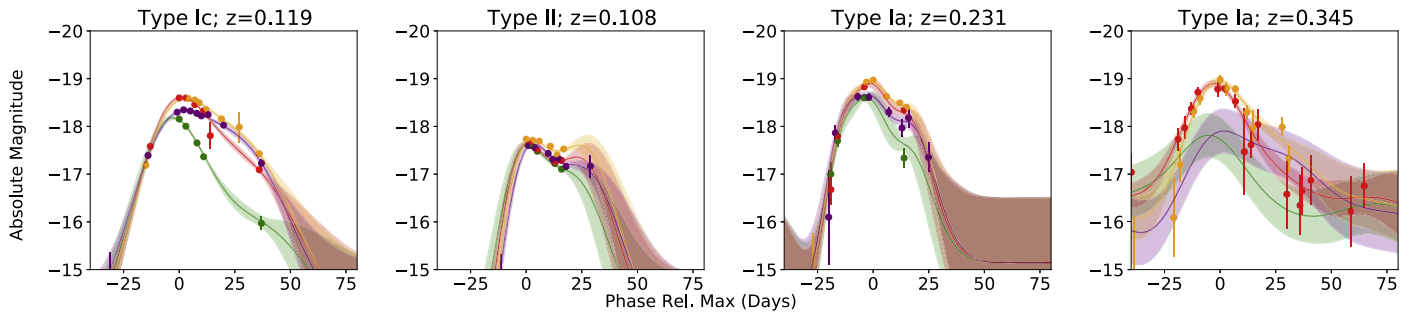
---

[23] https://github.com/villrv/SuperRAENN

Figure 4. Examples of four spectroscopically classified SNe and their associated GP-interpolated light curves in the four PS1 filters (*g*: green; *r*: red; *i*: orange; *z*: purple). Solid lines represent the mean GP prediction, while the shaded regions represent the 1σ estimated uncertainties. Note that the leftmost example lacks *g* and *z* band, but the GP provides reasonable light-curve estimates.

The architecture of the RAENN does *not* require uniformly sampled light curves. However, it does require that each observation is made in all four filters. For example, if an observation is made in *g* band, we need to provide interpolated values for *riz* bands for that time.

To interpolate the *griz* light curves, we fit a GP using the open-source Python package `George` (Ambikasaran et al. 2015). GPs are a nonparametric model that has been previously used to interpolate and classify SN light curves (see, e.g., Lochner et al. 2016; Revsbech et al. 2018; Boone 2019). GPs define a prior over a family of functions, which is then conditioned on the light-curve observations. A key assumption is that the posterior distribution describing the light curve is Gaussian, described by a mean, $\mu(t)$, and a covariance matrix, $\Sigma(t)$, given by $\Sigma_{i,j} = \kappa(x_i, x_j)$ with kernel κ. We use a 2D squared exponential kernel to simultaneously fit all four filtered light curves:

$$\kappa(t_i t_j f_i f_j; \sigma, l_t l_f) = \sigma^2 \exp\left[-\frac{(t_i - t_j)^2}{2l_t^2}\right]$$

$$\times \exp\left[-\frac{(f_i - f_j)^2}{2l_f^2}\right], \qquad (2)$$

where *f* is an integer between 0 and 3 that represents the *griz* filters, and the parameters l_t and l_f are characteristic correlation length scales in time and filter integer, respectively. This fitting process accounts for the measured data uncertainties, making it robust to low-confidence outliers.

We independently optimize the kernel parameters for each SN using the `minimize` function implemented in `scipy`, with initial values of $l_t = 100$ days and $l_f = 1$. We find that our choice of initialization values has little effect on the resulting best fit. We find that l_t is typically about 1 week, and l_f is typically 2–3, indicating that the filters are highly correlated. Examples of the GP interpolation for SNe Ia, SNe Ic, and SNe II are shown in Figure 4. The GP is able to produce reasonable interpolated light curves, even in cases with sparse and noisy data, and provide reasonable error estimates.

A similar GP method was implemented by Boone (2019) to classify a variety of SN types in the Photometric LSST Astronomical Time-series Classification (PLAsTiCC; The PLAsTiCC team et al. 2018; Kessler et al. 2019) data set. Instead of an integer, Boone (2019) used the rest-frame central wavelength of each filter for each object. We avoid this added layer of complexity because the *k*-corrections and time-evolving SN SED change the weighted central filter wavelength. However, the simple 2D kernel still allows the four bands to share mutual information.

Our light curves contain several years of data, most of which are nondetections. To limit our input data, we keep data points (of any significance) within 100 days of peak flux (in whichever filter is brightest). For ease of optimization, the light curves need to contain the same number of data points. The data must be a consistent size during the back-propagation step of optimization for the RAENN for each iteration (see next section). Our longest light curve contains 169 data points, so we pad all light curves to match this length. We do so by appending a value dimmer than the estimated absolute limiting flux (we use $m_{\rm lim} = 25$) to 100 days after the last detection in the light curve.

We note that using luminosity-based light curves (rather than magnitudes) is an alternative preprocessing choice. Luminosity-based light curves would remove the need to renormalize the light curves to a chosen limiting magnitude. We find that using luminosity-based light curves results in worse performance of the RAENN, likely due to the orders-of-magnitude differences in scale between events.

3.2. Unsupervised Learning: A Recurrent Autoencoder Neural Network (RAENN)

To extract unique features from the complete (unlabeled and labeled) PS1-MDS photometric sample, we construct an RAENN, inspired by the work of Naul et al. (2018), who uses a similar method to classify variable stars.

Neural networks are a class of machine learning algorithms that use many latent layers to model complex functions. These and other machine learning algorithms are becoming increasingly common in astronomy (see Ntampaka et al. 2019 for an overview). Autoencoders (AEs; Kramer 1991) are a class of neural network architectures that learn a compressed representation of input data. By training an AE to return the original data given a limited set of variables, it learns an "encoded" version of the data.

In astrophysics, AEs have been used for feature learning in galaxy SEDs (Frontera-Pons et al. 2017), image denoising (Lucas et al. 2018; Ma et al. 2019), and event classification (Naul et al. 2018; Pasquet et al. 2019). AEs are also increasingly being used in the astrophysics literature for dimensionality reduction (see, e.g., Ralph et al. 2019 and Portillo et al. 2020 for recent examples).

Here, our model is designed to address several concerns of SN light curves: (1) the temporal irregularity of data, (2) data across multiple filters, and (3) streaming data that update on a
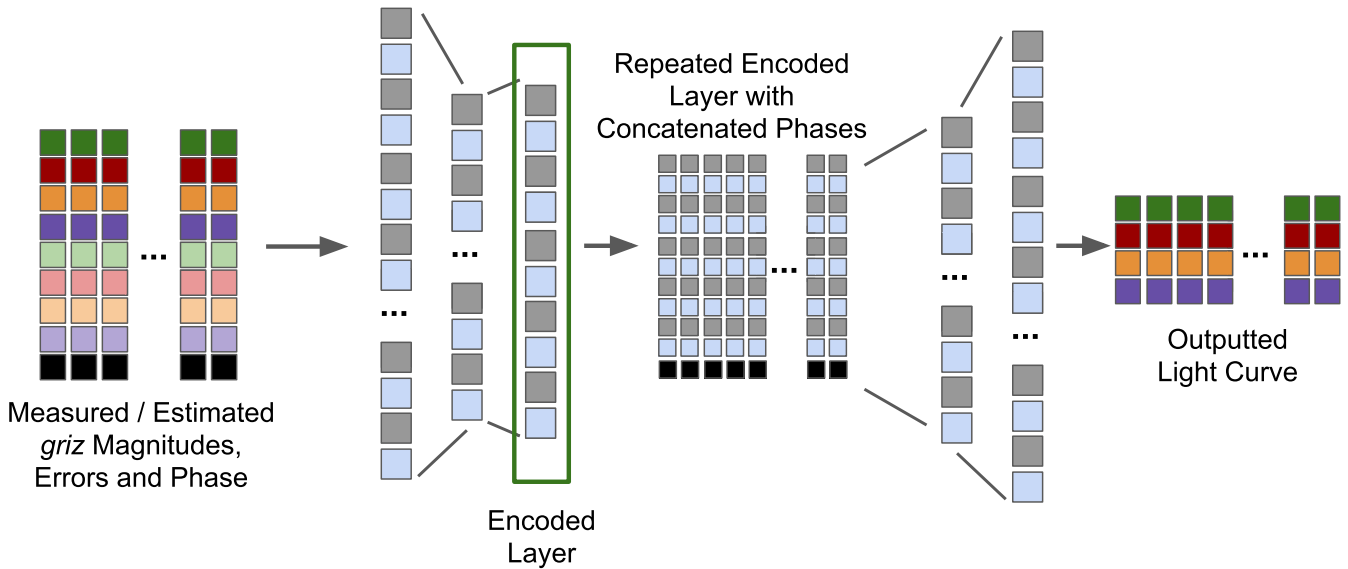
Figure 5. Diagram of the RAENN architecture. The preprocessed GP-interpolated light curves are fed into the encoder, which encodes the light curve into an encoding vector. This vector is then repeated, and new time values are appended to each copy. The final light curve is then predicted at each new time value. The RAENN is trained by comparing the input light curve with the predicted light curve. The values from the encoded layer are inputted into the RF as features and used to classify the SN light curves.

given cadence. The last point is not a concern for our PS1-MDS archival data set, but it will become important as LSST comes online and discovers thousands of SNe nightly.

The RAENN uses the GP light curves as input, by codifying the light curves as matrices of size $9 \times T_0$, where $T_0 = 169$, as described in the previous section. The nine values are as follows: one time value, relative to maximum (in whichever filter is brightest); four magnitude values ($griz$) at that time; and four magnitude uncertainties. Recall that the magnitude values are either measured or estimated from the GP. For the uncertainties, we use the 1σ errors for the measured points. For the GP points, we use a large error of 1 mag. We note that the GP produces estimate errors, but we find that, in practice, using this larger error bar leads to better performance. We leave exploration of utilizing error bars to future work. We emphasize that, while $T_0 = 169$ for training, the RAENN architecture allows a user to input a light curve of any size without needing to pad the light curve.

The RAENN architecture is divided into an encoder and a decoder. Our encoder is a series of fully connected layers that decrease in size until the final encoded layer with size N_E (i.e., the number of neurons used to fully encode the SN light curve). We note that N_E is a free parameter of our model that needs to be optimized. Similarly, the fully connected layer has N_N neurons, where $N_N > N_E$ and is also a tunable parameter. Following the encoded layer, the decoder half of the architecture mirrors the encoder with increasing layer sizes.

A novel feature of our architecture is the inclusion of a repeat layer immediately after our final encoding layer (the layer of size N_E). In this layer, we repeat the encoded version of the light curve T_N times. To each copy, we append the time of each data point, relative to peak brightness in one filter (whichever filter happens to be brightest). One way to view the purpose of this layer is to imagine the autoencoder as two functions. The first function (the encoder) takes in the original data points, including observation times, and outputs a set of N_E values. This is similar to the idea of taking a light curve and fitting it to a model with N_E free parameters. A second function (the decoder) takes in a set of N_E values and T_N times to generate a light curve at the T_N times. This architecture allows us to generate a light curve at different T_N times, e.g., interpolated or extrapolated light curves, which is further explored in Section 5. In this work, we choose $T_N = T_0$, namely, we repeat the encoded values to match the original light-curve length.

Our autoencoder utilizes gated recurrent neurons (GRUs; Rumelhart et al. 1988; Cho et al. 2014). In addition to the typical hidden weights that are optimized during training, recurrent neurons have additional weights that act as "memory" of previous input. GRUs in particular utilize an *update* value (called a gate) and a *reset* gate. The values of these neurons determine how the current and previous inputs affect the value of the output. With each light-curve data point, the gates become updated with new information that informs the next prediction. This class of neurons is useful for our light curves with various numbers of observed data points. Our GRU neurons use the tanh activation functions with a hard sigmoid for the gate activation function.

Our RAENN is implemented in Keras (Chollet 2015) with a Tensorflow back end (Abadi et al. 2016). A diagram of the architecture is shown in Figure 5 and is outlined as follows:

1. *Input Layer*: Input light curve of size $T_0 \times 9$ with each $griz$ data point labeled with a time (one value) in days relative to light-curve peak (four values) and an uncertainty (four values).
2. *Encoding Layer*: Encoding layer with N_N neurons, where N_N is a hyperparameter.
3. *Encoded Layer*: Encoded light curve with N_E neurons, where N_E is a hyperparameter.
4. *Repeat Layer*: Layer to repeat encoded light curve to match with new time array, with size $T_0 \times N_E$.
5. *Concatenate Layer*: Layer to concatenate new times to encoded light curve, with size $T_0 \times (N_E + 1)$.
6. *Decoding Layer*: Decoding layer with N_N neurons.
7. *Decoded Layer*: $T_0 \times 4$ decoded $griz$ light curve.

To optimize the free parameters (the weights) of the RAENN model, we must define a *loss function*. Our loss function is a simple mean square error function:

$$\mathcal{L} = \sum_{i=0}^{N} \frac{[F_{i,\text{True}}(t, f) - F_{i,\text{Predicted}}(t, f)]^2}{N}, \qquad (3)$$

where F is the SN flux as a function of time t and filter f. Although we feed uncertainties into the network, we find that excluding flux errors in our loss function substantially improved the ability of the RAENN to match the input light curves. We minimize our loss function using the gradient-descent-based optimizer, `Adam` (Kingma & Ba 2014), finding an optimal learning rate of 10^{-4}, which is a typical value.

We randomly split our unlabeled data set into training (2/3) and test (1/3) sets. We optimize the number of neurons in both the encoding and decoding layers (fixed to be the same number, N_N) and the number of encoding neurons (N_E) through a grid search, allowing N_N to vary from 20 to 160 in intervals of 20 and N_E to vary between 2 and 24 in intervals of 2. We find that, when optimizing over final classification F1-score (defined below), purity, and completeness, our results are relatively insensitive to N_E and N_N for values of $N_E \sim 10$ and $N_N \sim 100$. For our final model, we use $N_E = 8$ and $N_N = 120$, which is *not* our optimal model but a *representative* model. Utilizing our optimal model without creating a test set (in addition to a training and validation set) would likely overestimate performance. Given our limited data set, we are unable to properly optimize our hyperparameters and thus present representative results. Our best-performing model ($N_E = 4$ and $N_N = 160$) performs slightly better, with a macro-average completeness and purity of 71% and 69%, respectively. We note that N_N is slightly below the maximum number of data points in our set of light curves (where the longest light curve has 169 observed data points). The number of encoding neurons N_E is similar to the number of free parameters for the analytical model used in Villar et al. (2019) to capture the shape of a single-filter SN light curve.

We contrast our architecture with methods from Naul et al. (2018) and Pasquet et al. (2019), who present similar methodologies. Naul et al. (2018) use a similar GRU-based RAENN to classify variable stars with unevenly sampled light curves in one filter from the All Sky Automated Survey Catalog of Variable stars (Pojmanski 2002). The flux and time since last observation (Δt) are sequentially read into the recurrent layers. The same time array is fed into the decoder for output. In our case, we feed in a time series across four filters and give a time array relative to peak rather than relative to the previous data point. This is more natural in our problem, in which the SNe have a clear beginning and end, versus the periodic signals of variable stars. Additionally, our architecture allows us to give the decoder a different time series to allow for interpolation or extrapolation of the data.

Pasquet et al. (2019) use a semisupervised method to classify simulated SN light curves from the SPCC (Kessler et al. 2010). They use an AE with convolutional layers by transforming the light curves into "light-curve images" (see Pasquet-Itam & Pasquet 2018). Rather than interpolate the light curves, Pasquet et al. (2019) apply a mask to filters that are missing data at a certain time. In contrast, we interpolate our light curves but assign interpolated values a large uncertainty of 1 mag, as explained above. We found that the method of transforming light curves into images and masking across four filters led to unstable training and poorer performance. This is likely due to the large data gaps in the *real* PS1-MDS light curves, compared to the high-cadence (2 days for each filter) *simulated* light curves of SPCC. Since the LSST data are expected to more closely resemble the PS1-MDS light curves than the SPCC simulated events, we expect our method to be more robust in a real-life application.

3.3. Supervised Learning: Random Forest Classifier

As a final step, we use the encoded light curves as features for a supervised classification method. Following Villar et al. (2019), we train an RF classifier on the PS1-MDS spectroscopically classified SNe, including the RAENN encodings as features.

In addition to the encoding (eight features), we use the following 36 features based on the GP-interpolated light curves:

1. The *griz* rise times in the rest frame, calculated 1, 2, and 3 mag below peak (12 features).
2. The *griz* decline times in the rest frame, calculated 1, 2, and 3 mag below peak (12 features).
3. The *griz* peak absolute magnitudes (four features).
4. The median *griz* slope between 10 and 30 days post-peak in observer frame. This area was chosen by eye to specifically help the model differentiate between SNe II and SNe Ibc (four features).
5. The integral of *griz* light curves (four features).

We measure these values from the GP-interpolated light curves rather than the decoded light curves. The decoded light curves are, at best, approximations of the GP-interpolated light curves. Therefore, using them would only result in noisier features. The decoded light curves are necessary, however, as a means to train the RAENN to extract the N_E encoding neuron values. We note that for some features, e.g., the rise and decline times, the feature values are heavily dependent on the GP extrapolation in cases where there are no measured data. Including GP errors in the supervised portion of our analysis could help capture this intrinsic uncertainty in the underlying light curve, but we leave that exploration to future work.

These features were chosen through trial and error while optimizing classification accuracy. We find that inclusion of all features leads to our optimal classification accuracy, although we do explore how well our classifier performs with the RAENN features alone in the following section.

We pass these features through an RF classifier, utilizing 350 trees in the RF and the Gini information criterion. The number of trees was determined based on trial-and-error optimization. To counteract the imbalance across the five spectroscopic classes, we tested several algorithms to generate synthetic data to augment our training set. Following Villar et al. (2019), we use a Synthetic Minority Over-sampling Technique (SMOTE; Chawla et al. 2002) and a multivariate Gaussian (MVG) fit. We additionally test using a kernel density estimate of the training set, using a Gaussian kernel with bandwidth equal to 0.2 (or 20% of the whitened feature standard deviation). We find that the MVG with a halved covariance matrix performs best. We test our classifier using leave-one-out cross-validation, in which we remove one SN from the sample, oversample the remaining objects by generating new objects using the MVG, and then apply the trained RF to the single, removed event and recording the result. For each object, our RF reports probabilities
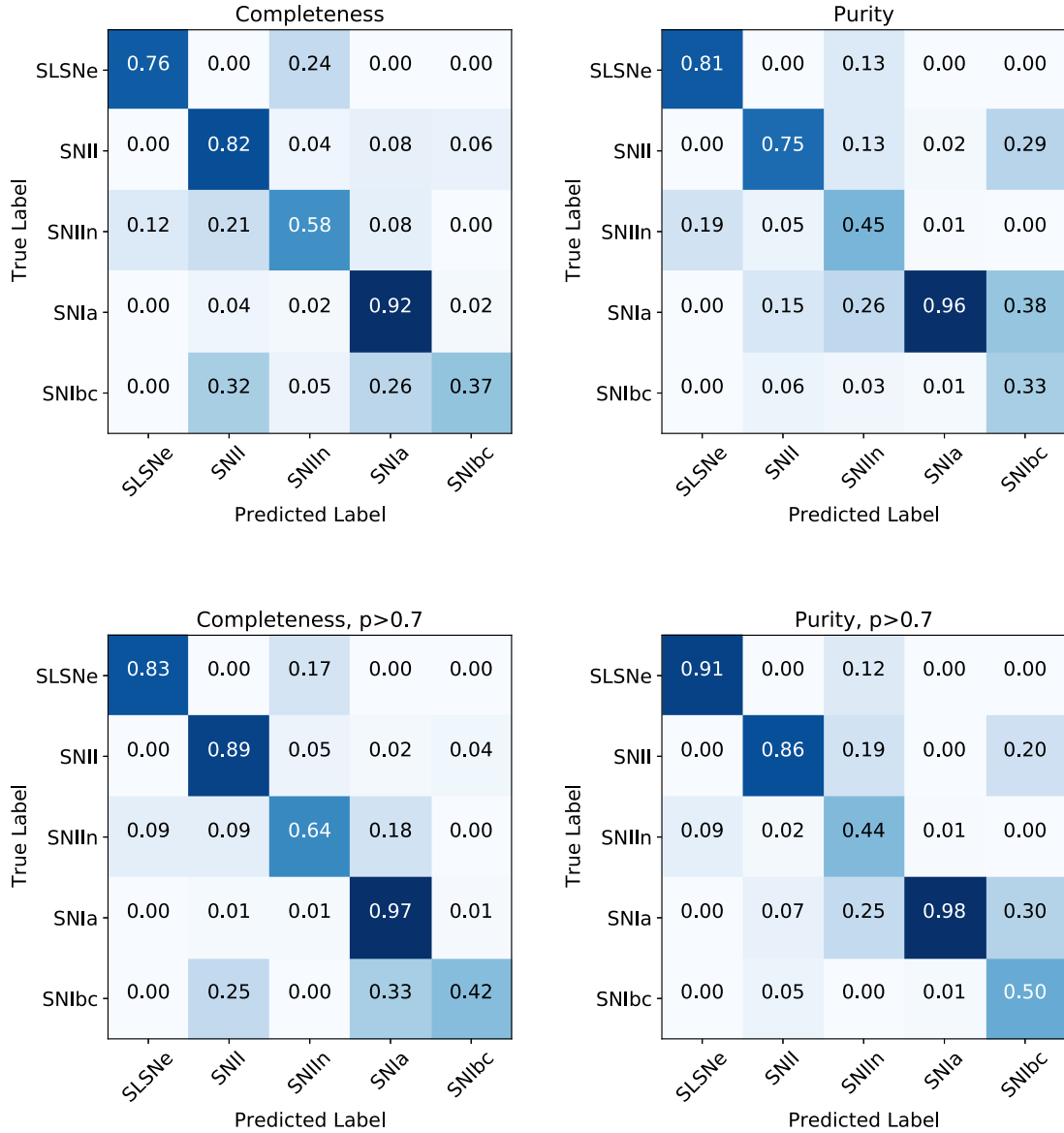
Figure 6. Confusion matrices for the full set of 557 spectroscopically classified SNe. In the bottom panel, we include only objects where the maximum probability is ⩾0.7 (438 events). Left panels: completeness-based confusion matrices, in which each row is normalized to equal 1. Completeness quantifies how much of a spectroscopic class the classifier has correctly classified. Right panels: purity-based confusion matrices, in which each column is normalized to equal 1. Purity quantifies how much a photometric class is composed of the true spectroscopic class. By restricting our classes to the high-confidence objects (bottom panels), both our completeness and purity increase.

associated with each class, which are calculated using the fraction of trees that vote for each class. We take the class with the highest probability as the predicted SN type.

4. Classification Results

There are several metrics to measure the success of a classifier. We focus on three metrics: the purity, completeness, and accuracy. We define the three, calculated for a single class, below:

$$\text{Purity} = \frac{TP}{TP + FP}$$
$$\text{Completeness} = \frac{TP}{TP + FN}$$
$$\text{Accuracy} = \frac{TP + TN}{TS}, \qquad (4)$$

where TP (FP) is the number of true (false) positives, TN (FN) is the number true (false) negatives, and TS is the total sample size. We optimize the hyperparameters of our classifier using the F1-score, defined here as the class-averaged harmonic mean of the purity and completeness.

4.1. Spectroscopic Sample

We visualize the completeness and purity of the spectroscopic sample using confusion matrices in Figure 6. A confusion matrix compares our RAENN label (horizontal axis, in the case of the completeness matrix) with the spectroscopic label (vertical). Results are shown for leave-one-out cross-validation, in which one event is removed from the sample for training and the trained model is applied to the left-out event. As with Villar et al. (2019), we find that our classifier performs best for SNe Ia (92% completeness), SLSNe (76%), and SNe II

(82%) and worst for SNe Ibc (37%). Our class-averaged classification completeness is 69% across the five SN types. This is worse than the performance of Villar et al. (2019), who find a class-averaged completeness of 80%. Our class-averaged purity is 66%, again slightly worse than the average purity of 72% found in Villar et al. (2019). When limiting the sample to only objects in which the classification probability is $\geqslant 0.7$ (a total of 438 objects), we find that our performance increases, with a class-averaged completeness of 75% and a class-averaged purity of 74% with a loss of 20% of the sample size.

Next, we explore the classification confidence reported by our algorithm. The confidence estimates are directly outputted by the RF. With larger data sets, one can calibrate the outputted uncertainties using, e.g., an additional logistic function. Given our small data set, we do not perform any additional calibration. In Figure 7 we plot the cumulative fractions of SNe in our training set, grouped by their spectroscopic and photometric classifications. The majority of high-confidence objects are SNe Ia, with nearly half of the spectroscopic SNe Ia having a confidence ($p > 0.98$). Similarly, half of the SLSNe have high-confidence identifications ($p > 0.8$). SNe Ibc and SNe IIn have the lowest confidence on average, with the majority of events having $p < 0.5$. This is likely reflective of the fact that SNe Ibc and SNe IIn span a wide range of observed properties, including overlap with SNe Ia.

Figure 7 also indicates the misclassified objects. Ideally, we want our misclassifications to largely occur in low-confidence objects. This is the case for SLSNe, SNe Ia, and SNe II. For SNe IIn and SNe Ibc, the misclassifications occur even for high-confidence events. This indicates that for SNe Ia, SNe II, and SLSNe, misclassifications are likely tied to data quality. In contrast, misclassifications of SNe IIn and SNe Ibc seem to be due to intrinsic overlap of the classes in feature space with other SNe (mainly SNe Ia).

We additionally attempt to sort events based on the number of data points, rather than classification confidence (see the right column of Figure 7). Our photometric data set has, on average, fewer $>5\sigma$ data points compared to our spectroscopic data set ($\sim$15 vs. $\sim$30 data points on average). Because of this mismatch and the lack of a strong correlation between number of points and classification confidence, we do not further explore how cutting sparse light curves affects our final classification accuracy.

We next turn our attention to the performance of our classifier when constrained to only data-driven (RAENN) features. Using the same set of RAENN features *without* any additional information, we produce the confusion matrices shown in Figure 8. We find a class-averaged completeness of 53% (and purity of 50%), approximately 20% worse than including the additional features. The overall breakdown is similar to our final confusion matrix, with the worst-performing classes being Type Ibc and Type IIn. We find that our RAENN-only classification is more inclined to label events as SNe Ia, likely a bias from the fact that our SN data set used to train the RAENN is highly dominated by SNe Ia. One could correct for this bias by adding some type of prior on the relative observational rates of the SN classes explored here. If we run our classification algorithm *without* the RAENN features, we find that SuperRAENN performs similarly (slightly worse, with a macro-averaged purity and completeness of 66% and 62%, respectively), implying that the RAENN has not picked up on uniquely helpful features independent from our hand-selected feature set. To be clear, the

intent of RAENN is not necessarily to outperform hand-selected features but to create model-independent features in real time. In this work, we determine final classifications with the RAENN and hand-selected features to provide the highest-confidence photometric classifications. Improvements to classifications based solely on RAENN features are left to future work.

While not optimized for Type Ia versus non-SN Ia classification, we explore how well our classifier (using the full set of features) performs when we collapse the confusion matrix into just two classes. In Figure 9, we show the completeness and purity confusion matrices for Type Ia versus core-collapse SN (CCSN) classifications, finding $\approx$90% completeness and $>$80% purity in both classes.

The RF classifier allows us to measure the relative "importance" of the 44 features used to classify the SNe. We define importance as the decrease in the Gini impurity, which accounts for how often a feature is used to split a node and how often a node is reached in the forest (Breiman et al. 1984). We show the importance of each RF feature in Figure 10, along with the measured importance for a normal random variable. The peak magnitudes and decline rates are the most important features for classification. However, the RAENN features also have significant influence on the final classifications, with two RAENN features appearing in the top 10 important features.

The feature importance unfortunately loses some quantitative meaning if the features are correlated, which is the case with our features. When two features are highly correlated, one may be arbitrarily measured as more important, so the general trends are more meaningful than precise order. We show the magnitudes of the feature correlations in Figure 11 to better understand the underlying correlations. There are clear correlations between features derived in multiple bands (e.g., the peak magnitude in g band is highly correlated to that in r band). However, we also see correlations between the RAENN features and the more traditional light-curve features. About half of the RAENN features seem strongly correlated with the peak magnitudes, while two others seem well correlated with rise and decline times. A more detailed exploration of the physical interpretation of the RAENN feature space may be worthwhile but is beyond the scope of this work.

4.2. Classifying the Complete Photometric Data Set

We apply our trained classification algorithm to the PS1-MDS data set of SN-like transients that pass our quality cuts described in Section 2. We report the probabilities of each class type for each light curve in Table 2. Error bars for each class probability are calculated by running the trained RF classifier 25 times with unique random seeds. We show the class breakdown of the complete photometric set (2885 SNe) in Figure 12. Excluding the spectroscopic sample, we present 2315 new SNe, with 1435 (61.9%) SNe Ia, 459 (19.9%) SNe IIP, 272 (11.7%) SNe Ibc, 112 (4.8%) SNe IIn, and 37 (1.6%) SLSNe. Of these, 1311 are high-confidence ($p > 0.7$) photometric classifications. A cumulative plot of the confidences grouped by each photometric class is shown in Figure 7; the distribution of these probabilities largely matches the spectroscopic sample.

A sample of SNe from each photometric class is shown in Figure 13, including high- and low-probability examples. For the low-probability examples, it seems that even well-sampled light curves can have low confidence scores, likely because the features of their light curves reside on a region of feature space
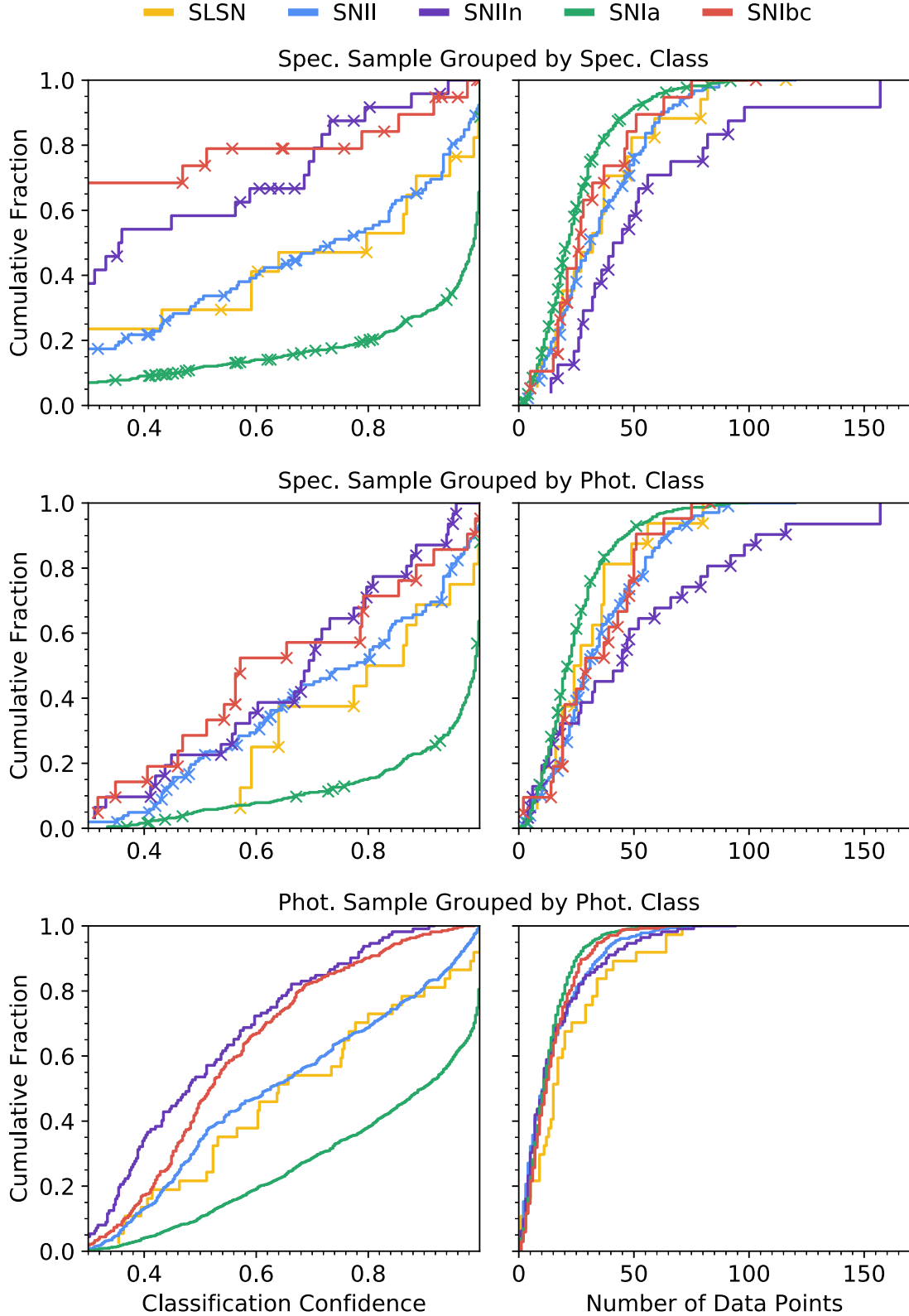
Figure 7. Top: cumulative fraction of the spectroscopic SN sample as a function of classification confidence (left) and number of $>5\sigma$ data points (right), grouped by spectroscopic class. Misclassifications are marked with crosses. Middle: cumulative fraction of the spectroscopic SN sample, grouped by photometrically identified class. As expected, most misclassifications occur at low confidence. At our chosen high-confidence cutoff ($p > 0.7$), we find that the samples are largely pure. Bottom: cumulative fraction of the photometric SN sample, grouped by photometrically identified class. The distributions based on classification confidence follow a similar trend to those seen in the spectroscopic sample, with SNe Ia and SLSNe having the highest fraction of high-confidence events. However, the photometric set has significantly more points on average when compared to the spectroscopic data set.
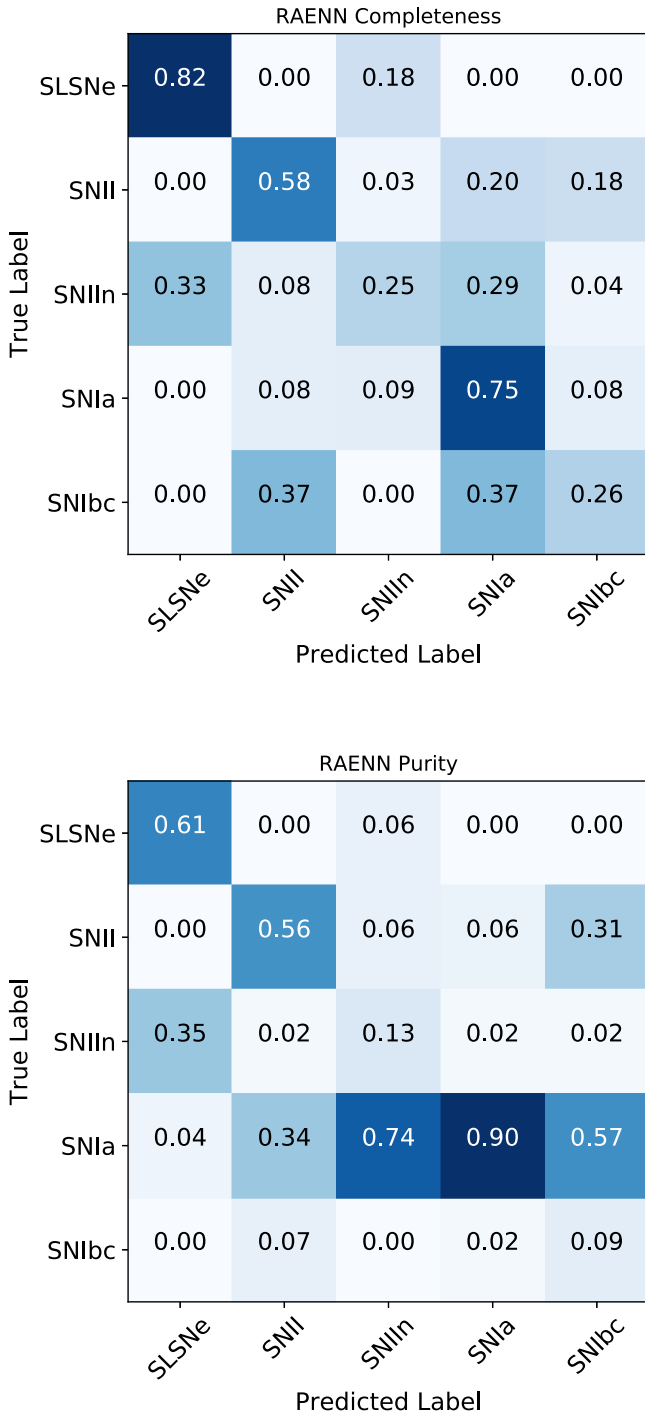
Figure 8. Completeness and purity confusion matrices, generated from classifying the spectroscopic data set using only RAENN features and leave-one-out cross-validation. Even without additional features, the classifier performs similarly to other simulation-based classifiers such as those presented in Muthukrishna et al. (2019a) and Boone (2019).

Figure 9. Confusion matrices for a simpler Type Ia SN vs. non-Ia (CCSN) classification, generated by collapsing the complete confusion matrices.

in which various SN classes reside. This is highlighted in Figure 14 in luminosity–decline space. Several misclassified events (shown as squares) have a high number of data points (represented in misclassified events as marker size).

The redshift distributions of the new, photometrically classified events are roughly consistent with those seen in the spectroscopic sample (Figure 2), with Type Ibc peaking at $z \sim 0.19$, Type II peaking at $z \sim 0.21$, Type Ia and Type IIn peaking at $z \sim 0.42$, and SLSNe peaking at $z \sim 0.58$.

We compare the overall photometric breakdown of SN types to that of the ZTF Bright Transient Survey (Fremling et al. 2020), which spectroscopically classified 761 SNe with peak g- or r-band magnitude of <18.5. Fremling et al. (2020) find that their magnitude-limited survey breaks down into 72% SNe Ia, 16% "normal" SNe II (Type IIP/L), 3% SNe IIn (including their Type IIn and SLSN-II category), 5% SNe Ibc, and 1.6% Type I SLSNe. This is a similar breakdown to that found in our spectroscopic sample. Comparing to our photometric set, we find a slightly higher fraction of SNe II and SNe IIn and a lower fraction of SNe Ia (all within ~20% of the ZTF BTS values), as shown in Figure 12. For our high-confidence ($p > 0.7$) sample (also shown in Figure 12), our class breakdowns are closer to those of our spectroscopic and the ZTF BTS sample, with a slight overabundance of SNe Ia (≈78%). Based on our
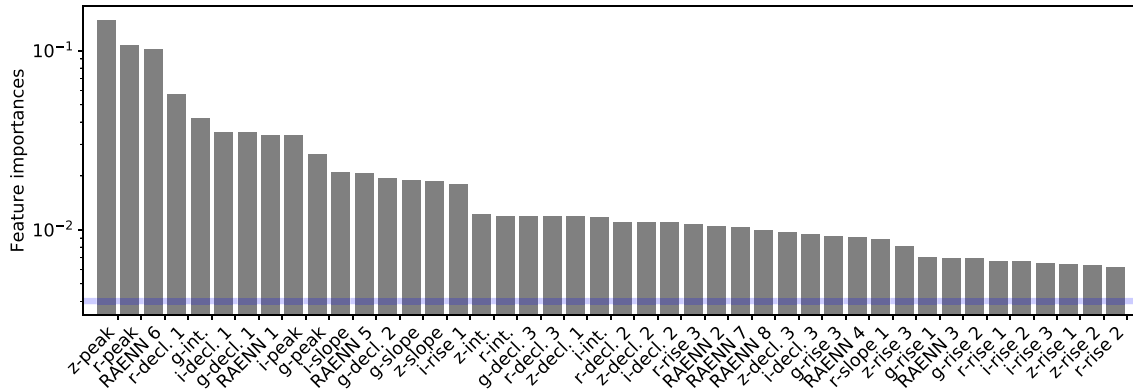
Figure 10. Feature importance (gray). The blue horizontal line shows the importance measure for a normally distributed random variable; features at or below this line can be considered largely unimportant to the final classification. In our case, all features are considered important by the RF.
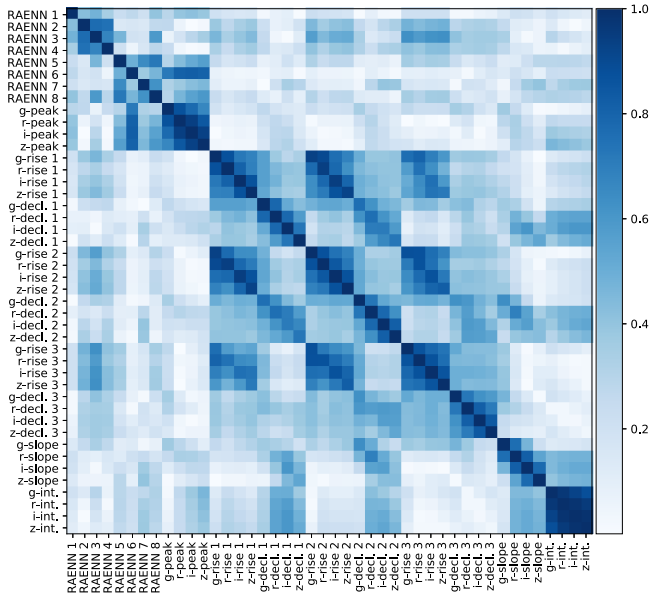
Figure 11. Absolute values of the covariance matrix of the various features used in our classification method, where darker blue represents a stronger absolute correlation. Unsurprisingly, the same features derived from different bands (e.g., the peak g-band flux vs. the peak r-band flux) are highly correlated. The RAENN features are also correlated to the physically motivated parameters, with some being strongly correlated to peak magnitudes, some to rise and decline times, and some to the post-peak slope.

Table 2
SN Classification

Event Name	p_{SLSN}	p_{II}	p_{IIn}	p_{Ia}	p_{Ibc}
PSc000001	$0.00_{0.00}^{0.00}$	$1.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$
PSc000006	$0.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$	$1.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$
PSc000010	$0.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$	$1.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$
PSc000011	$0.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$	$1.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$
PSc000014	$0.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$	$1.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$
PSc000034	$0.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$	$1.00_{0.00}^{0.00}$	$0.00_{0.00}^{0.00}$

Note. Spectroscopically identified SNe have probabilities of 1.

(This table is available in its entirety in machine-readable form.)

understanding of how our classifier performs on the training set, we can understand the biases present (e.g., that some spectroscopic SNe Ibc are classified photometrically as SNe Ia). We can use these known biases, encoded within the

confusion matrices, to correct our class breakdown. Mathematically, this is calculated as the dot product of the purity matrix and our original class breakdown. Applying this correction to the photometric data set, the class breakdown is well aligned with the breakdown of our spectroscopic sample, as shown in Figure 12. This study should *not* be used to rigorously study the observational breakdown of SN classes; however, the fact that our $p > 0.7$ sample is in relatively good agreement with the ZTF BTS provides some evidence that our photometric sample is correctly labeled.

5. Discussion

5.1. Classification of Other Transients

Our algorithm assumes that every SN belongs in one of five classes: SLSNe, SNe II, SNe IIn, SNe Ia, and SNe Ibc. Yet what does our algorithm do for transients that do not fall in these five classes? Here we address this question for a number of spectroscopically classified extragalactic transients. We summarize the photometric classification for these rare transients in Table 3.

Drout et al. (2014) presented a sample of 10 extragalactic transients discovered with PS1-MDS with redshift measurements that rise too rapidly to be powered solely with 56Ni.[24] Following Rest et al. (2018), we refer to these as FELTs. FELTs have a broad range of peak magnitude ($-16 \gtrsim M \gtrsim -20$), which is reflected in the distribution of photometric classifications. Of these 10 objects, six have "high-confidence" ($p > 0.7$) classifications in one of our five categories, four of which are SNe Ia and two of which are SNe II. The other four objects are classified as low-confidence Type Ia (one object), Type II (two objects), and Type Ibc (one object). As expected, the higher-luminosity objects are those classified as Type Ia, while the lower-luminosity objects are classified as Type II. The majority of objects have Type Ibc as their second-highest classification. Based on this analysis, FELTs are likely a (small) contaminant of both SNe II and SNe Ia in our sample, and our algorithm would need to be retrained to specifically classify FELTs.

Two known TDEs were discovered in PS1-MDS: PS1-10jh (PSc040777; Gezari et al. 2012) and PS1-11af (PSc120170; Chornock et al. 2014). Both objects are classified as SNe IIn with $p \sim 0.8$ and $p \sim 0.6$, respectively. This makes intuitive

---

[24] Drout et al. (2014) present an additional four objects that lack a confident redshift estimate (the "bronze" sample), which we exclude from our analysis.
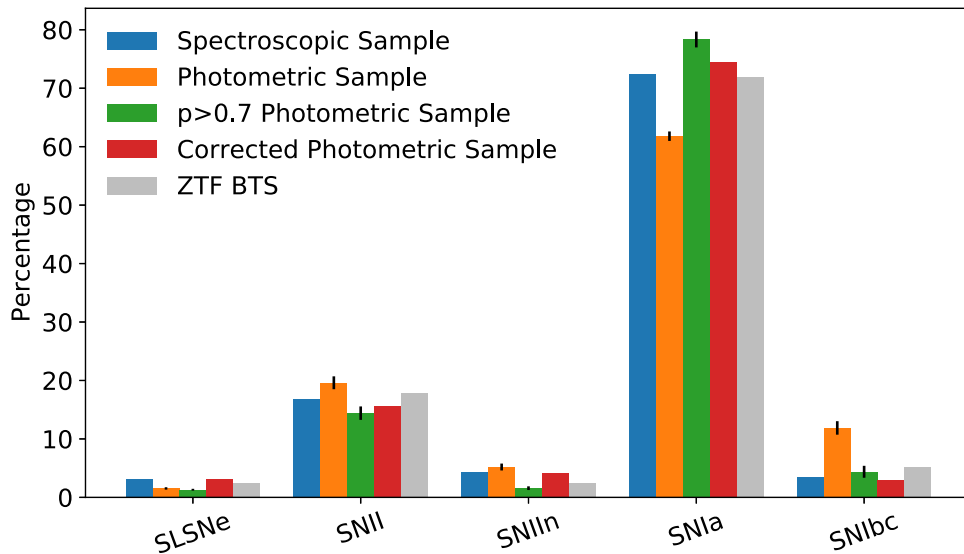
Figure 12. Breakdown of SN subclasses in the spectroscopic and photometric samples. There is a significantly smaller fraction of SNe Ia in our photometric sample (orange) vs. our spectroscopic sample (blue), implying that we have misclassified SNe Ia as CCSNe. If we limit our photometric sample to the high-confidence ($p > 0.7$) events (green), the class breakdowns are better aligned. Using our confusion matrix, we can correct the photometric class breakdown for known biases (see text for details; red), which also better aligns our class breakdowns. Finally, we compare our results to the ZTF Bright Transient Survey, finding good agreement between the spectroscopic class breakdown and corrected photometric class breakdown (Fremling et al. 2020).

sense, as the light curves tend to be long-lived and bright like some SNe IIn. Both objects have Type Ia and Type II as their next most likely classifications. Based on these, it may be possible to search for TDEs in our sample within the photometric Type IIn sample.

We highlight four other SNe that do not fit in our five categories. PS1-10afx (PSc080333) is a lensed SN Ia (Chornock et al. 2013; Quimby et al. 2014), which peaks at -22 mag. We classify PS1-10afx as a high probability ($p \sim 0.9$) SLSN. PS1-12sk (PSc370290) is an SN Ibn (Sanders et al. 2013) that peaks at $M \sim -19$. We classify PS1-12sk as a low-probability Type Ia ($p \sim 0.6$) or Type IIn ($p \sim 0.4$). We classify PS1-12sz (PSc370330) as a likely SN IIb using SNID; PS1-12sz peaks at $M \sim -18.5$. We photometrically classify this object as a low-probability Type Ibc ($p \sim 0.6$). Finally, SN 2009ku (PS0910012) is a spectroscopically identified Type Iax (Narayan et al. 2011) that peaks at $M \sim -18.5$. We classify this object as a low-probability Type Ia ($p \sim 0.5$) or Type Ibc ($p \sim 0.3$).

5.2. Potential Biases

As discussed in Section 2, our spectroscopic sample is somewhat brighter and at a lower redshift than our test set. This difference may introduce biases in our final classifications, although this effect should be minimal considering the small ($\sim$1 mag) difference between the two sets. De-redshifting the SNe removes some of this bias, by removing knowledge of the underlying redshift as a feature. The relative fractions of SN subtypes may evolve with redshift as host properties change (see, e.g., Graur et al. 2017 for an exploration of the correlations between host properties and SN type). Our spectroscopic and photometric sets differ most greatly at $z \gtrsim 0.5$ (see Figure 2). In this redshift range, average host metallicity is not expected to drastically shift (Lilly et al. 2003), implying a small potential bias. A separate bias may arise from the fact that our photometric sample relies on a measured spectroscopic redshift. At higher redshift, our galaxy redshift measurements become increasingly uncertain as dominant

emission lines shift out of the optical band and intrinsically dim hosts fall below our observational limits. In contrast, rest-frame UV features of SNe (especially SLSNe) remain in the optical band, making it easier to confidently measure a distance from SN spectra. In the future, this problem can be mitigated with photometrically derived host galaxy redshifts.

As expected, the relative observed fraction of SN subtypes evolves with redshift owing to the magnitude limit of the survey. We trace this evolution in Figure 15. We show the cumulative fraction (integrating from $z = 0$) of each subclass as a function of redshift. Each subclass peaks in order of luminosity function. The dimmest subclass, SNe II, dominates the sample for $z < 0.3$, peaking near $z \sim 0$.

Using the high-redshift ($z > 0.75$) sample, we can test whether redshift information is playing an unwanted role in our training. The spectroscopic sample at $z \gtrsim 0.75$ is solely made up of SLSNe; however, we do not expect *all* high-z objects to be SLSNe. Given a typical limiting magnitude of $m_{r,lim} \sim 23.3$, the corresponding absolute magnitude is ~ -20 at $z = 0.75$. At this sensitivity, we expect to find SLSNe, SNe IIn, and potentially bright SNe Ia (if the limiting magnitude is slightly deeper). For $z > 0.75$, we find that our photometric sample (a total of 28 SNe) is 68% SLSNe, 18% SNe IIn, and 14% SNe Ia (with all SNe Ia occurring at $z < 0.85$), implying that our classifier has not learned to simply classify all high-z events as SLSNe. The high-z SNe Ia, in particular, have noisy light curves that peak at $M \sim -20$.

5.3. Comparison to Other Works

We first compare our results to H20, which extends the work of Villar et al. (2019) to classify the PS1-MDS photometric sample using features extracted from analytical fits to the light curves. Overall, H20 (and Villar et al. 2019) achieve better performance at the cost of a more computationally expensive feature extraction method. While our algorithm takes roughly seconds to run on new events after training, Superphot takes roughly several minutes to fit an analytical model to the data
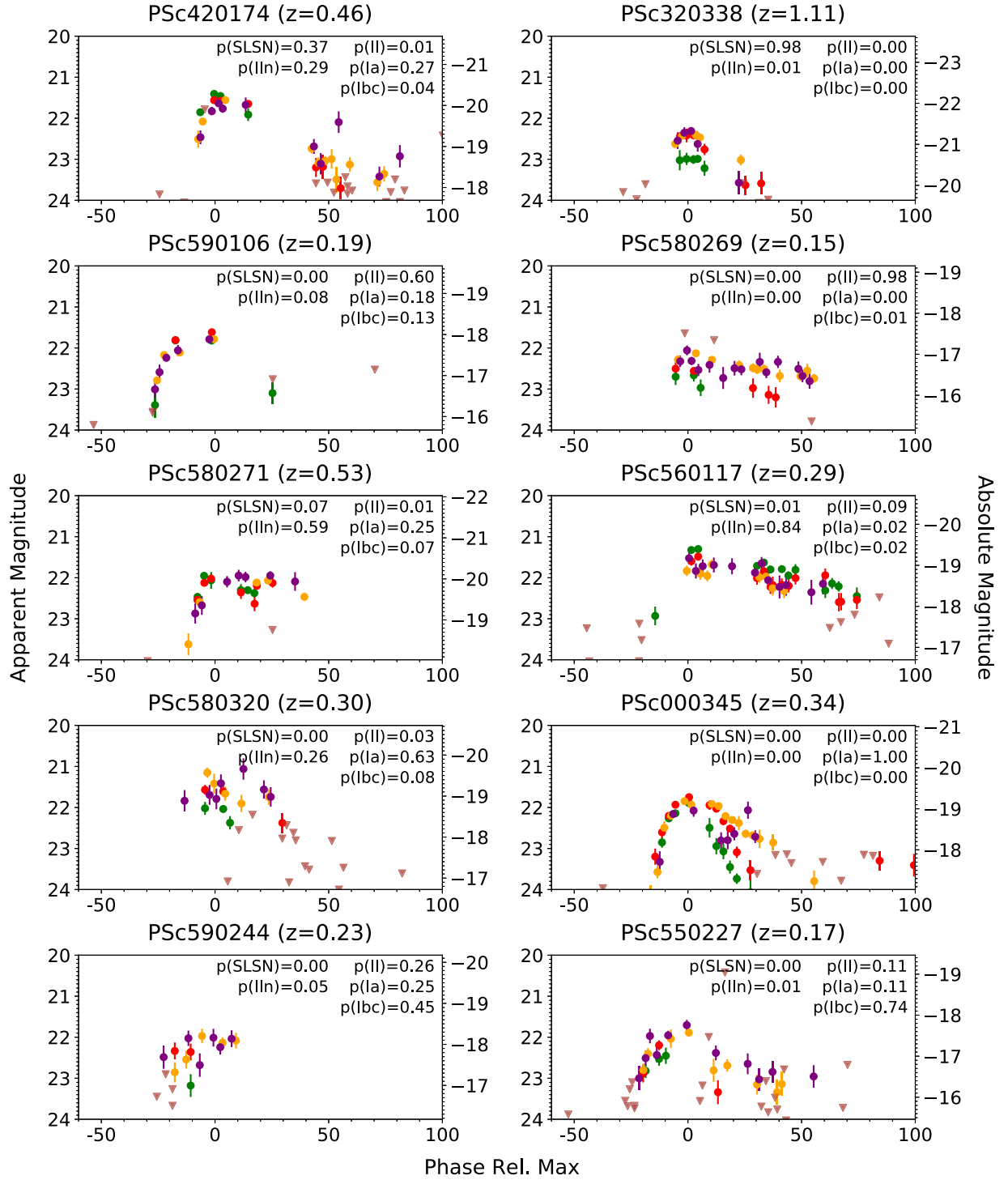
Figure 13. Sample of SNe from our photometric sample, sorted by low (left column) vs. high (right column) confidence and photometrically identified SN class (rows). Here we show only $>3\sigma$ detections and otherwise show magnitudes as upper limits (triangles). Low confidence in classification appears to be due to either poor data quality or confusion between multiple classes.

set. We agree with 74% of the photometric classifications of H20. If we compare the top two labels, the algorithms agree on 95% of classifications. Indeed, often the top two classification choices are flipped for either algorithm, occurring most often with SNe II/Ibc and SNe IIn/Ia. We find stronger agreement if we exclude objects with low classification confidence, namely, using only $p > 0.7$ in both algorithms, our classifications agree 84% of the time (with 1597 objects remaining after the cut, i.e., a loss of ∼50% of the sample). The

agreement increases further for even higher probability cuts of $p > 0.8$ (>0.9), with 88% (92%) agreement with 1249 (888) objects remaining. Most classification disagreements lie in the Type Ibc/IIn categories, which have low-confidence classifications. We find that our algorithm is more likely to classify SNe as Type Ia, likely a bias built into the unsupervised step of training on the complete data set (which is dominated by SNe Ia). A more detailed comparison of these two results is provided in H20.
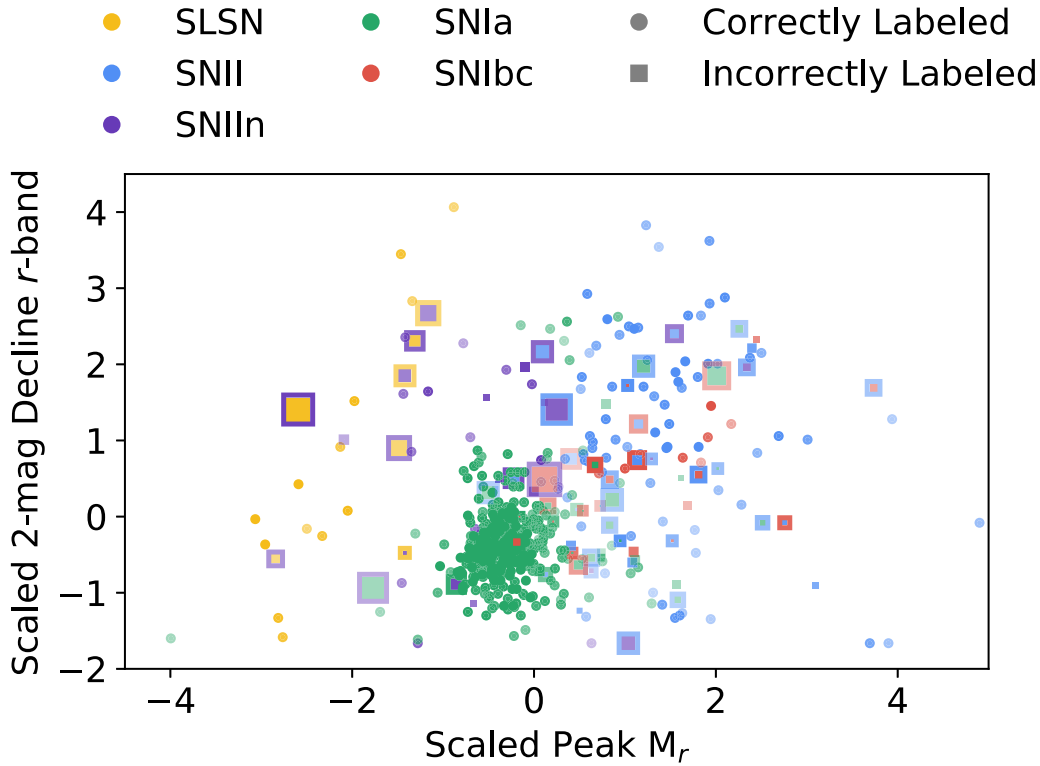
Figure 14. Spread of training set in an example feature subspace (peak magnitude vs. decline time). Points are filled in to represent their spectroscopic label, while the outline color represents the *photometric* classification. For the misclassified events, the opacity represents the classification confidence (with opaque points being most confident), and the marker size represents the number of 3σ data points in the PS1-MDS light curve. Several events that are misclassified with a large number of data points (with and without high confidence) lie in areas of this feature space *not* occupied by members of their spectroscopic class.

Table 3
Rare Transient Classification

Event Name	Spec. Class	References	p_{SLSN}	p_{II}	p_{IIn}	p_{Ia}	p_{Ibc}
PSc040777 (PS1-10jh)	TDE	Gezari et al. (2012)	0.06	0.06	**0.79**	0.06	0.03
PSc120170 (PS1-11af)	TDE	Chornock et al. (2014)	0.12	0.08	**0.62**	0.14	0.04
PSc080333 (PS1-10afx)	Lensed Ia	Chornock et al. (2013)	**0.88**	0.00	0.09	0.02	0.01
PSc370290 (PS1-12sk)	Ibn	Sanders et al. (2013)	0.00	0.00	0.41	**0.57**	0.02
PS0910012 (SN 2009ku)	Iax	Narayan et al. (2011)	0.00	0.13	0.04	**0.55**	0.28
PSc010411 (PS1-10ah)	FELT	Drout et al. (2014)	0.00	**0.74**	0.01	0.12	0.13
PSc091902 (PS1-10bjp)	FELT	Drout et al. (2014)	0.00	**0.56**	0.27	0.14	0.03
PSc150020 (PS1-11qr)	FELT	Drout et al. (2014)	0.00	0.02	0.06	**0.82**	0.09
PSc340012 (PS1-11bbq)	FELT	Drout et al. (2014)	0.03	0.00	0.04	**0.80**	0.13
PSc350224 (PS1-12bb)	FELT	Drout et al. (2014)	0.00	0.28	0.13	0.17	**0.42**
PSc350352 (PS1-12bv)	FELT	Drout et al. (2014)	0.00	0.00	0.00	**0.95**	0.04
PSc440088 (PS1-12brf)	FELT	Drout et al. (2014)	0.00	0.14	0.11	**0.49**	0.25
PSc570006 (PS1-13duy)	FELT	Drout et al. (2014)	0.00	0.02	0.08	**0.79**	0.12
PSc570060 (PS1-13dwm)	FELT	Drout et al. (2014)	0.00	**0.76**	0.03	0.09	0.13
PSc580304 (PS1-13ess)	FELT	Drout et al. (2014)	0.00	**0.51**	0.02	0.37	0.11

Note. The most likely classes are shown in bold.

(This table is available in machine-readable form.)

Villar et al. (2019) discuss the difficulty in comparing our results to the broader literature. In short, previous works have largely focused on Type Ia versus CCSN classification (e.g., Ishida & de Souza 2013; Jones et al. 2017; Brunel et al. 2019) or have been trained and tested on simulated data (e.g., Kessler et al. 2010; Muthukrishna et al. 2019a; Möller & de Boissière 2020). In the case of Type Ia versus CCSN classification, we achieve an accuracy of ≈92%, similar to (but somewhat worse than) specialized classifiers (Jones et al. 2017; Brunel et al. 2019). When comparing to works based on simulated data, we

caution that not all simulated data sets are suitable for multiclass SN classification. In particular, the SNPCC training set (Kessler et al. 2010) lacks the SN diversity necessary to accurately train classifiers and will lead to artificially promising results. PLAsTiCC (The PLAsTiCC team et al. 2018; Kessler et al. 2019) is better suited for this task, and we encourage future work to be built on this data set or the PS1-MDS data set presented here.

We next compare our results to Jones et al. (2017), who presented a PS1-MDS sample of 1169 likely SNe Ia, focusing
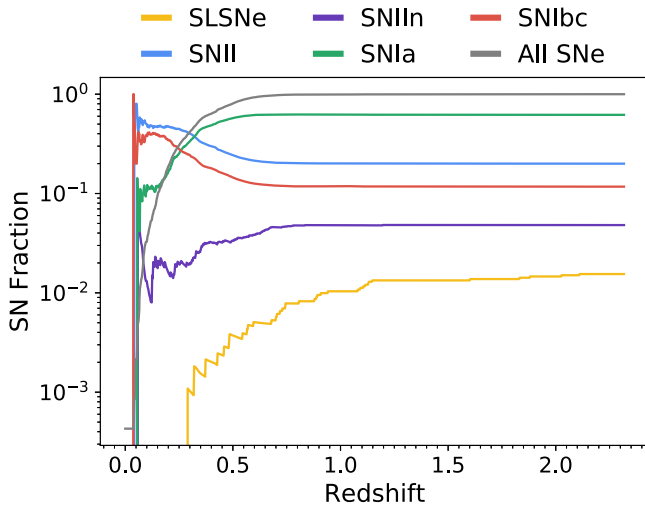
Figure 15. Observed SN subclass cumulative fraction as a function of redshift (colored) and the overall cumulative distribution (gray).

on Type Ia versus non-Ia classification. Jones et al. (2017) used four classification algorithms: the template-matching algorithm PSNID (Sako et al. 2011), a nearest-neighbor approach using the PSNID templates, an algorithm based on fitting light curves to SALT2 templates (Guy et al. 2007), and a method, GALSNID (Foley & Mandel 2013), that only utilizes host galaxy properties. Jones et al. (2017) similarly removed objects with unreliable host redshifts and potential active galactic nucleus hosts, but unlike our analysis, they removed objects at $z > 0.75$. Of their 1169 identified SNe Ia, only 1046 SNe pass our quality cuts to be classified in this work. For these, we find 95% agreement. Of the remaining 48 SNe, we identified Type Ia as the second-highest choice in 24 cases. Of the remaining 24 cases, 15 have low Type Ia probabilities ($p < 0.8$ from Jones et al. 2017) or classification probabilities based entirely on host galaxy. It is worth noting that our classifier, similar to Jones et al. (2017), achieves 96% purity in SNe Ia, making it likely usable for cosmological studies (Jones et al. 2018).

We compare our results to those trained on PLAsTiCC—in particular, Boone (2019), Muthukrishna et al. (2019a), and Gabruseva et al. (2020). These classifiers present average completenesses of ≈0.88 for SLSNe (higher than our score), ≈0.5 for SNe II/IIn (lower than our averaged Type II/IIn score), ≈0.92 for SNe Ia (similar to our score), and ≈0.46 for SNe Ibc (similar to our score given low-number statistics). These results are based on simulated data that lack the complexity of real data, so it is encouraging that our algorithm performs similarly or outperforms these works. It would be interesting and useful to the community to know how these algorithms perform on the PS1-MDS data set, but we leave this for future work.

5.4. RAENN Architecture: Limitations and Benefits

We now turn to the architecture of the RAENN itself and its use in future surveys. The recurrent neurons allow our neural network to generate light-curve features that can be updated in real time, in addition to extrapolating and interpolating light curves. We highlight the accuracy of the RAENN light-curve model as a function of light-curve completeness in Figure 16, parameterized by the rms error. In particular, we train the RAENN model on the available observations up to a certain epoch and then compare the complete model estimate to the

complete light curve. We find that `SuperRAENN` performance drastically improves post-peak, but that it can provide accurate light-curve estimates somewhat before peak. We leave the optimization of `SuperRAENN` to *classify* SNe in real-time data streams to future work. To explore why `SuperRAENN` improves near peak, we track how the RAENN features change as the light curves evolve. In Figure 17, we plot the values of representative encoding values of an SN Ia; we find similar behavior for other types of SNe. The encodings vary smoothly until settling on the final values ∼10 days post-peak. For CCSNe (Type II and Type Ibc), the encoding values also vary smoothly but trend more slowly to the final values.

The ability of the RAENN to extrapolate light curves without built-in physical assumptions allows it to search for anomalous events in real time for the purpose of spectroscopic and multiwavelength follow-up. Given the millions of events expected from LSST, it is essential to search for unexpected or previously unknown physical effects. One concern is that our algorithm is potentially not robust to noisy live-streaming data; in other words, our algorithm must be able to distinguish between anomalous data and noisy data. We check the stability of our encoded values as a function of scaled white noise by adding white noise to a light curve. We then use our RAENN to encode the noisy light curve and record the scatter of the encoded values. We report the results of this test in Figure 18, in which we show the scaled scatter of the encoded values as a function of the magnitude of the injected noise. The scatter grows linearly with noise; however, even with 1 mag of scatter added to the light curve, the overall scatter of the encoded values only increases to 30% of the overall spread of class's features. This implies that the RAENN is largely robust to noise.

Several steps need to be taken to allow our architecture to work on streaming data. First, we use phases relative to maximum light, which will be unavailable during the rise of the SN. A shift to a time measurable early in the light curve, like time of first detection, will allow the RAENN to otherwise perform as designed. Similarly, the features utilized during the supervised portion of our classifier rely on the full light curve being available. All features can be estimated from extrapolated RAENN light curves, or a new set of features may be used on streaming data. Finally, although not necessary, our RAENN could output uncertainties on the SN light curves by converting the network into a *variational* AE, which is designed to simultaneously find an encoding space and uncertainties on the encoded data. This more complex architecture would likely require a larger training set to be reliable. Finally, we note that an algorithm like RAENN could be used in conjunction with an active oracle (a software that recommends new observations to improve classification) such as REFITT (Sravan et al. 2020), in order to actively optimize classification accuracy.

Finally, we note that we build the architecture and `Super-RAENN` pipeline assuming that measured redshift will be available. In practice, LSST will build an archive of redshift estimates as template images become more sensitive and photo-z algorithms become more accurate. We test how reliant our method is on redshift estimates by training a model (with the same hyperparameters as our final pipeline) in which all estimates are assumed to be equal (equivalent to working in flux space rather than luminosity space). We find that our algorithm performs, unsurprisingly, worse without redshift information, obtaining a macro-averaged completeness of 0.50
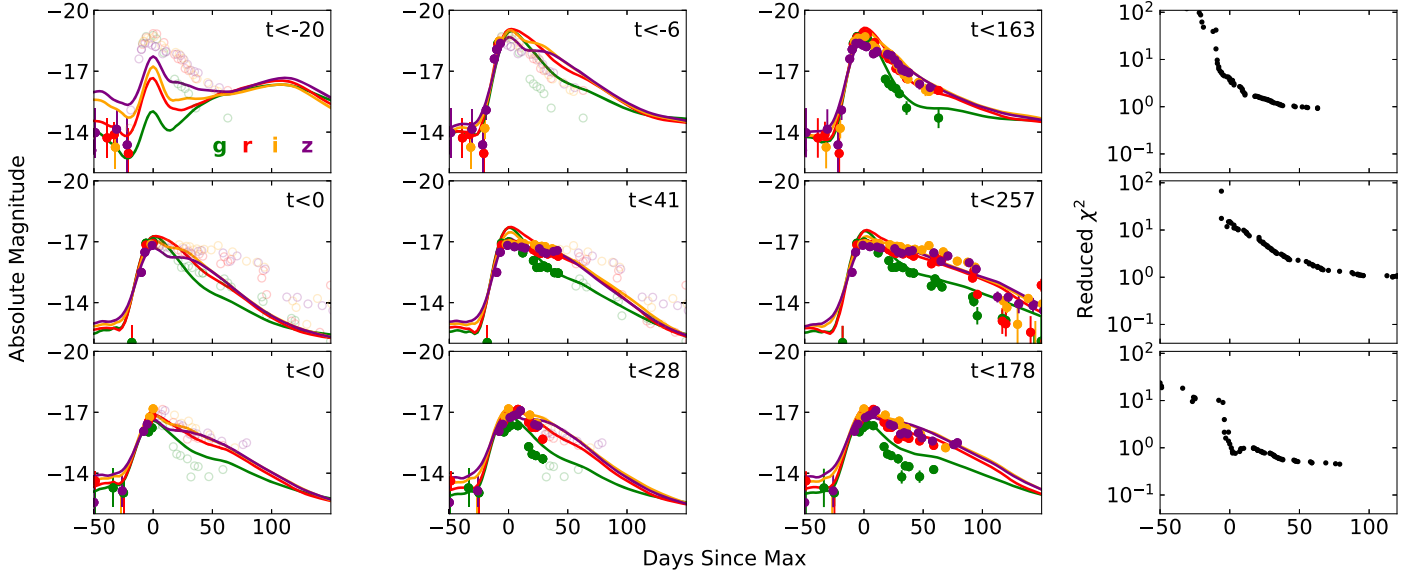
Figure 16. Examples of an SN Ia (top row), SN II (middle row), and SN Ibc (bottom row). Filled points represent observations used to generate the RAENN model (colored lines), while open points are the full data set to guide the eye. In the rightmost column, we show the rms error as a function of SN phase, as more data are being included in the RAENN model. Interestingly, the rms reaches ∼1 near peak for all SNe shown. We emphasize that the RAENN model has been optimized to classify complete SN light curves rather than partial light curves.
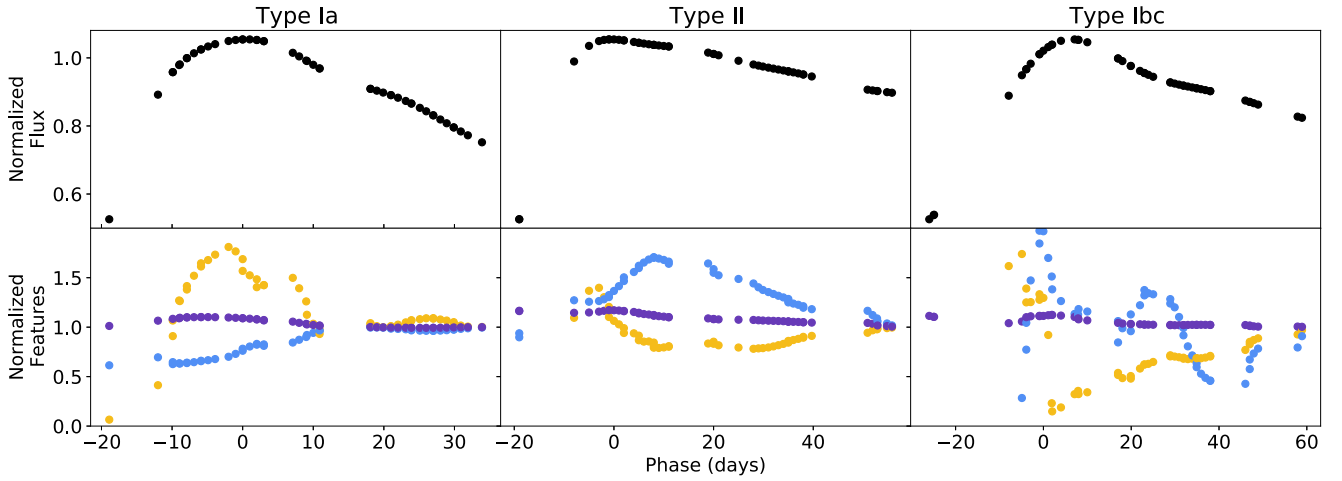
Figure 17. Top: normalized, GP-interpolated *r*-band light curves of spectroscopically classified SNe Ia, SNe II, and SNe Ibc. Bottom: representative set of three (orange, purple, blue) normalized AE features as a function of SN phase. To generate these features, we run the light-curve data through the RAENN up to a certain phase. As shown, the values vary smoothly and then settle on the final values about 1 week post-peak for SNe Ia, but they vary for slightly longer for non-Ia SNe.
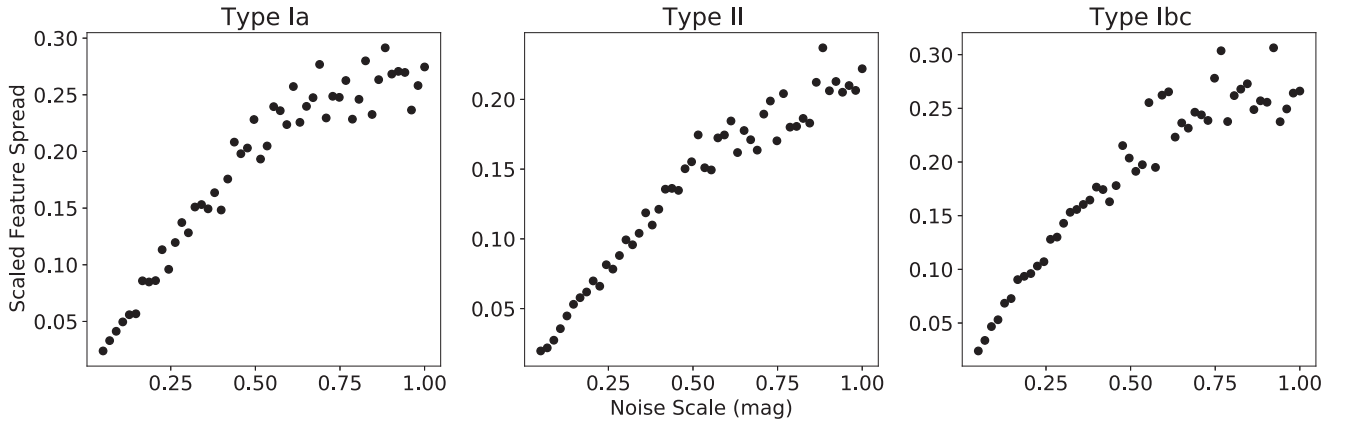
Figure 18. Average spread of the RAENN features for spectroscopic SNe Ia, SNe II, and SNe Ibc as a function of light-curve noise. For every noise scale, we run 100 simulations, adding random noise to the light curve. We then track the average spread of each parameter. We scale this spread by the total spread in the Type Ia class. Even with an injected error of 1.0 mag, the spread in the RAENN feature space only reaches 30% of the total spread throughout the Type Ia class in feature space, implying that the method is robust to noise.

and purity of 0.49 across the five classes. The highest-performing classes are Type Ia and Type Ibc (which are similar to our redshift-dependent pipeline). The pipeline performs significantly worse on SLSNe and SNe II without redshift information.

6. Conclusions

Deep learning-based classifiers are becoming increasingly important for classification of archival SN light curves. In this paper we present a novel, semisupervised approach to light-curve classification, which utilizes spectroscopically labeled and unlabeled SN data from the PS1-MDS. Our key conclusions are as follows:

1. We present the light curves of 5243 SN-like events discovered with PS1-MDS.
2. We present the spectroscopic classifications of 557 SNe, including 17 Type I SLSNe, 93 SNe II, 24 SNe IIn, 404 SNe Ia, and 19 SNe Ibc.
3. We measure and report the spectroscopic redshifts for 2885 SN-like events used in our unsupervised training set.
4. We present a new, open-source photometric classification algorithm, `SuperRAENN`. `SuperRAENN` uses a semi-supervised approach and novel neural network architecture to classify irregularly sampled SN light curves.
5. Using `SuperRAENN`, we extract learned, nonlinear features from the sparse light curves. We use these features and others to classify the complete set of 2885 SN-like objects in the PS1-MDS data set with host galaxy redshifts.
6. We achieve high (87%) accuracy for our spectro-scopically labeled sample. We find best performance for SLSNe, SNe Ia, and SNe II owing to their distinctive regions of feature space. We find worst performance for SNe Ibc, likely due to the small sample size (just 19 events) and their significant overlap with SNe Ia and the subset of rapidly declining SNe II (formerly IIL).
7. Compared to previous studies, we find that our general classifier performs as well as or can outperform classifiers trained on synthetic data sets.
8. We perform simple tests for classification bias and method robustness to noise, finding our method robust to both.

In addition to these key results, we highlight several lessons learned from this study. We find that both Type IIn and Type Ibc classes suffer from poor accuracy, likely due to substantial overlap with SNe Ia in feature space. This finding has also been shown in Villar et al. (2019) and H20, implying that this is a general problem for classifiers. Additionally, rare transients, e.g., FELTs, abnormal Type Ia classes, etc., can be hidden as high-confidence events in another class *or* low-confidence events across several classes. Adapting preexisting classifiers to new classes should be taken on a case-by-case basis. Finally, we find that a mixture of hand-selected and data-driven (in our case, RAENN) features can improve classification accuracy, but hand-selected features seem to generally outperform data-driven features.

Finally, we note that several modifications to our presented classifier will allow it to work with live, rather than archival, data streams such as ZTF and LSST. We perform simple tests and find that our classifier performs optimally around peak, although we have not optimized for this purpose. Finally, the RAENN architecture may also be utilized to search for anomalous events in real time. We plan to explore this in future work.

Facilities: ADS, NED, PS1, TNS.

Software: Astropy (Astropy Collaboration et al. 2018), extinction (Barbary 2016), keras (Chollet 2015), Matplotlib (Hunter 2007), NumPy (Oliphant 2006), RVSAO (Kurtz & Mink 1998), scikit-learn (Pedregosa et al. 2012), SciPy (Virtanen et al. 2020).

ORCID iDs

V. Ashley Villar ● https://orcid.org/0000-0002-5814-4061
Griffin Hosseinzadeh ● https://orcid.org/0000-0002-0832-2974
Edo Berger ● https://orcid.org/0000-0002-9392-9681

Michelle Ntampaka ⬤ https://orcid.org/0000-0002-0144-387X

David O. Jones ⬤ https://orcid.org/0000-0002-6230-0151

Ryan Chornock ⬤ https://orcid.org/0000-0002-7706-5668

Maria R. Drout ⬤ https://orcid.org/0000-0001-7081-0082

Robert P. Kirshner ⬤ https://orcid.org/0000-0002-1966-3942

Ragnhild Lunnan ⬤ https://orcid.org/0000-0001-9454-4639

Raffaella Margutti ⬤ https://orcid.org/0000-0003-4768-7586

Dan Milisavljevic ⬤ https://orcid.org/0000-0002-0763-3885

Yen-Chen Pan ⬤ https://orcid.org/0000-0001-8415-6720

Armin Rest ⬤ https://orcid.org/0000-0002-4410-5387

Daniel M. Scolnic ⬤ https://orcid.org/0000-0002-4934-5849

Eugene Magnier ⬤ https://orcid.org/0000-0002-7965-2815

Nigel Metcalfe ⬤ https://orcid.org/0000-0001-9034-4402

Richard Wainscoat ⬤ https://orcid.org/0000-0002-1341-0952

Christopher Waters ⬤ https://orcid.org/0000-0003-1989-4879

References

Abadi, M., Barham, P., Chen, J., et al. 2016, arXiv:1603.04467
Ade, P. A., Aghanim, N., Armitage-Caplan, C., et al. 2014, A&A, 571, A16
Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2007, ApJS, 172, 634
Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2015, ITPAM, 38, 252
Astropy Collaboration, Price-Whelan, A. M., Sipöcz, B. M., et al. 2018, AJ, 156, 123
Balestra, I., Mainieri, V., Popesso, P., et al. 2010, A&A, 512, A12
Barbary, K. 2016, extinction v0.3.0, Zenodo, doi:10.5281/zenodo.804967
Becker, A. 2015, HOTPANTS: High Order Transform of PSF ANd Template Subtraction, Astrophysics Source Code Library, ascl:1504.004
Blake, C., Brough, S., Couch, W., et al. 2008, A&G, 49, 5.19
Blondin, S., & Tonry, J. L. 2007, ApJ, 666, 1024
Boone, K. 2019, AJ, 158, 257
Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. 1984, Classification and Regression Trees (Boca Raton, FL: CRC Press)
Bronder, T. J., Hook, I. M., Astier, P., et al. 2008, A&A, 477, 717
Brunel, A., Pasquet, J., PASQUET, J., et al. 2019, Electronic Imaging, 2019, 90
Cannon, R., Drinkwater, M., Edge, A., et al. 2006, MNRAS, 372, 425
Cappellaro, E., Benetti, S., Pastorello, A., et al. 2012a, CBET, 3274, 1
Cappellaro, E., Benetti, S., Pastorello, A., et al. 2012b, CBET, 3311, 1
Challis, P., Foley, R. J., Chornock, R., et al. 2010, CBET, 2350, 1
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
Charnock, T., & Moss, A. 2017, ApJL, 837, L28
Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2002, Journal of Artificial Intelligence Research, 16, 321
Cho, K., Van Merriënboer, B., Gulcehre, C., et al. 2014, arXiv:1406.1078
Chollet, F. 2015, keras, https://github.com/fchollet/keras, GitHub
Chomiuk, L., Chornock, R., Soderberg, A. M., et al. 2011, ApJ, 743, 114
Chornock, R., Berger, E., Gezari, S., et al. 2014, ApJ, 780, 44
Chornock, R., Berger, E., Rest, A., et al. 2013, ApJ, 767, 162
Chornock, R., Huber, M., Foley, R. J., et al. 2010a, CBET, 2414, 1
Chornock, R., Huber, M., Rest, A., et al. 2010b, ATel, 2680, 1
Chornock, R., Marion, G. H., Calkins, M., et al. 2012, CBET, 3066, 1
Colless, M., Peterson, B. A., Jackson, C., et al. 2003, arXiv:astro-ph/0306581
Cowie, L. L., Barger, A. J., & Hu, E. M. 2010, ApJ, 711, 928
Drake, A. J., Djorgovski, S. G., Graham, M. J., et al. 2013, CBET, 3451, 1
Dressler, A., & Gunn, J. E. 1992, ApJS, 78, 1
Drinkwater, M. J., Jurek, R. J., Blake, C., et al. 2010, MNRAS, 401, 1429
Drout, M. R. 2016, PhD thesis, Harvard Univ.
Drout, M. R., Chornock, R., Soderberg, A. M., et al. 2014, ApJ, 794, 23
Fabricant, D., Fata, R., Roll, J., et al. 2005, PASP, 117, 1411
Ferrante, F. V., Silverman, J. M., Vinko, J., et al. 2013, CBET, 3413, 1
Filippenko, A. V. 1997, ARA&A, 35, 309
Finkelstein, S. L., Cohen, S. H., Malhotra, S., et al. 2009, ApJL, 703, L162
Foley, R. J., & Mandel, K. 2013, ApJ, 778, 167
Fraser, M., Kotak, R., Wright, D., et al. 2012, CBET, 3171, 1
Fremling, U., Miller, A., Sharma, Y., et al. 2020, ApJ, 895, 32
Frontera-Pons, J., Sureau, F., Bobin, J., & Le Floc'h, E. 2017, A&A, 603, A60
Gabruseva, T., Zlobin, S., & Wang, P. 2020, JAI, 9, 2050005
Garcet, O., Gandhi, P., Gosset, E., et al. 2007, A&A, 474, 473

Gezari, S., Chornock, R., Rest, A., et al. 2012, Natur, 485, 217
Graur, O., Bianco, F. B., Huang, S., et al. 2017, ApJ, 837, 120
Guillochon, J., Parrent, J., Kelley, L. Z., & Margutti, R. 2017, ApJ, 835, 64
Guy, J., Astier, P., Baumont, S., et al. 2007, A&A, 466, 11
Harutyunyan, A. H., Pfahler, P., Pastorello, A., et al. 2008, A&A, 488, 383
Hewett, P. C., & Wild, V. 2010, MNRAS, 405, 2302
Hillebrandt, W., & Niemeyer, J. C. 2000, ARA&A, 38, 191
Hosseinzadeh, G., Dauphin, F., Villar, V. A., et al. 2020, ApJ, 905, 93
Howell, D. A., Sullivan, M., Perrett, K., et al. 2005, ApJ, 634, 1190
Hunter, J. D. 2007, CSE, 9, 90
Im, M., Faber, S. M., Gebhardt, K., et al. 2001, AJ, 122, 750
Ishida, E. E., & de Souza, R. S. 2013, MNRAS, 430, 509
Ivezic, Z. & LSST Science Collaboration 2010, The LSST System Science Requirements Document (v5.1.1)
Jedicke, R., Tonry, J., Veres, P., et al. 2012, AAS/Division for Planetary Sciences Meeting Abstracts, 44, 210.12
Jones, D., Scolnic, D., Foley, R., et al. 2019, ApJ, 881, 19
Jones, D., Scolnic, D., Riess, A., et al. 2017, ApJ, 843, 6
Jones, D., Scolnic, D., Riess, A., et al. 2018, ApJ, 857, 51
Jones, D. H., Read, M. A., Saunders, W., et al. 2009, MNRAS, 399, 683
Kaiser, N., Burgett, W., Chambers, K., et al. 2010, Proc. SPIE, 7733, 77330E
Karhunen, K., Kotilainen, J. K., Falomo, R., & Bettoni, D. 2014, MNRAS, 441, 1802
Kasliwal, M. M. 2012, PASA, 29, 482
Kessler, R., Bassett, B., Belov, P., et al. 2010, PASP, 122, 1415
Kessler, R., Narayan, G., Avelino, A., et al. 2019, PASP, 131, 094501
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Kramer, M. A. 1991, AIChE, 37, 233
Kulkarni, S. 2018, ATel, 11266, 1
Kurtz, M. J., & Mink, D. J. 1998, PASP, 110, 934
Lamareille, F., Brinchmann, J., Contini, T., et al. 2009, A&A, 495, 53
Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, A&A, 439, 845
Lilly, S. J., Carollo, C. M., & Stockton, A. N. 2003, ApJ, 597, 730
Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, ApJS, 184, 218
Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, ApJS, 172, 70
Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, ApJS, 225, 31
Lucas, J., Calef, B., & Kyono, T. 2018, in The Advanced Maui Optical and Space Surveillance Technologies Conf., ed. S. Ryan (Kihei, HI: AMOS), 51
Lunnan, R., Chornock, R., Berger, E., et al. 2014, ApJ, 787, 138
Lunnan, R., Chornock, R., Berger, E., et al. 2018, ApJ, 852, 81
Ma, Z., Zhu, J., Li, W., & Xu, H. 2019, ApJS, 240, 34
Magnier, E. A., Chambers, K. C., Flewelling, H. A., et al. 2020a, ApJS, 251, 3
Magnier, E. A., Schlafly, E. F., Finkbeiner, D. P., et al. 2020b, ApJS, 251, 6
Mink, D. J., Wyatt, W. F., Caldwell, N., et al. 2007, in ASP Conf. Ser. 376, adass XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco, CA: ASP), 249
Möller, A., & de Boissière, T. 2020, MNRAS, 491, 4277
Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019a, PASP, 131, 118002
Muthukrishna, D., Parkinson, D., & Tucker, B. E. 2019b, ApJ, 885, 85
Narayan, G., Foley, R., Berger, E., et al. 2011, ApJL, 731, L11
Naul, B., Bloom, J. S., Pérez, F., & van der Walt, S. 2018, NatAs, 2, 151
Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5
Nicholl, M., Guillochon, J., & Berger, E. 2017, ApJ, 850, 55
Norris, R. P., Afonso, J., Appleton, P. N., et al. 2006, AJ, 132, 2409
Ntampaka, M., Avestruz, C., Boada, S., et al. 2019, arXiv:1902.10159
Oliphant, T. E. 2006, A Guide to NumPy (USA: Trelgol Publishing)
Owen, F. N., & Morrison, G. E. 2009, ApJS, 182, 625
Pasquet-Itam, J., & Pasquet, J. 2018, A&A, 611, A97
Pasquet, J., Pasquet, J., Chaumont, M., & Fouchez, D. 2019, A&A, 627, A21
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2012, arXiv:1201.0490
Pojmanski, G. 2002, arXiv:astro-ph/0210283
Portillo, S. K., Parejko, J. K., Vergara, J. R., & Connolly, A. J. 2020, arXiv:2002.10464
Quimby, R. M., Aldering, G., Wheeler, J. C., et al. 2007, ApJL, 668, L99
Quimby, R. M., Oguri, M., More, A., et al. 2014, Sci, 344, 396
Ralph, N. O., Norris, R. P., Fang, G., et al. 2019, PASP, 131, 108011
Rest, A., Garnavich, P., Khatami, D., et al. 2018, NatAs, 2, 307
Rest, A., Narayan, G., Berger, E., et al. 2009, CBET, 2012, 1
Rest, A., Scolnic, D., Foley, R. J., et al. 2014, ApJ, 795, 44
Rest, A., Stubbs, C., Becker, A. C., et al. 2005, ApJ, 634, 1103
Revsbech, E. A., Trotta, R., & van Dyk, D. A. 2018, MNRAS, 473, 3969
Richards, J. W., Homrighausen, D., Freeman, P. E., Schafer, C. M., & Poznanski, D. 2012, MNRAS, 419, 1121
Ross, N. P., Shanks, T., Cannon, R. D., et al. 2008, MNRAS, 387, 1323

Rovilos, E., Fotopoulou, S., Salvato, M., et al. 2011, A&A, 529, A135
Rumelhart, D. E., Hinton, G. E., Williams, R. J., et al. 1986, Natur, 323, 533
Sako, M., Bassett, B., Connolly, B., et al. 2011, ApJ, 738, 162
Sanders, N. E., Soderberg, A. M., Foley, R. J., et al. 2013, ApJ, 769, 39
Sanders, N. E., Soderberg, A. M., Gezari, S., et al. 2015, ApJ, 799, 208
Scarlata, C., Colbert, J., Teplitz, H. I., et al. 2009, ApJL, 704, L98
Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103
Scodeggio, M., Guzzo, L., Garilli, B., et al. 2018, A&A, 609, A84
Scolnic, D. M., Jones, D., Rest, A., et al. 2018, ApJ, 859, 101
Shappee, B., Prieto, J., Stanek, K., et al. 2014, AAS Meeting, 223, 236.03
Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, AJ, 146, 32
Smith, A. G., Hopkins, A. M., Hunstead, R. W., & Pimbblet, K. A. 2012,
    MNRAS, 422, 25
Smith, N., Mauerhan, J. C., & Prieto, J. L. 2014, MNRAS, 438, 1191
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757
Sravan, N., Milisavljevic, D., Reynolds, J. M., Lentner, G., & Linvill, M. 2020,
    ApJ, 893, 127
Stalin, C. S., Petitjean, P., Srianand, R., et al. 2010, MNRAS, 401, 294
Stubbs, C. W., Doherty, P., Cramer, C., et al. 2010, ApJS, 191, 376
Sullivan, M., Le Borgne, D., Pritchet, C., et al. 2006, ApJ, 648, 868

Szokoly, G. P., Bergeron, J., Hasinger, G., et al. 2004, ApJS, 155, 271
Tajer, M., Polletta, M., Chiappetti, L., et al. 2007, A&A, 467, 73
Takats, K., Pignata, G., Cappellaro, E., et al. 2013, CBET, 3728, 1
The PLAsTiCC team, Allam, T. J., Bahmanyar, A., et al. 2018, arXiv:1810.00001
Tonry, J., & Davis, M. 1979, AJ, 84, 1511
Tonry, J. L., Stubbs, C. W., Lykke, K. R., et al. 2012, ApJ, 750, 99
Trump, J. R., Impey, C. D., Elvis, M., et al. 2009, ApJ, 696, 1195
Uomoto, A., & Kirshner, R. 1985, A&A, 149, 7
Villar, V. 2020, superRAENN, Zenodo, v1.0, doi:10.5281/zenodo.3968715
Valenti, S., Fraser, M., Young, D., et al. 2010, CBET, 2214, 1
Valenti, S., Graham, M. L., Howell, D. A., et al. 2013, CBET, 3458, 1
Villar, V. A., Berger, E., Metzger, B. D., & Guillochon, J. 2017, ApJ, 849, 70
Villar, V. A., Berger, E., Miller, G., et al. 2019, arXiv:1905.07422
Villar, V. A., Hosseinzadeh, G., Berger, E., et al. 2020, Light Curves of Pan-
    STARRS1 SN-like Transients, Zenodo, v1.0., doi:10.5281/zenodo.3974950
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Waters, C. Z., Magnier, E. A., Price, P. A., et al. 2020, ApJS, 251, 4
Wen, Z. L., & Han, J. L. 2015, ApJ, 807, 178
Young, D., Valenti, S., Dunstall, P., et al. 2010, ATel, 2667, 1
Zheng, W., Yuan, F., Vinko, J., et al. 2010, CBET, 2297, 1