The Astrophysical Journal, 884:83 (15pp), 2019 October 10 https://doi.org/10.3847/1538-4357/ab418c

Supernova Photometric Classification Pipelines Trained on Spectroscopically Classified Supernovae from the Pan-STARRS1 Medium-deep Survey

V. A. Villar[1], E. Berger[1], G. Miller[1], R. Chornock[2], A. Rest[3], D. O. Jones[4], M. R. Drout[5], R. J. Foley[6], R. Kirshner[1,7], R. Lunnan[8], E. Magnier[9], D. Milisavljevic[10], N. Sanders[11], and D. Scolnic[12]
[1] Center for Astrophysics / Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138-1516, USA; vvillar@cfa.harvard.edu
[2] Astrophysical Institute, Department of Physics and Astronomy, 251B Clippinger Lab, Ohio University, Athens, OH 45701-2942, USA
[3] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[4] Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 92064, USA
[5] Department of Astronomy and Astrophysics, University of Toronto, 50 George Street, Toronto, ON M5S 3H4, Canada
[6] Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064, USA
[7] Gordon and Betty Moore Foundation, 1661 Page Mill Road, Palo Alto, CA 94028, USA
[8] The Oskar Klein Centre & Department of Astronomy, Stockholm University, AlbaNova, SE-106 91 Stockholm, Sweden
[9] Institute for Astronomy, University of Hawaii at Manoa, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
[10] Department of Physics and Astronomy, Purdue University, 525 Northwestern Avenue, West Lafayette, IN 47906, USA
[11] WarnerMedia Applied Analytics, 535 Boylston Street., Boston, MA 02116, USA
[12] Department of Physics, Duke University, 120 Science Drive, Durham, NC 27708, USA
Received 2019 May 17; revised 2019 September 2; accepted 2019 September 3; published 2019 October 14

Abstract

Photometric classification of supernovae (SNe) is imperative as recent and upcoming optical time-domain surveys, such as the Large Synoptic Survey Telescope (LSST), overwhelm the available resources for spectrosopic follow-up. Here we develop a range of light curve (LC) classification pipelines, trained on 513 spectroscopically classified SNe from the Pan-STARRS1 Medium-Deep Survey (PS1-MDS): 357 Type Ia, 93 Type II, 25 Type IIn, 21 Type Ibc, and 17 Type I superluminous SNe (SLSNe). We present a new parametric analytical model that can accommodate a broad range of SN LC morphologies, including those with a plateau, and fit this model to data in four PS1 filters ($g_{P1}r_{P1}i_{P1}z_{P1}$). We test a number of feature extraction methods, data augmentation strategies, and machine-learning algorithms to predict the class of each SN. Our best pipelines result in ≈90% average accuracy, ≈70% average purity, and ≈80% average completeness for all SN classes, with the highest success rates for SNe Ia and SLSNe and the lowest for SNe Ibc. Despite the greater complexity of our classification scheme, the purity of our SN Ia classification, ≈95%, is on par with methods developed specifically for Type Ia versus non-Type Ia binary classification. As the first of its kind, this study serves as a guide to developing and training classification algorithms for a wide range of SN types with a purely empirical training set, particularly one that is similar in its characteristics to the expected LSST main survey strategy. Future work will implement this classification pipeline on ≈3000 PS1/MDS LCs that lack spectroscopic classification.

Key words: supernovae: general – surveys – techniques: photometric

1. Introduction

Optical time-domain astronomy has entered a new era of large photometric data sets thanks to current and upcoming deep and wide-field surveys, such as the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; Kaiser et al. 2010), the Asteroid Terrestrial-impact Last Alert System (Jedicke et al. 2012), the Zwicky Transient Facility (Kulkarni 2018), the Large Synoptic Survey Telescope (LSST; Ivezic et al. 2008), and the *Wide Field Infrared Survey Telescope* (Spergel et al. 2015). The current surveys are already discovering $\sim 10^4$ supernovae (SNe) per year, a hundred-fold increase over the rate of discovery only a decade ago. LSST will increase this discovery rate to $\sim 10^6$ SNe per year.

SNe have traditionally been classified based on their spectra (Filippenko 1997). In the early days this was accomplished through visual inspection, then with template-matching techniques (e.g., SNID; Blondin & Tonry 2007), and most recently with deep learning techniques (e.g., Muthukrishna 2016). However, given the current discovery rate, and the anticipated LSST discovery rate, spectroscopic follow-up is severely limited. The consequences of this fact are twofold. First, we need a way to effectively identify "needles" in the haystack—the events that will yield the greatest scientific return with detailed follow-up observations (e.g., spectroscopy, radio, X-ray). Second, we need to devise methods to extract as much information and physical insight as possible from the "haystack" of SNe for which no spectroscopy or other data will be available. Here, we specifically focus on the latter issue and explore the question: given complete optical light curves (LCs), can we classify SNe into their main spectroscopic classes (Ia, Ibc, IIP, etc.)?

Previous studies in this area have largely focused on the simpler task of separating thermonuclear SNe Ia from non-SNe Ia, motivated by the use of SNe Ia as standardizable cosmological candles, and taking advantage of their uniformity (e.g., Möller et al. 2016; Kimura et al. 2017). Separating the classes of core-collapse SNe (CCSNe) is a broader and more challenging problem. First, unlike SNe Ia, CCSNe exhibit broad diversity between and within each class in terms of basic properties such as luminosity, timescale, and color (e.g., Drout et al. 2011; Taddia et al. 2013; Sanders et al. 2015; Nicholl et al. 2017; Villar et al. 2017). This is due to their wide variety of progenitor systems, energy sources, and circumstellar environments. Second, the overall diversity of CCSNe is less thoroughly explored, due to small sample sizes and few published uniform studies. As a consequence, most previous

works on photometric classification of CCSNe have relied on simulated data sets to train and test classification algorithms (e.g., Richards et al. 2011; Charnock & Moss 2017; Kimura et al. 2017; Ishida et al. 2019). Simulated data sets are based on strong assumptions about the underlying populations of each SN class and often do not reflect the true event diversity, or the effects of actual survey conditions.

Here, we approach the question of SN photometric classification using a large and uniform data set of 513 spectroscopically classified SNe from the Pan-STARRS1 Medium-Deep Survey (PS1-MDS). Importantly, the characteristics of this data set in terms of filters, depth, and cadence are the closest available analog to the LSST main survey design. We fit the observed LCs with a flexible analytical model that can accommodate all existing LC shapes, using a Markov chain Monte Carlo (MCMC) approach. We then train and evaluate 24 classification pipelines that span different feature extraction, data augmentation, and classification methods. We further use the posteriors of our MCMC fits to determine overall uncertainties on our classifications.

The paper is organized as follows. In Section 2 we introduce the PS1-MDS data set utilized here. In Section 3 we describe our analytical LC model and iterative MCMC fitting approach. In Section 4 we describe the key components of our various classification pipelines, including feature extraction, data augmentation, and classification approaches. We present the results of our classifications in Section 5, compare to previous classifications efforts in Section 6, and discuss limitations and future directions in Section 7.

Throughout this paper, we assume a flat ΛCDM cosmology with $\Omega_M = 0.286$, $\Omega_\Lambda = 0.712$ and $H_0 = 69.3$ km s^{-1} Mpc^{-1} (Hinshaw et al. 2013).

2. PS1-MDS Supernova LCs and Spectroscopic Classifications

PS1 is a wide-field survey telescope with a 1.8 m diameter primary mirror located on Haleakala, Hawaii (Kaiser et al. 2010). The PS1 1.4 gigapixel camera is an array of 60 4800 × 4800 pixel detectors with a pixel scale of 0.″258 and an overall field of view of 7.1 deg^2. The PS1 survey used five broadband filters, $g_{\rm P1}$, $r_{\rm P1}$, $i_{\rm P1}$, $z_{\rm P1}$, $y_{\rm P1}$. The details of the filters and the photometry system are given in Stubbs et al. (2010) and Tonry et al. (2012).

The PS1-MDS, conducted in 2010c2014, consisted of 10 single-pointing fields for a total area of about 70 deg^2 (Chambers et al. 2016). About 25% of the overall survey observing time was dedicated to the MDS fields, which were observed with a cadence of about 3 days per filter in $g_{\rm P1}r_{\rm P1}i_{\rm P1}z_{\rm P1}$ to a 5σ depth of ≈23.3 mag per visit. The typical sequence consisted of $g_{\rm P1}$ and $r_{\rm P1}$ on the same night, followed by $i_{\rm P1}$ and then $z_{\rm P1}$ on subsequent nights. Observations in the $y_{\rm P1}$-band were concentrated near full Moon with a shallower 5σ depth of ≈21.7 mag; we did not use the $y_{\rm P1}$-band data in this study due to its significantly shallower depth and poorer cadence.

The reduction, astrometry, and stacking of the nightly images were carried out by the PS1 Image Processing Pipeline (Magnier et al. 2016a, 2016b; Waters et al. 2016). The nightly stacks were then transferred to the Harvard FAS Research Computing Odyssey cluster for a transient search using the `photpipe` pipeline, previously used in the SuperMACHO and ESSENCE surveys (Rest et al. 2005; Miknaitis et al. 2007) and described in detail in our previous analyses of PS1-MDS data (Rest et al. 2014; Jones et al. 2018; Scolnic et al. 2018).

In the full PS1-MDS data set have identified 5235 likely SNe (Jones et al. 2017, 2018). During the course of the survey, spectroscopic observations were obtained for over 500 events using the MMT 6.5 m telescope, the Magellan 6.5 m telescopes, and the Gemini 8 m telescopes. We further obtained spectroscopic host galaxy redshifts for 3147 SN-like transients. The transients spectroscopically and photometrically classified as SNe Ia were published in Jones et al. (2017); the LCs and photometric classification of the remaining objects will be presented in future work. Similarly, the bulk of the SNe IIP (76 events) were published in Sanders et al. (2015), and the superluminous SNe I (SLSNe I) (17 events) were published in Lunnan et al. (2018). Here we focus on 513 spectroscopically classified events, which were classified using the SNID software package (Blondin & Tonry 2007). The sample contains 357 SNe Ia, 93 IIP/L SNe, 25 SNe IIn, 21 SNe Ibc, and 17 SLSNe I.[13]

Our sample is limited to events with high-confidence spectroscopic classifications with a statistically useful number of members in each class. As part of the PS1-MDS we discovered several other rare transients, including tidal disruption events (Gezari et al. 2012; Chornock et al. 2014) and fast-evolving luminous transients (Drout et al. 2014), but the sample sizes for those are too small for inclusion in this study. It is possible that SNID misclassification exists in our data set; e.g., low signal-to-noise (S/N) events are more likely to match to SNe Ia (Blondin & Tonry 2007). To partially counteract this, we check each member of our Type Ibc and SLSNe classes (our smallest classes) by eye to ensure high purity. Finally, we note that the magnitude limit for our spectroscopic follow-up was generally shallower by about 1.5 mag relative to the PS1-MDS nominal per-visit depth. This does not affect our ability to test classifiers on the spectroscopic sample itself, but will be considered when extending our method to the full photometric data set in future work (see Section 6).

The LCs range from a minimum of 3 to ≈150 total data points in any filter with a signal-to-noise ratio of S/N > 3, with a median of about 30 data points in each LC. We have only eliminated events with LCs that contain fewer than two 3σ detections in three or more filters, eliminating seven SNe from our sample[14] (six SNe Ia and one SN II) leaving 506 remaining SNe for our training set.

In Figure 1 we plot the peak absolute $i_{\rm P1}$ magnitude versus redshift for our spectroscopic sample. The sample spans $M_i \approx -14.5$ to -22.5 and extends in redshift to $z \approx 1.6$, with only the brightest classes (SLSNe and Type IIn) being observed at $z \gtrsim 0.6$. Specifically, we find a range of $M_i \approx -14.5$ to -18.5 mag for the SNe II, ≈-16.5 to -19.5 mag for the SNe Ibc, -16 to -20.5 for the SNe IIn, and ≈-20.5 to -22.5 for the SLSNe.

[13] Three of the 17 SLSNe (PS1-12cil, PS1-10ahf, and PS1-13or) do not have spectroscopic host redshift measurements. Lunnan et al. (2018) estimated their redshifts (0.32, 1.10 and 1.52, respectively) from strong rest-frame UV features for the $z > 1$ objects and SN Ic-like post-peak features for PS1-12cil.

[14] For completeness, we ran our final classifier on these LCs as well and found that five of the six SNe Ia, as well as the one SN II, were actually correctly classified, albeit with a low classification confidence.

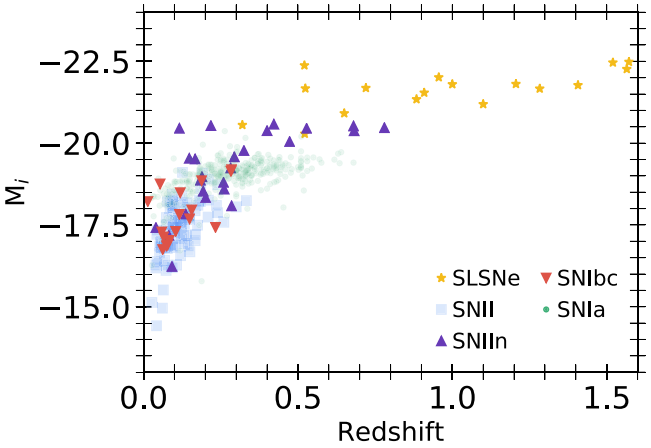

Figure 1. Peak $i_{\rm P1}$-band absolute magnitude vs. redshift for the sample of PS1-MDS spectroscopically classified SNe used in this study. We apply a cosmological k-correction to the peak magnitudes, but do not correct for the intrinsic spectral energy distribution of the various SNe. The sample includes five SN classes: Ia (green circle), Ibc (red downward triangle), II (blue square), IIn (purple upward triangle), and SLSNe (yellow star).

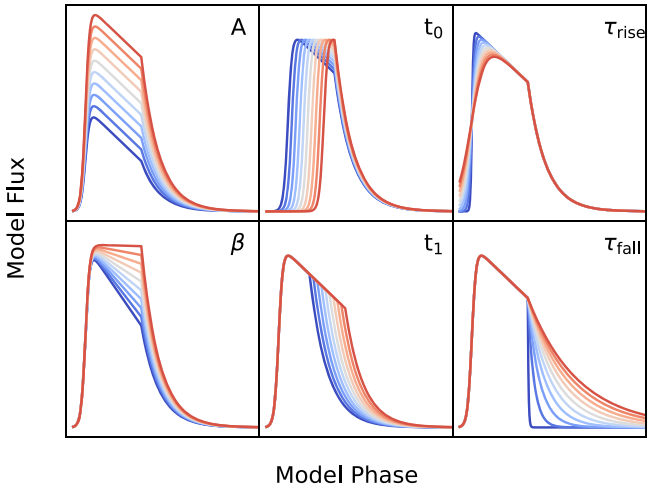

Figure 2. Example model light curves (LCs) based on Equation (1) highlighting how each of the free parameters affects them. The parameters are individually varied from low (blue) to high (red) values.

3. Analytical LC Model and Fitting

Rather than interpolating data points, a common method to standardize data is to fit a simple parametric model to the LCs (e.g., Bazin et al. 2009; Newling et al. 2011; Karpenka et al. 2013). However, the majority of existing analytical LC models are best-suited for SNe Ia and have limited flexibility for the full observed range of SN LC shapes. Here we present and fit our data with a new parametric piecewise model that is designed to be flexible enough for a broad range of LC morphologies:

$$F = \begin{cases} \dfrac{A + \beta(t - t_0)}{1 + e^{-(t-t_0)/\tau_{\rm rise}}} & t < t_1 \\ \dfrac{(A + \beta(t_1 - t_0))e^{-(t-t_1)/\tau_{\rm fall}}}{1 + e^{-(t-t_0)/\tau_{\rm rise}}} & t \geqslant t_1. \end{cases} \quad (1)$$

The model contains seven free parameters, whose effects on the resulting LCs are shown in Figure 2. Although each parameter has a unique and interpretable effect, some degeneracies between the parameters exist. For example, the parameter A affects the amplitude of the LC, although its value does not exactly correspond to the peak flux. Similarly, t_0 acts as a temporal shift in the LC, but does not directly correspond to the time of explosion or the time of peak. The parameters, $t_{\rm rise}$, t_1, and $t_{\rm fall}$ control the rise, plateau onset, and fall time of the LC, respectively. For the purposes of fitting, we reparameterize t_1 into a new parameter $\gamma \equiv t_1 - t_0$, which better represents the plateau duration of the LC and results in fewer degeneracies when fitting. Finally, the parameter β controls the slope of the plateau phase.

This functional form is similar to those presented in Bazin et al. (2009) (with five free parameters) and Karpenka et al. (2013) (with six free parameters), but incorporates a plateau component. In Figure 3 we show examples of fits to an SN IIP and an SN Ia with our model, the Bazin model, and the Karpenka model. Our model provides a better fit to both the fast rise time and plateau phase of the SN IIP LC, and is flexible enough to also fit the smoother LC of a SN Ia. We note that Sanders et al. (2015) presented a similar piecewise model with 11 free parameters to fit a sample of 76 PS1-MDS SNe II; however, Lochner et al. (2016) found that this model was not robust when fitting data without the use of informative priors, due to the large number of free parameters. Additionally, the sharp transitions between rise and decline in the Sanders model make it difficult to fit CCSNe with smooth peaks.

One common SN Ia LC feature missing from our model is the second peak in the red LCs at about one month post-explosion (e.g., Kasen 2006; Mandel et al. 2011; Dhawan et al. 2015). We find that this feature manifests itself as a "plateau" in our analytical model in the i- and z-bands. However, as we show in Section 5, our classification pipelines can reliably classify SNe Ia without explicitly including a second peak in our model.

We fit the LCs using PyMC (V2; Patil et al. 2010), a Python module that implements a Metropolis–Hastings MCMC sampling algorithm. We assume uniform priors on all parameters with the exception of γ. We found that LCs typically fall in one of two solutions: LCs with a long plateau (in the case of SNe IIP) and LCs that lack a plateau (all other types). To best reflect this fact, we set the prior of γ to a double Gaussian peaked at 5 and 60 days. This prior helps to remove a degeneracy in which a steep exponential decline can resemble a linear decline. The priors are listed in Table 1. We use a standard likelihood function, incorporating both the observational error and a scalar white noise scatter term added in quadrature. We find that several of our model parameters are correlated (degenerate) with one another. In particular, the amplitude (A) is negatively correlated with both the rise time ($t_{\rm rise}$) and plateau duration (γ) but negatively correlated with the start time (t_0). Additionally, duration is negatively correlated to both the rise time and start time, while the rise time is positively correlated with the start time.

We fit the LC in each of the four filters independently, in the observer frame, but use an iterative fitting routine to incorporate combined information from all filters. We first run the MCMC to convergence on each filter independently with the same set of priors. We then combine the marginalized posteriors (i.e., we ignore parameter covariances) from each filter and use the combined posterior as a new prior for a second iteration of fitting. We can apply this process repeatedly, but we find that a single iteration is sufficient for the vast majority of events. Our iterative procedure is essential for fitting LCs in which some filters have significantly fewer data points, a situation that is common in photometric surveys

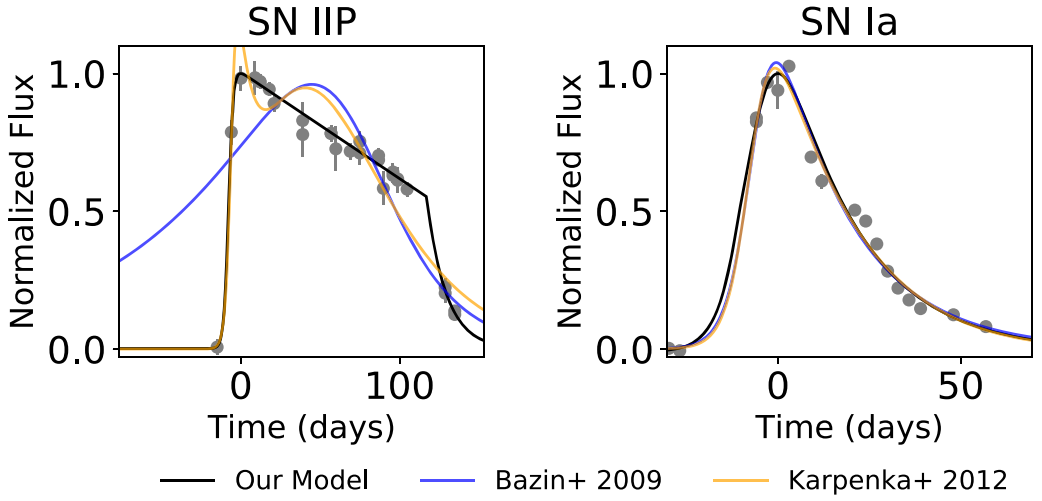

Figure 3. Comparison of our analytical LC model (Equation (1); black line) to that of Bazin et al. (2009) (blue line) and Karpenka et al. (2013) (yellow line) for *i*-band lightcurves of both an SN IIP (left) and an SN Ia (right). Our model performs similarly for an SN Ia, but is superior at fitting SNe with a LC plateau.

Table 1
Parameter Descriptions and Priors

Parameter	Description	Prior
$\tau_{\rm rise}$ (days)	Rise time	U(0.01, 50)
$\tau_{\rm fall}$ (days)	Decline time	U(1, 300)
t_0 (MJD)	"Start" time	U($t_{\rm min} - 50$, $t_{\rm max} + 300$
A	Amplitude	U(3σ, 100 $F_{\rm max}$)
β (flux/day)	Plateau slope	U($-F_{\rm max}/150$,0)
c (flux)	Baseline flux	U(-3σ, 3σ)
γ (days)	Plateau duration	(2/3)N(5,5) + (1/3)N(60, 30)

due to differences in relative sensitivity, the intrinsic colors and color evolution of SNe, and varying observing conditions. An example of the best-fit solutions given by the first and second iterations is shown in Figure 4. In this example, the peak times in *g*- and *i*-filters are in disagreement with *r*- and *z*-filters due to poorly sampled data in the former two. Following the second iteration, this disagreement is removed, leading to more realistic fits.

Representative LCs and their best fits are shown in Figure 5. The solutions are constrained for well-sampled LCs (e.g., the SNe Ia shown) but more poorly constrained for sparse LCs (e.g., the SLSNe shown). Crucially, because we have access to the full posterior of LC solutions, we can feed many samples of the posterior through our classification algorithm to quantify the classification uncertainty for each event.

Unless otherwise specified, we use the observer-frame LC fits to extract features. We then include the redshift to transform to absolute magnitudes, including a cosmological *k*-correction: $M = m - 5\log(d_L/10{\rm pc}) + 2.5\log(1+z)$, where d_L is the luminosity distance. We do not apply *k*-corrections to account for the intrinsic spectral energy distribution of the various SN types.

4. Classification Pipelines

For each SN, our MCMC fitting generates posterior distributions for the model LC parameters. To train a classifier, we need to extract features from the LCs generated by the fitted parameters. We test several methods of feature extraction, data augmentation and classification. We describe each method in the following subsections, and we compare the algorithms in

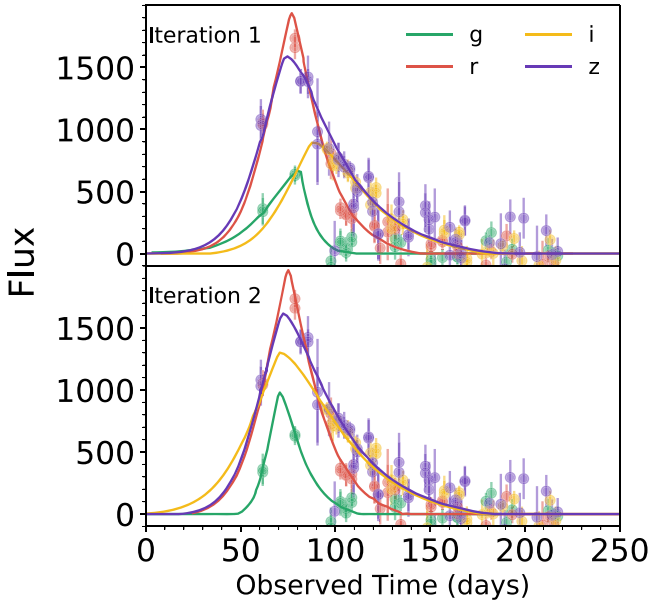

Figure 4. Example best-fit LCs in the four PS1 filters after the first (top) and second (bottom) MCMC iterations. Following the first iteration, the peak time varies significantly between the filters due to differences in the data quality and time sampling. The best-fit solution of the second iteration, using the combined posteriors from the first iteration, provides much better agreement in the LC properties.

terms of classification purity, completeness, and accuracy in Section 5. *Purity* (also called *precision*) is defined as the fraction of events in a predicted class that are correctly identified; for example, if our classifier predicts a total of 100 SNe Ia, but only 70 of those are spectroscopically classified as SNe Ia, the purity would be 0.7. *Completeness* (also called *recall*) is defined as the fraction of events in an observed class that are correctly identified; for example, if our sample contains 100 spectrosopically classified SNe Ia, but our classifier has only identified 70 of those events as SNe Ia, then our completeness would be 0.7. *Accuracy* is defined as the total fraction of events that are classified correctly as being a member or not of a given class; for example, if a sample of 100 SNe contains 70 spectrosopically classified SNe Ia, and our classifier correctly identifies the 70 SNe Ia but incorrectly

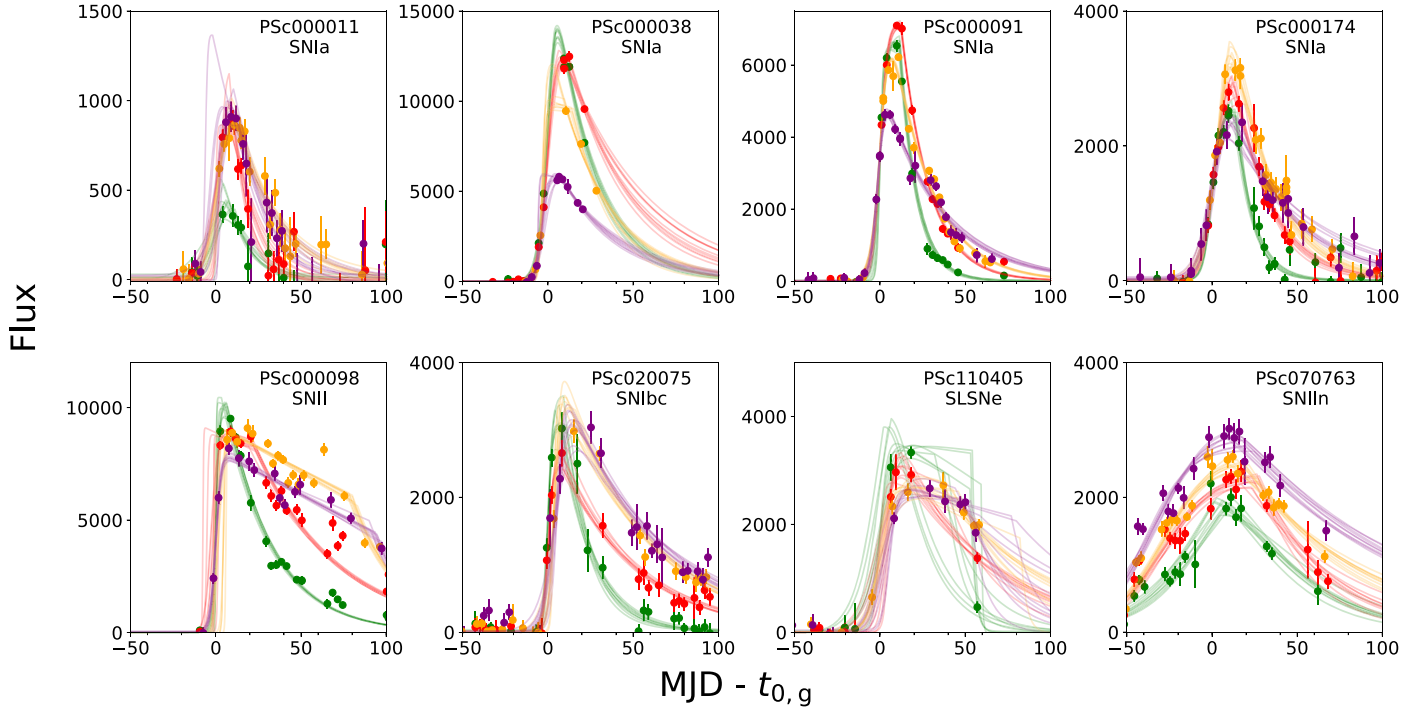

Figure 5. Example light curves and sample posterior draws of associated model fits in the four filters for various SN types. The model is described by Equation (1) and the fitting procedure is described in Section 3.

classifies 20 more CCSNe as SNe Ia, the overall accuracy is 0.8. The three terms are mathematically defined as follows:

$$\text{Purity} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Completeness} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TS}}, \tag{4}$$

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.

4.1. Feature Selection

Although our analytical model produces interpretable features for each LC (albeit ones that are somewhat degenerate) we would like to explore various methods of feature extraction, based on the analytical fits. In particular, we explore the following four types of features.

1. Model parameters (M). We use the analytical model parameters as features, as well as the peak absolute magnitude in each filter, including a cosmological k-correction but no correction for intrinsic SN colors and color evolution.
2. Hand-selected features (HS). We use HS interpretable features: the peak absolute magnitude in each filter, including a cosmological k-correction but no correction for intrinsic SN colors and color evolution, and the rest-frame rise and fall times by 1, 2, and 3 mag relative to peak (where we do *not* correct the rise and fall times for cosmological time-dilation).
3. Principal component analysis (PCA). We fit a PCA decomposition model to the full set of analytical model fits (without any redshift corrections) independently for each filter. We use the first six PCA components from each filter, corresponding to an explained variance within the LCs of ∼99.9%. We also use the peak absolute magnitude, including a cosmological k-correction, in each filter in addition to the PCA components.
4. LCs. We use the model LCs as the features. We renormalize the flux of each LC, correcting for luminosity distance; however, we find that neglecting time dilation corrections improves classification accuracy, and therefore we do not make these corrections. We down-sample each filter model to 10 observations logarithmic-spacing between t_0 and $t_0 + 300$ to decrease the number of features.

To provide some intuition, we highlight a sub-space of the HS features (M_{peak} versus duration time to rise and fall by 2 mag) in Figure 6. We find that some SN classes, such as SLSNe versus Type II, or Type Ia versus Type IIn, easily separate in the duration–luminosity feature space. However, other classes, such as Type Ibc versus Type Ia and IIP, have substantial overlap in this space, regardless of filter. This highlights that, while simple heuristics can be used as first-order classifiers for some SN classes, other classes are intrinsically difficult to disentangle from LC information alone.

4.2. Data Augmentation

Data augmentation is ubiquitous in machine-learning applications, as a larger data set can significantly improve the accuracy and generalizability of most classification algorithms. Data augmentation methods have already been utilized in the astrophysical context (e.g., Hoyle et al. 2015).

Here, we augment our training set with simulated events for two key reasons. First, our training set is unbalanced in terms of SN classes due to the differing observed rates of transients,

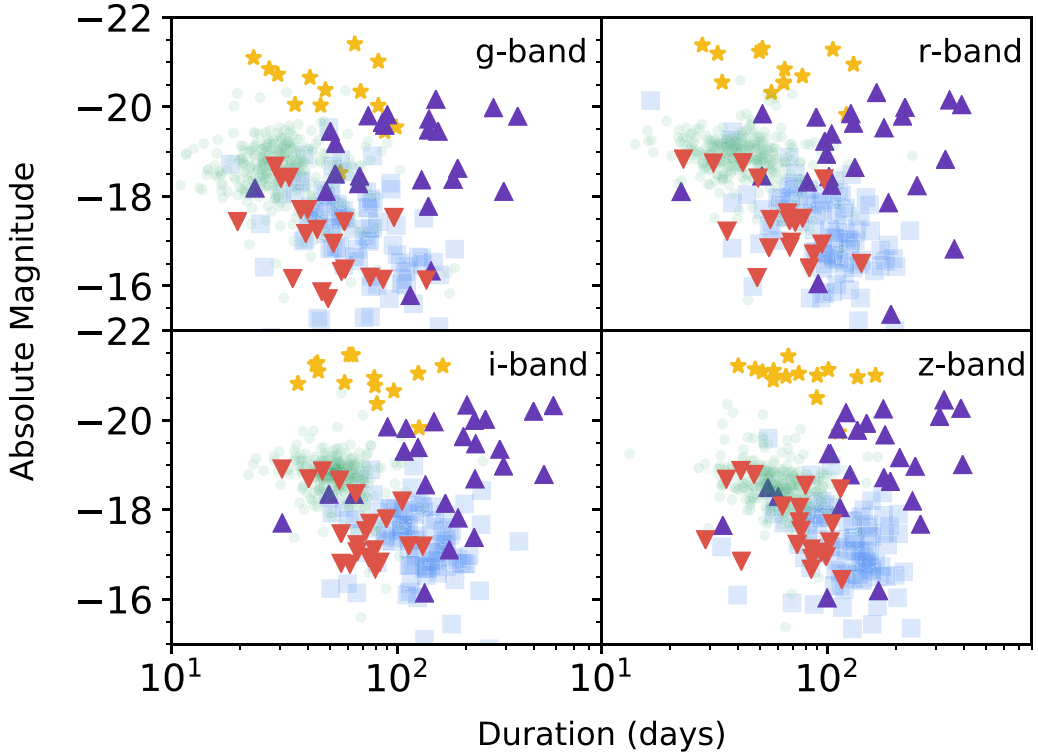

Figure 6. Duration–luminosity feature space of the data set in the four PS1 filters. Duration is defined as the total time for the light curves to rise and decline by 2 mag relative to the peak. The plotted values are from the median model fits to the light curves using Equation (1). The sample includes the five SN classes: Ia (green circle), Ibc (red downward triangle), II (blue square), IIn (purple upward triangle), and SLSNe (yellow star).

with SNe Ia representing ≈70% of our sample (and more generally, of any magnitude-limited optical survey). Classification algorithms trained on unbalanced training sets tend to over-classify all objects as the dominant class. This is because the algorithms can minimize the decision-making complexity by ignoring minority classes in favor of correctly classifying the majority class. In our case, a classification algorithm may preferentially label all objects as SNe Ia to achieve an overall high accuracy. Second, our training set is small in the context of machine learning, with the smallest class (SLSNe) containing just 17 events.

One approach to overcome this in the context of our method is to augment our training set with many draws from the MCMC posteriors. However, this would lead to clustering of solutions in feature space that may bias the training algorithms. Instead, we address the issue of a small and imbalanced training set by synthesizing more event samples using two techniques. First, we use the synthetic minority over-sampling technique (SMOTE; Chawla et al. 2002) to over-sample all the non-SN Ia classes to be equally represented as the SNe Ia. SMOTE creates synthetic samples in feature space by randomly sampling along line segments joining the k nearest neighbors of a sample, where k is a free parameter of the algorithm. Here we find that $k = 5$ performs well for sampling the minority classes. An example of the SMOTE resampling algorithm is shown in Figure 7. A key feature of SMOTE resampling is that it produces realistic samples within each class, but it cannot produce samples outside the extent of the original sample. While this prevents the generation of unphysical models, it may overly constrain the properties of classes with only a few samples (e.g., SLSNe).

Second, we augment the non-SN Ia classes by fitting the feature space of each class to a multivariate-Gaussian (MVG) and resampling from the fitted MVG. This is similar to the SMOTE algorithm in that it allows for the generation of new events that encompass a larger potential feature space. However, one key difference is that this method allows for synthesized events beyond the feature boundaries seen in the data. While this may lead to some unphysical models, it better reflects the potential spread in LC parameters in poorly sampled classes. An example of MVG resampling is shown in Figure 7. Both augmentation methods aim to increase our training set in a way which is representative of the set and therefore makes no attempt to correct for potential biases. This can potentially lead to increased misclassifications if our labeled training set is unrepresentative of a future test set; however, we expect no such effects within the training set of 513 objects.

4.3. Classification

Following the work of Lochner et al. (2016), we test three classification algorithms: a support vector machine (SVM), a random forest (RF), and a multilayer perceptron (MLP). We optimize the hyperparameters of each algorithm independently using a grid search. Each algorithm and its tunable hyperparameters are described below. We use the `scikit-learn` python package throughout the classification portion of our pipeline.

4.3.1. Support Vector Machine

An SVM classifies the training set by finding the optimal hyperplane in feature space to minimize the number of

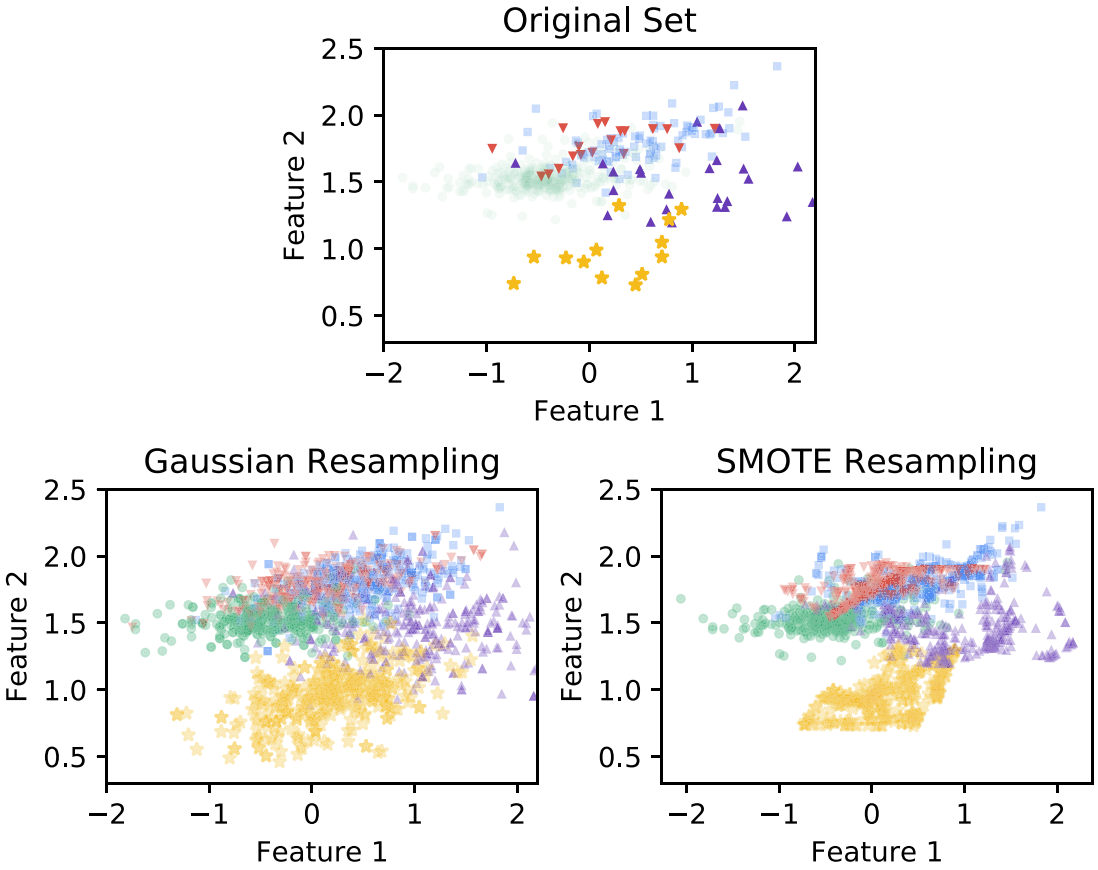

Figure 7. Top: original data set plotted in terms of feature 1 vs. feature 2, indicating both the span of the various SN classes and the imbalance in number of events per class. Bottom: augmented data set using multivariate-Gaussian resampling (left) and SMOTE resampling (right).

misclassified samples. In particular, the SVM will select a hyperplane that maximizes the distance between class samples nearest the hyperplane (also known as the support vectors). In the majority of cases, the classes are not linearly separable within the feature space alone (i.e., there may be significant overlap between classes). Instead, the features are expanded into an infinite basis function using the so-called kernel trick (Aizerman 1964), allowing one to find a feature space in which the separating hyperplane is linear. We optimize the kernel and a regularization term using a coarse grid search, allowing the kernel size to logarithmically range from $\sigma = 1$ to $\sigma = 100$ and the normalization to logarithmically range from 1 to 1000. We find that a radial basis function kernel with width $\sigma = 10$ typically results in optimal classification, with normalization values ranging depending on the pipeline.

4.3.2. Random Forest

RF classifiers (Breiman 2001) are built on the idea of a decision tree, which is a model that generates a set of rules to map input features to classes. This mapping is based on a series of branching decisions based on feature values (e.g., "is the peak g-band magnitude brighter than -19?"). While single trees are theoretically sufficient for classification problems, they often lead to over-fitting due to the specialized branching required for each class. RFs overcome this problem by combining decision trees that are trained on different subsets of the training data and features. The ensemble of decision trees is then used as the classifier. There are a number of free parameters within an RF, including the number of decision trees, the number of nodes for each tree, and the splitting rules for each node. Through a grid search of hyperparameters, we find that 100 decision trees utilizing the Gini impurity (the probability that a randomly chosen SN from a labeled class is misclassified) as a splitting criterion and allowing nodes to be split until all leaves are pure results in the highest accuracy.

4.3.3. Multilayer Perceptron

A fully connected MLP is the simplest artificial neural network (e.g., Schmidhuber 2015). It is composed of a series of layers of neurons, where each neuron is the dot product of the previous layer and a set of optimizable weights, passed through a nonlinear activation function. A "fully connected" MLP means that each neuron is connected to all neurons in the preceding layer. The nonlinear activation function is what allows a MLP to model nonlinear mappings between the feature set and classes. MLPs have many tunable parameters, including the number of layers, the number of neurons within each layer, the learning rate, and a regularization term. We optimize the hyperparameters using a grid search, finding that two layers with 10 neurons each typically perform best, and use the `Adam` optimization algorithm (Kingma & Ba 2014) to train the MLP.

An example of a complete pipeline, excluding the MCMC fitting step, is available on GitHub.[15]

[15] https://github.com/villrv/ps1ml

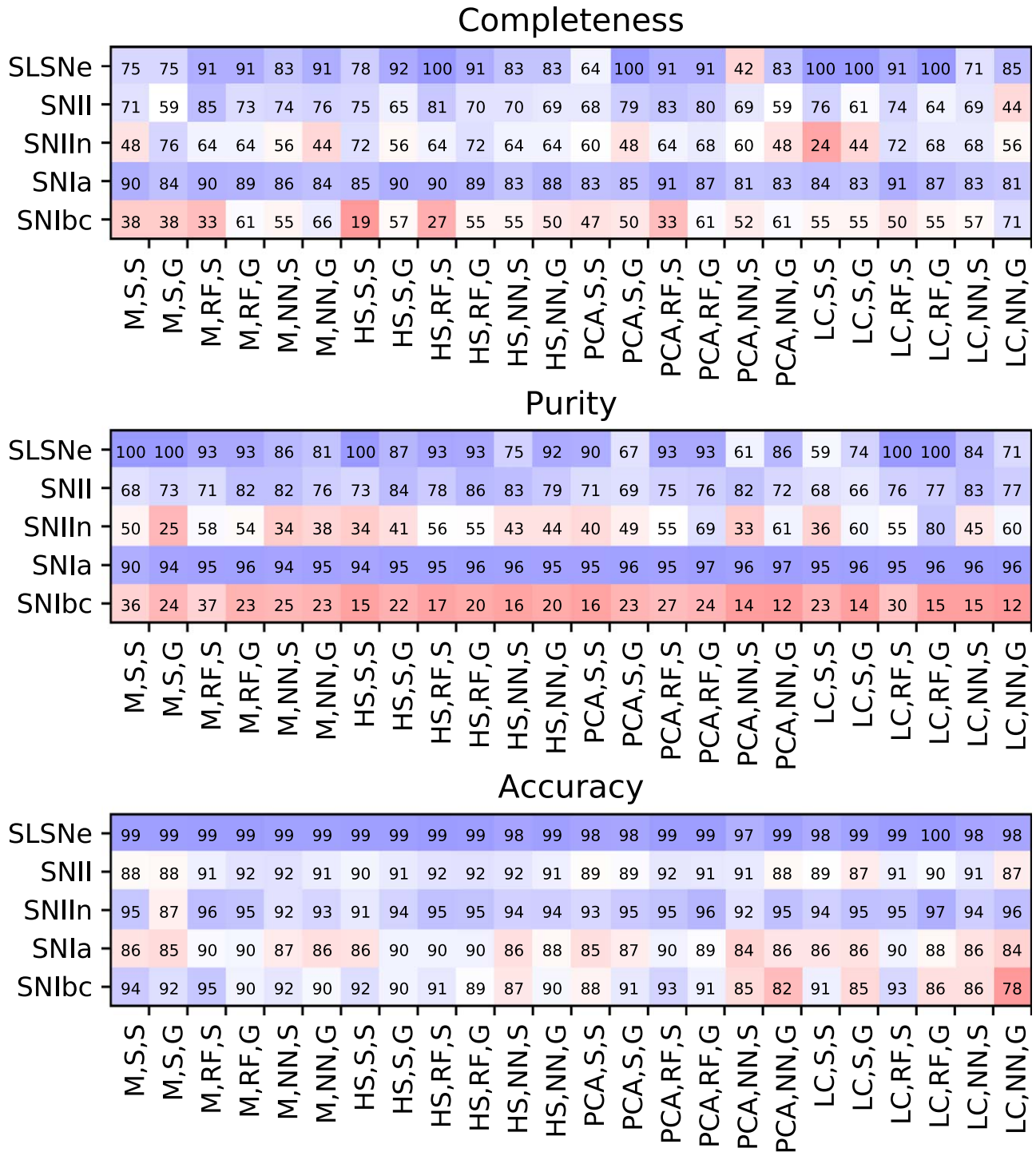

Figure 8. Completeness (top), purity (middle) and accuracy (bottom) for each of the five spectroscopic SN classes across the 24 classification pipelines. Each pipeline is encoded by its feature extraction method (M, HS, PCA, LC), data augmentation method (SMOTE (here S), MVG (here G)) and classification method (SVM, RF, MLP (here NN)).

5. Classification Results

We combine each of the four feature extraction methods (M, HS, PCA, and LC), two data augmentation methods (SMOTE and MVG), and three classification algorithms (SVM, RF, and MLP) to test a total of 24 classification pipelines. For each pipeline, we use the full data set to find the hyperparameters that optimize overall accuracy for the classification method. We optimize the hyperparameters over a coarse grid, due to the computational costs of performing a large grid search. We then perform leave-one-out cross-validation by iteratively removing one object from the sample, performing data augmentation on the remaining data set, and training a classifier on the new set. We then test the trained classifier on the median posterior values of the removed object and record the predicted label. Due to computational costs, we only utilize the full posteriors for classification error estimation using our optimal pipeline.

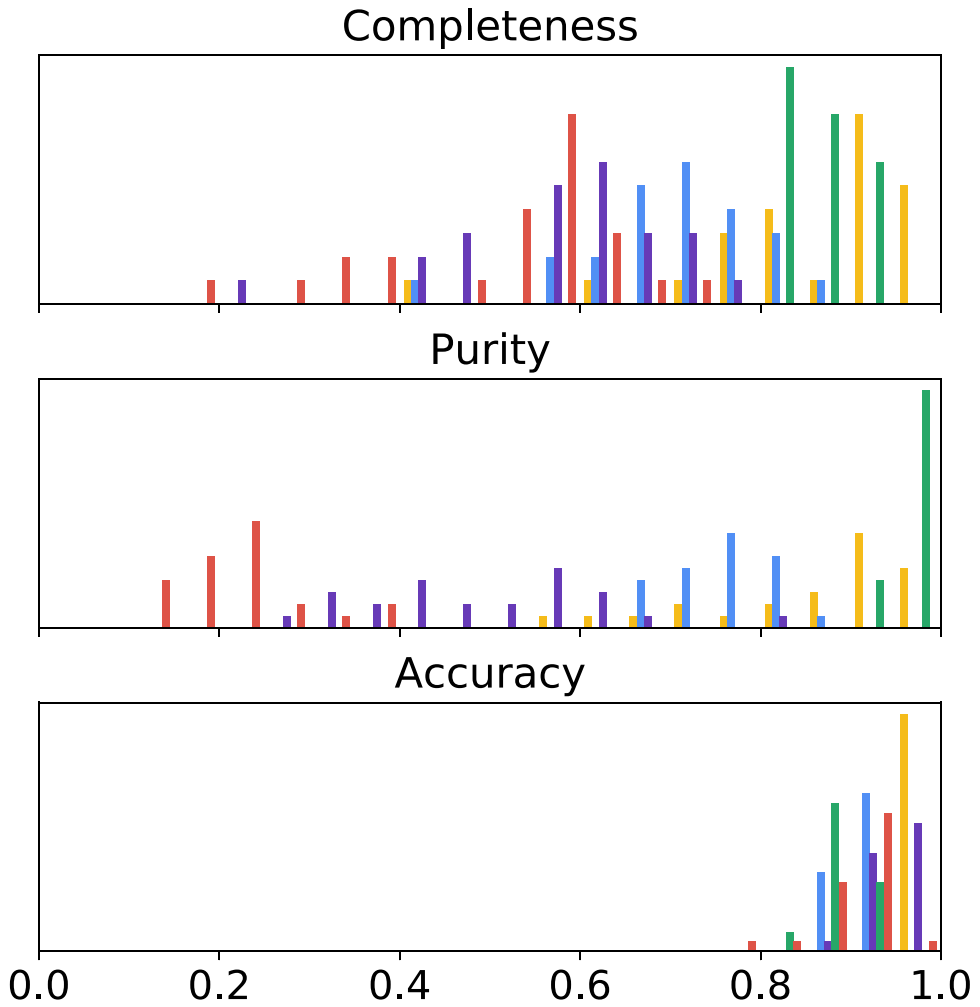

Figure 9. Histograms of completeness (top), purity (middle) and accuracy (bottom) for each of the five spectroscopic SN classes across the 24 classification pipelines.

5.1. General Trends

In Figures 8 and 9 we plot the purity, completeness, and accuracy for each of the 24 pipelines and each of the five SN classes. Figure 8 provides a matrix representation with the percentage score noted for each combination of pipeline and SN class, while Figure 9 shows the same results in histogram format to aid in visualizing the range of completeness, purity, and accuracy values across the 24 pipelines for each SN class.

We find that SLSNe and SNe Ia are consistently the classes with the highest purity and completeness, reaching $\gtrsim$90% for the best classification pipelines. This is due to the fact that SLSNe are easily separable from the other classes due to their high luminosity and longer durations (Figures 1 and 6), while SNe Ia are tightly clustered in feature space due to their intrinstic uniformity.

In contrast, we find that SNe Ibc typically have the lowest purity and completeness, with ≈15%–35% and ≈25%–65%, respectively, and a much wider spread in performance for the various pipelines. The lower classification success rate is due to broader diversity within SNe Ibc, as well as their significant overlap with SNe Ia (e.g., Figure 6).

For SNe II we find high values of purity and completeness of ≈65%–85% and ≈60%–80%, respectively. This overall high success rate is mainly due to the presence of a plateau phase that helps to distinguish most SNe II from the other classes. However, the failed classifications are most likely due to the faster-evolving SNe II (often called Type IIL), which tend to be misclassified as Type Ibc or SNe Ia due to overlap in LC shapes (e.g., Figure 6).

Finally, for SNe IIn we find purity and completeness of ≈30%–80% and ≈45%–70%, respectively, reflecting the broad diversity of LC morphologies and luminosities, with some events overlapping similar areas in feature space with SNe Ia and Ibc (e.g., Figure 6). As for the SNe Ibc, we find quite a broad dispersion in performance between the various pipelines.

For the overall accuracy across the five SN classes, we find generally high values of ≈100% for SLSNe, ≈95% for SNe IIn, ≈90% for SNe II, ≈85%–95% for SNe Ibc, and ≈85%–90% for SNe Ia. These values are essentially independent of the classification pipeline used.

To further explore the relative performance of the various pipelines, in Figure 10 we plot the distribution of completeness across the full data set, grouping the classification pipelines by feature extraction method, classification method, and data augmentation method. We find that the classification method has the largest impact on completeness, with the RF classifiers performing noticeably better, and more uniformly, than the SVM and MLP (NN) classifiers. In terms of feature extraction we find that use of model parameters (M) and PCA are somewhat advantageous compared to HS features and the LC approach, although the PCA extraction leads to a broader range

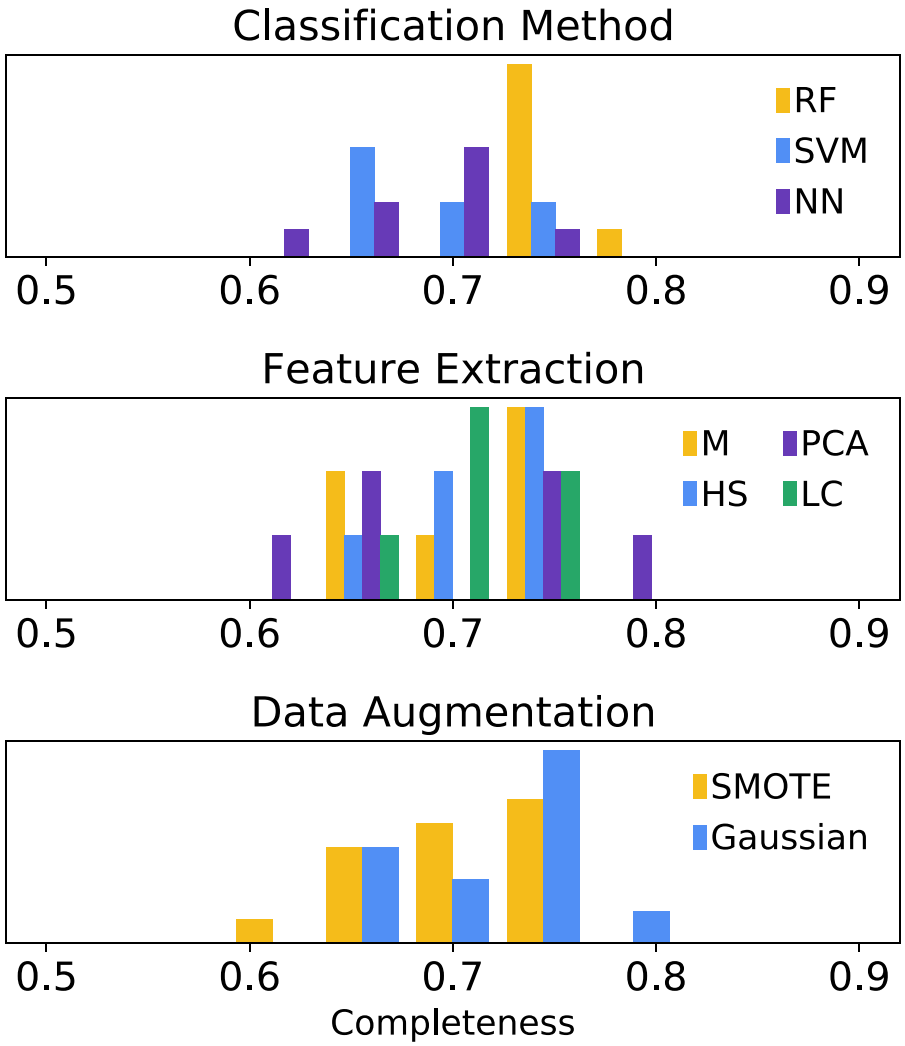

Figure 10. Histograms of completeness across all five SN classes, grouped by classification method (top), feature extraction method (middle) and data augmentation method (bottom).

of outcomes. Finally, the MVG augmentation method performs slightly better than SMOTE.

The top three pipelines in terms of purity, completeness and accuracy share RF classification and PCA feature extraction, with both MVG and SMOTE augmentation. Between these pipelines, the overall accuracy differs by $\lesssim$5% across the five SN classes. In addition to performing well, the RF classifier also has the advantage of allowing us to measure the relative important of each feature. For example, we test the relative importance of our HS and M features in the RF classification pipeline using the "gini importance," a measure of the average gini impurity decrease across descending nodes (Leo et al. 1984). We find that the peak magnitudes are the most important interpretable features, with durations and other parameters being roughly equally important.

For simplicity, below we focus on the results of our pipeline with the highest purity (72%) and completeness (78%) scores with an average accuracy of 93% across the five SN classes. This pipeline consists of PCA feature extraction, MVG data augmentation, and RF classifier; however, we emphasize that this pipeline does not significantly outperform the others. In Figure 11 we present the final confusion matrix for this pipeline across the full training set. The confusion matrix is a quick-look visualization of how each class is correctly or incorrectly classified. We generate the confusion matrix using the full posteriors for each SN, so the probability densities have been effectively smoothed out across the matrix. To specifically assess the role of poor-quality classifications, we show the confusion matrix for the full sample, as well as separately for classifications with a confidence of $p > 0.8$ only (representing $\sim$85% of the original sample). In practice, one can optimize pipeline parameters to maximize sample purity, completeness, or some other metric.

5.2. Assessing Misclassifications

Although the overall completeness for each SN class is high, we note several common misclassifications. First, Type II and Ia are the most likely classes to be misclassified as SNe Ibc. The SNe II that are misclassified as SNe Ibc are typically either poorly sampled or are rapidly evolving (the so-called IIL events). Second, Type Ibc, IIn, and II SNe are the most likely classes to be misclassified as SNe Ia. This is again due to specific events in those diverse classes that occupy the region in feature space that overlaps with the uniform SNe Ia. Finally, Type IIn and Ibc SNe are the most likely classes to be misclassified as SNe II, again due to overlaps in feature space. Comparing the full sample to the subset of events with high

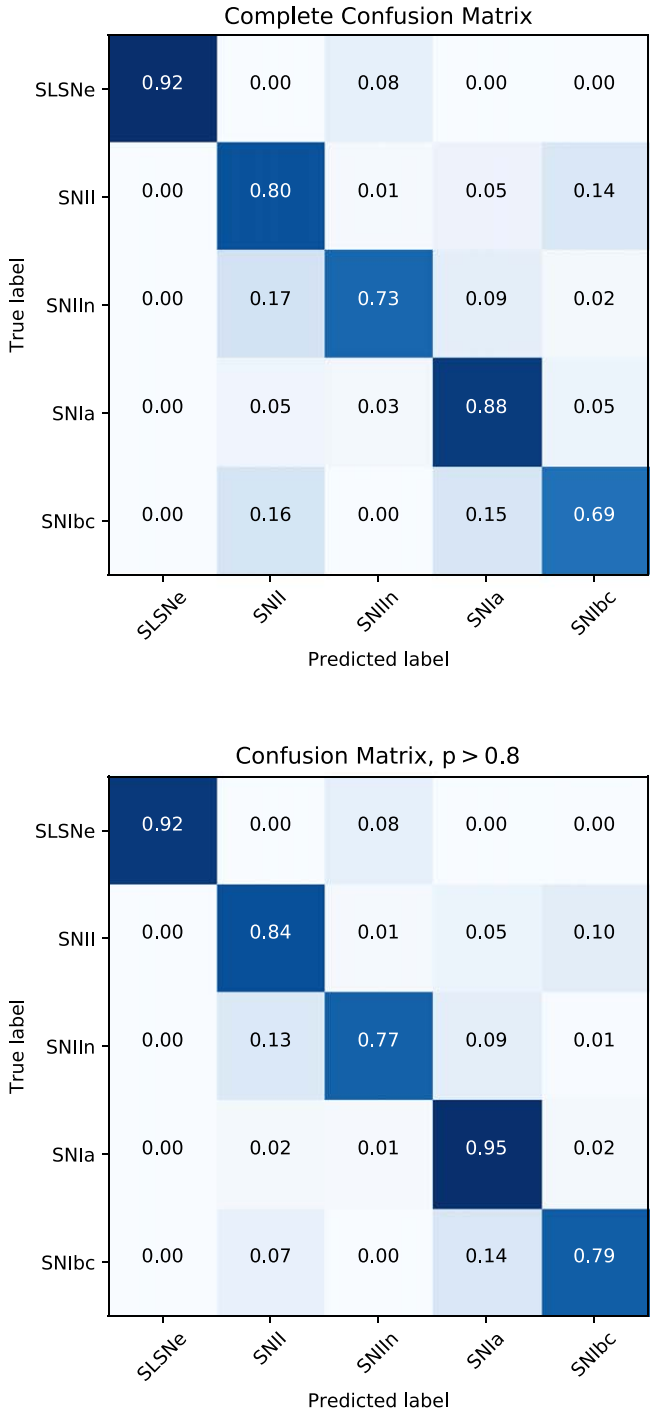

Figure 11. Confusion matrix for one of our best performing classification pipelines (PCA feature extraction, MVG data augmentation, and RF classifier) calculated using the full posterior distributions for each SN. We show the confusion matrix for both the full SN sample of 513 objects (top) and only for the 429 events with a high classification confidence probability of $p > 0.8$ (bottom).

classification confidence ($p > 0.8$) we find that the fraction of misclassified events indeed declines (most notably for SNe Ibc and Ia), indicating that some misclassifications are simply due to poorly sampled LCs. However, the overall trends for which classes are most likely to be misclassified as others remains the same, indicating that there is an inherent limitation to the classification success rate that is due to real overlaps in feature space.

We highlight several SNe that are misclassified, but with high confidence, in Figure 12. In these examples, a spectroscopic SN II with a rapid linear decline is misclassified as an SN Ibc, a slightly dim SN Ia is misclassified as an SN Ibc, and a fairly luminous SN Ibc is misclassified as an SN Ia. In each of these cases, the posterior of the fitted LCs is narrow, leading to little variability (i.e., high confidence) in the final classification. These events indicate that even with good photometric data quality there is inherent overlap of SNe in feature space that leads to misclassification.

The misclassifications of SNe are further highlighted in Figure 13. Each panel in the top part of Figure 13 represents a spectroscopically classified class, while in the bottom part each panel represents a photometrically assigned class. The misclassified events in both cases are labeled to provide insight into the most common misclassification. In all panels the ordinate represents the overall classification certainty, based on many draws from the posteriors of each event. In all cases, the majority of misclassifications occur at the low-confidence end ($p < 0.8$), but there are also high-confidence misclassifications.

We explore the role of data quantity in Figure 14, where we plot the classification accuracy as a function of total LC data points for all five SN classes. We again find that misclassifications are more likely in the regime of low number of data points, specifically $\lesssim$20 data points. However, as noted above, there are also high-confidence misclassifications for events with a large number of data points.

6. Comparison to Previous Photometric Classification Approaches

The photometric classification of optical transients has been previously explored in the existing literature. Previous studies on machine-learning methods have focused almost exclusively on the binary problem of Type Ia versus non-SN Ia classifications (e.g., Campbell et al. 2013; Ishida & de Souza 2013; Jones et al. 2017), or have been trained and tested on simulated data sets (e.g., Kessler et al. 2010; Tonry et al. 2012; Möller et al. 2016; Muthukrishna 2016; Charnock & Moss 2017; Möller & de Boissière 2019). We highlight the strengths and weaknesses of both approaches (which we note are disjoint) compared to our methodology. We emphasize that classification pipelines should ideally be compared using the same data set and set of labels. No machine-learning method, including the one presented in this paper, can be applied to a new test set without retraining or careful consideration of training-versus-test set biases. This is especially crucial when comparing our method to those created for simulated data sets, which have known biases, uncertainties, and simulated physics.

Identification of SNe Ia from photometric LCs is essential for precision cosmology in the era of large photometric surveys (Scolnic et al. 2014; Jones et al. 2017), which is why many studies have specifically focused on SN Ia classification. However, the binary problem of Type Ia versus non-SN Ia classification is much narrower (and simpler) than full classification of CCSN classes. As standardizable candles, SNe Ia are fairly homogeneous with observational variations (excluding reddening) that are well described by two observable features: stretch and peak luminosity. As a result, it is easier to separate the small area of feature space corresponding to SNe Ia from other transients. Studies that focus on this approach achieve a classification accuracy of $\gtrsim$0.95 (e.g., Ishida & de Souza 2013; Charnock & Moss 2017;

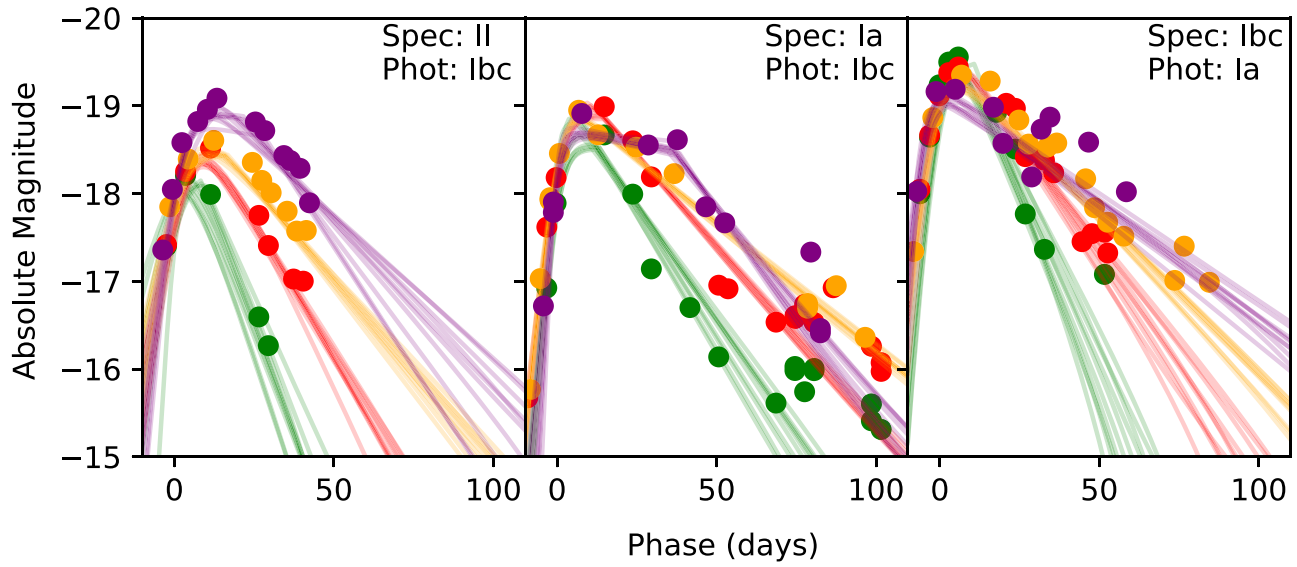

Figure 12. Light curves of three SNe classified incorrectly, but with high confidence ($p > 0.9$). We note the spectroscopic and photometric classification of each event. Given the high data quality, these misclassifications are due to inherent overlap between SNe in feature space.

Jones et al. 2017; Narayan et al. 2018; Pasquet et al. 2019). Although our pipeline is trained and tested on an empirical data set for five distinct SN classes, we find that our achieved purity (≈95%), completeness (≈90%) and accuracy (≈95%) for SN Ia classification are actually comparable to methods that specifically train on the binary classification. However, we note that studies such as Möller et al. (2016) achieve this high purity rate without redshift information, which our method currently requires.

The vast majority of previous photometric classification studies used simulated data sets to train classifiers. This is largely due to the fact that few homogeneous photometric data sets with large numbers of spectroscopically classified SNe exist. Most studies that train on simulated data sets use the Supernova Photometric Classification Challenge (SNPCC) training set (Kessler et al. 2010). The SNPCC data set consists of 20,000 simulated SNe with *griz* LCs, generated from templates of SNe Ia, Ibc, IIP, and IIn (they do not include SLSNe). This data set was presented as a community-wide classification challenge in preparation for the Dark Energy Survey, and was widely successful, with the top algorithms reaching an average SN Ia classification purity of ≈80% and completeness of ≈95%. Works such as Möller & de Boissière (2019) and Moss (2018) have reported average classification accuracies of ≈90% for CCSNe classes (similar to our reported accuracies here). Similarly, Lochner et al. (2016) report an average Type Ia classification accuracy of ∼84% using SALT2 LC features. They further break down the CCSNe subclass into Type Ibc and Type II, where they report accuracies of ∼63% and ∼93%, respectively. We caution that the SNPCC data set is not representative of the real diversity we encounter in ongoing and future surveys, and should not be used as a benchmark for CCSN classification. In particular, to generate synthetic LCs, Kessler et al. (2010) fit well-sampled real LCs from each CCSN class with a Bazin function. Then they stretch Nugent CCSN templates[16] to match the Bazin LCs. Variations within each class are included from both the sample of templates available and from random color variations derived from the Hubble scatter of SNe Ia and the peak luminosity derived from Richardson et al. (2002). While the collection of simulated SNe Ia likely samples the full phase space of LCs, the non-Type Ia templates used to build the model LCs were severely limited. For example, only two Type IIn SN templates were used to generate 800 template LCs, and only 16 Type Ibc SN templates were used to generate 3200 LCs. Because of this, we can expect methods that rely on this data set to overestimate the accuracy of classifications for CCSN classes.

A new classification challenge, PLAsTiCC (Allam et al. 2018; Kessler et al. 2019), is a more realistic simulated data set that can be used as a benchmark for CCSN classification, although it too largely relies on theoretical models. Recent work by Muthukrishna et al. (2019) find an average completeness of ≈65% over the five SN classes that we have classified here (although we note that the PLAsTiCC challenge combines Type IIP/L and SNe IIn into one class). Our average completeness is significantly higher, at ≈77%.

7. Limitations and Future Directions

The challenge of photometric classification for optical transients is broad and cannot be solved with one classification method alone. Like all methods, our classification pipeline aims to solve a simplified version of this problem: given a complete LC, a redshift, and a list of SN classes, what is the type of a given transient? Here we highlight several improvements that can be made to our pipeline, and more broadly outline outstanding problems in the field of transient classification.

Our pipeline requires a redshift, which simplifies the problem of classification by anchoring the absolute magnitudes of every LC. In our training set these redshifts were obtained from spectra of the transients and their host galaxies. However, in ongoing and future surveys we expect that spectroscopic redshifts (from the SNe or host galaxies) will be rare. On the other hand, LSST will provide photometric redshifts (photo-z) for all galaxies with $m < 27.5$ mag, with an expected root-mean-square scatter of $\sigma_z/(1+z) \lesssim 0.05$ for galaxies with $m < 25.3$ (Abell et al. 2009), and a fraction of outliers of <10% (Graham et al. 2018). A classification algorithm that can

[16] https://c3.lbl.gov/nugent/nugent_templates.html

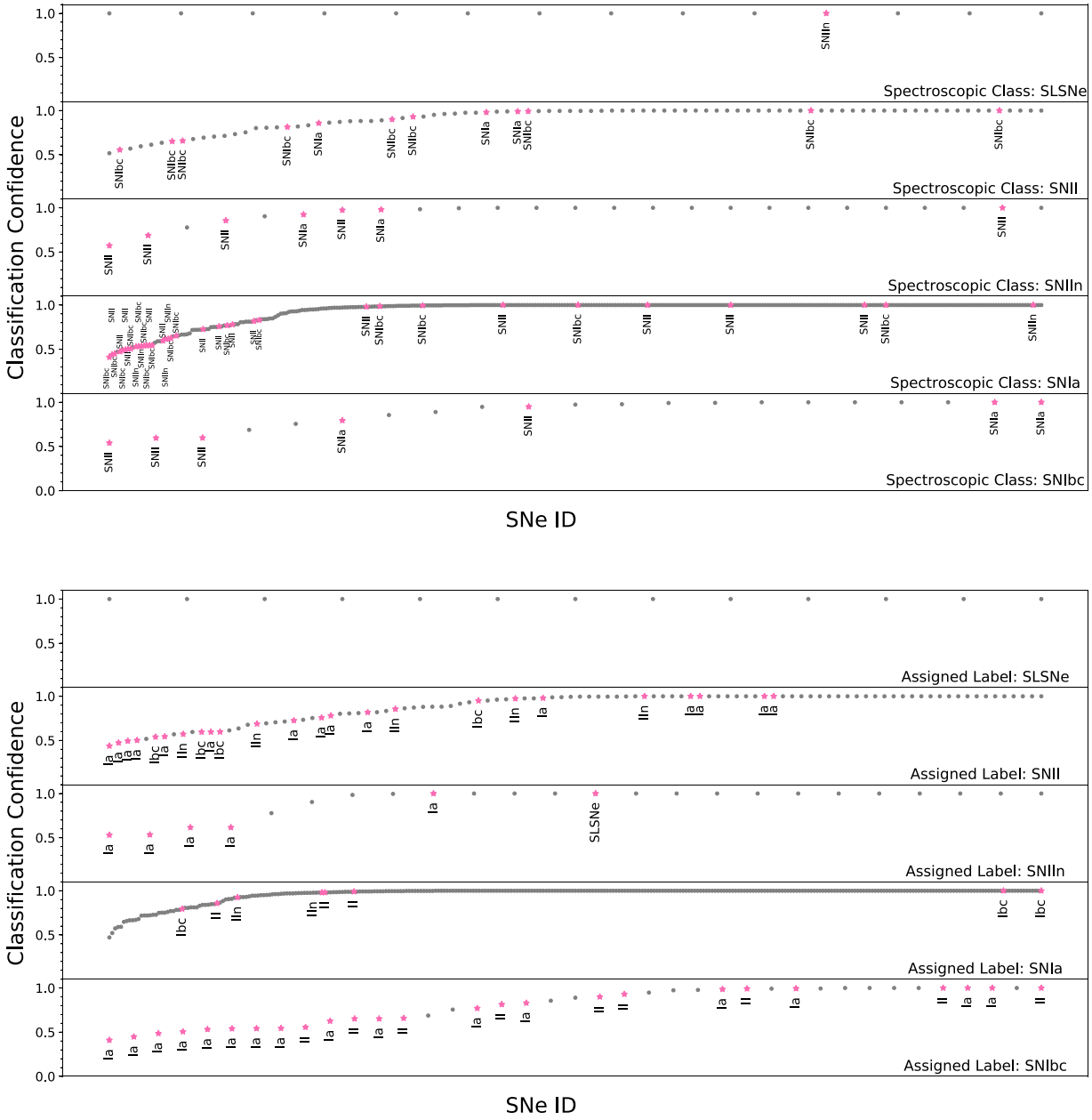

Figure 13. Top: for each *true* spectroscopic class, we show the correct classifications (gray) and misclassifications (pink), with the classification confidence plotted on the ordinate. The misclassified label is given next to each misclassified event. We find that the bulk of the misclassifications are concentrated at low classification confidence. Bottom: for each *assigned* label, we again show the correct classifications (gray) and misclassifications (pink). The correct class label of each misclassified event is given.

associate a transient to its host galaxy will therefore be able to utilize the photo-z value. We anticipate that the additional uncertainty in the model fits due to the photo-z uncertainty will not be a dominant factor. We additionally note that by including redshift information as a feature (even when doing so indirectly) we have limited the use of our pipeline to surveys of similar depth.

Additionally, our classification pipelines best utilizes full LCs, and are thus most naturally applicable for after-the-fact classification. The most natural use is on the yearly samples of $\sim 10^6$ transients from LSST to enable large-scale population studies, as well as targeted studies of specific subsets (e.g., host galaxies of SLSNe). While our method can work on partial LCs for real-time classification, its performance in this context is yet to be evaluated. Several studies that have explored the specific issue of real-time classification have found that recurrent neural networks perform well for this purpose (e.g., Charnock & Moss 2017; Möller & de Boissière 2019; Muthukrishna et al. 2019).

Our algorithm currently relies exclusively on information derived from the transient LCs (other than the redshift). However, useful contextual information about an SN can be extracted from the host galaxy. For example, SLSNe prefer low-metallicity, dwarf galaxies (Lunnan et al. 2014), other CCSN classes span a wide range of star-forming galaxies, and SNe Ia are found in both star-forming and elliptical galaxies. Simple galaxy features, such as Hubble type, color, and SN offset can be easily incorporated into the classification pipeline (e.g., Foley & Mandel 2013). This will be explored in follow-up work.

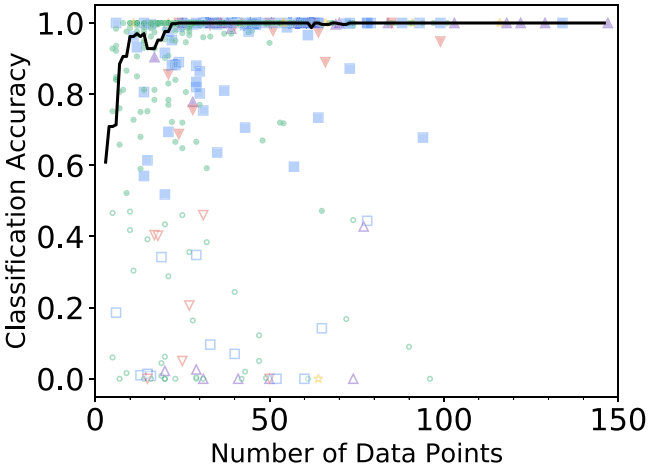

Figure 14. Classification accuracy as a function of number of LC data points. The colors and shapes reflect the SN classes, and the black line represents a smoothed median to guide the eye. Filled symbols are SNe classified correctly, while open symbols are misclassified events. We find that misclassifications are more prevalent for LCs with fewer points, but also that some events are misclassified even with tens of data points, as also highlighted in Figure 12.

Furthermore, our algorithm is limited to classification within known SN classes (in this case five classes). To add additional classes under our current framework, we would need to incorporate new data into the training set and retrain the classification algorithms. Our pipeline is amenable to rapid training, so it is feasible to incorporate more classes in this way. For a more complex classification pipeline (e.g., one involving a large neural network), one could incorporate new classes cheaply using "one-shot" learning (Lv et al. 2006), in which a classifier learns the characteristics of a new class using very limited examples. However, the addition of new classes will not solve the issue of how to identify unforeseen classes of transients and entirely new phenomena. Such a classifier is challenging to train, since outlier events are (by definition) rare.

Because our original training set is imbalanced and small, we needed to augment our data set with simulated events drawn from the observed populations. For completeness, we test our best classification pipeline (PCA feature extraction and RF classifier) on the original training set without data augmentation. As expected, we find that we can classify classes with the most samples (Type Ia and II SNe) or those that are well-separated in feature space (SLSNe), as well as or better than our classification pipeline with data augmentation. However, the completeness of the minority classes, like Type Ibc and IIn SNe, falls by 20%–40%. This is a good indication that data augmentation in the extracted feature space is a potential solution to the imbalanced classes.

Our method neglects the possibility of a biased spectroscopic sample, for example, if the spectroscopic sample contains only the brighter end of the luminosity function for rare transients. In our presumed classification case in which we have access to the full LCs, one can use the full data set to detect and minimize the effects of selection bias without knowing the true underlying distribution. For example, one can re-weight the importance of each SN in the spectroscopic training sample to better reflect the distribution of features from the full data set (using, e.g., Huang et al. 2007 and Cortes et al. 2008). A detailed study of the effect of observational biases on transient classification is essential, but beyond the scope of this work.

Finally, we note that classification is only the first step in understanding the uncovered transients. Even for the currently rare SLSN class, LSST will discover $\sim 10^4$ events per year (Villar et al. 2018). Additional data cuts that remove LCs with a minimal *information content* (or those from which we cannot extract physical parameters) may be necessary in order to realistically fit a representative set of LCs.

8. Conclusions

Given increasingly large data sets and limited spectroscopic resources, photometric classification of SNe is a pressing problem within the wide scope of time-domain astrophysics. Here we used the PS1-MDS spectroscopiccaly classified SNe data set (513 events) to test a number of classification pipelines, varying the features extracted from each LC, the augmentation method to bolster the training set, and the classification algorithms. We used a flexible analytical model with an iterative MCMC process to model the $g_{\rm P1}$, $r_{\rm P1}$, $i_{\rm P1}$, $z_{\rm P1}$ LCs of each event, and to generate posterior distributions. We find that several pipelines (e.g., PCA feature extraction, MVG resampling, and RF classifier) perform well across the five relevant SN classes, achieving an average accuracy of about 90% and a SN Ia purity of about 95%.

Our study is the first to use an empirical data set to classify multiple classes of SNe, rather than just Type Ia versus non-SN Ia classification. Our overall results rival similar pipelines trained on simulated SN data sets, as well as those that utilize only a binary classification. This indicates that we can use this approach to generate robust samples of both common and rare SN type (e.g., Type IIn, SLSNe) from the LSST.

Finally, we highlight several areas for future exploration and improvement of our classification approach, including the use of contextual information and the possible application to real-time classification. We plan to extend this work and other classification approaches to the full set of PS1-MDS SN photometric LCs in future work.

The Berger Time-Domain Group is supported in part by NSF grant AST-1714498 and NASA grant NNX15AE50G. V.A.V. acknowledges support by the National Science Foundation through a Graduate Research Fellowship. The UCSC team is supported in part by NASA grant NNG17PX03C; NSF grants AST-1518052 and AST-1815935; the Gordon & Betty Moore Foundation; the Heising-Simons Foundation; and by a fellowship from the David and Lucile Packard Foundation to R. J.F. R.L. is supported by a Marie Skłodowska-Curie Individual Fellowship within the Horizon 2020 European Union (EU) Framework Programme for Research and Innovation (H2020-MSCA-IF-2017-794467). Some of the computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University. The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central

University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

Facilities: ADS, Pan-STARRs.

Software: Astropy (Astropy Collaboration 2018), Matplotlib (Hunter 2007), NumPy (van der Walt et al. 2011), SciPy (Oliphant 2007), Scikit-learn (?), SNID (Blondin & Tonry 2007).

ORCID iDs

V. A. Villar https://orcid.org/0000-0002-5814-4061
E. Berger https://orcid.org/0000-0002-9392-9681
E. Magnier https://orcid.org/0000-0002-7965-2815
D. Milisavljevic https://orcid.org/0000-0002-0763-3885

References

Abell, P. A., Burke, D. L., Hamuy, M., et al. 2009, arXiv:0912.0201
Aizerman, M. A. 1964, Automation and Remote Control, 25, 821
Allam, T., Jr, Bahmanyar, A., Biswas, R., et al. 2018, arXiv:1810.00001
Astropy Collaboration 2018, AJ, 156, 123
Bazin, G., Palanque-Delabrouille, N., Rich, J., et al. 2009, A&A, 499, 653
Blondin, S., & Tonry, J. L. 2007, ApJ, 666, 1024
Breiman, L. 2001, Machine Learning, 45, 5
Campbell, H., DAndrea, C. B., Nichol, R. C., et al. 2013, ApJ, 763, 88
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, J. Artif. Intell. Res., 16, 321
Chornock, R., Berger, E., Gezari, S., et al. 2014, ApJ, 780, 44
Cortes, C., Mohri, M., Riley, M., & Rostamizadeh, A. 2008, arXiv:0805.2775
Dhawan, S., Leibundgut, B., Spyromilio, J., & Maguire, K. 2015, MNRAS, 448, 1345
Drout, M. R., Chornock, R., Soderberg, A. M., et al. 2014, ApJ, 794, 23
Drout, M. R., Soderberg, A. M., Gal-Yam, A., et al. 2011, ApJ, 741, 97
Filippenko, A. V. 1997, ARA&A, 35, 309
Gezari, S., Chornock, R., Rest, A., et al. 2012, Natur, 485, 217
Graham, M. L., Connolly, A. J., Ivezić, Ž, et al. 2018, AJ, 155, 1
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
Hoyle, B., Rau, M. M., Bonnett, C., Seitz, S., & Weller, J. 2015, MNRAS, 450, 305
Huang, J., Gretton, A., Borgwardt, K. M., Schölkopf, B., & Smola, A. J. 2007, in Advances in Neural Information Processing Systems, ed. B. Schölkopf, J. Platt, & T. Hofmann (Cambridge, MA: MIT Press), 601
Hunter, J. D. 2007, CSE, 9, 90
Ishida, E. E., & de Souza, R. S. 2013, MNRAS, 430, 509
Ishida, E. E. O., Beck, R., González-Gaitán, S., et al. 2019, MNRAS, 483, 2
Ivezic, Z., Axelrod, T., Brandt, W., et al. 2008, SerAJ, 176, 1
Jedicke, R., Tonry, J., Veres, P., et al. 2012, AAS/DPS Meeting, 44, 210.12
Jones, D., Scolnic, D., Riess, A., et al. 2017, ApJ, 843, 6
Jones, D., Scolnic, D., Riess, A., et al. 2018, ApJ, 857, 51
Kaiser, N., Burgett, W., Chambers, K., et al. 2010, Proc. SPIE, 7733, 77330E
Karpenka, N. V., Feroz, F., & Hobson, M. P. 2013, MNRAS, 429, 1278
Kasen, D. 2006, ApJ, 649, 939
Kessler, R., Bassett, B., Belov, P., et al. 2010, PASP, 122, 1415
Kessler, R., Narayan, G., Avelino, A., et al. 2019, PASP, 131, 094501
Kimura, A., Takahashi, I., Tanaka, M., et al. 2017, arXiv:1711.11526
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Kulkarni, S. 2018, ATel, 11266
Leo, B., Friedman, J. H., Olshen, R. A., & Stone, C. J. 1984, Classification and Regression Trees (Belmont, CA: Wadsworth), 151
Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, ApJS, 225, 31
Lunnan, R., Chornock, R., Berger, E., et al. 2014, ApJ, 787, 138
Lunnan, R., Chornock, R., Berger, E., et al. 2018, ApJ, 852, 81
Lv, F., Zhao, T., Nevatia, R., et al. 2006, ITPAM, 28, 1513
Magnier, E., Schlafly, E., Finkbeiner, D. P., et al. 2016a, arXiv:1612.05242
Magnier, E. A., Chambers, K., Flewelling, H., et al. 2016b, arXiv:1612.05240
Mandel, K. S., Narayan, G., & Kirshner, R. P. 2011, ApJ, 731, 120
Miknaitis, G., Pignata, G., Rest, A., et al. 2007, ApJ, 666, 674
Möller, A., & de Boissière, T. 2019, arXiv:1901.06384
Möller, A., Ruhlmann-Kleider, V., Leloup, C., et al. 2016, JCAP, 2016, 008
Muthukrishna, D. 2016, arXiv:1903.02557
Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, arXiv:1904.00014
Narayan, G., Zaidi, T., Soraisam, M. D., et al. 2018, ApJS, 236, 9
Newling, J., Varughese, M., Bassett, B., et al. 2011, MNRAS, 414, 1987
Nicholl, M., Guillochon, J., & Berger, E. 2017, ApJ, 850, 55
Oliphant, T. E. 2007, CSE, 9, 10
Pasquet, J., Chaumont, M., & Fouchez, D. 2019, A&A, 627, A21
Patil, A., Huard, D., & Fonnesbeck, C. J. 2010, J. Stat. Software, 35, 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
Richards, J. W., Homrighausen, D., Freeman, P. E., Schafer, C. M., & Poznanski, D. 2011, MNRAS, 419, 1121
Richardson, D., Branch, D., Casebeer, D., et al. 2002, AJ, 123, 745
Sanders, N. E., Soderberg, A. M., Gezari, S., et al. 2015, ApJ, 799, 208
Schmidhuber, J. 2015, NN, 61, 85
Scolnic, D., Jones, D., Rest, A., et al. 2018, ApJ, 859, 101
Scolnic, D., Rest, A., Riess, A., et al. 2014, ApJ, 795, 45
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757
Stubbs, C. W., Doherty, P., Cramer, C., et al. 2010, ApJS, 191, 376
Taddia, F., Stritzinger, M., Sollerman, J., et al. 2013, A&A, 555, A10
Tonry, J., Stubbs, C. W., Lykke, K. R., et al. 2012, ApJ, 750, 99
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Villar, V. A., Berger, E., Metzger, B. D., & Guillochon, J. 2017, ApJ, 849, 70
Villar, V. A., Nicholl, M., & Berger, E. 2018, ApJ, 869, 166
Waters, C., Magnier, E., Price, P., et al. 2016, arXiv:1612.05245