Draft version May 18, 2024
Typeset using LaTeX **twocolumn** style in AASTeX631

The Impact of Host Galaxy Properties on Supernova Classification with Hierarchical Labels

Ashley V. Villar,[1] Sebastian Gomez,[2] Edo Berger,[1] and Alex Gagliano[1]

[1] *Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138-1516, USA*
[2] *Space Telescope Science Institute, 3700 San Martin Dr, Baltimore, MD 21218, USA*

ABSTRACT

With the advent of the Vera C. Rubin Observatory, the discovery rate of supernovae will surpass the rate of supernovae with real time spectroscopic followup by three orders of magnitude. Classification methods, especially those which classify supernovae accurately and early, are essential in the selection of which supernovae to actively observe with spectroscopic and multi-wavelength instruments. The host galaxies of supernovae provide early clues about the classification of transient, even before any supernova light has been observed. In this work, we investigate the impact of observable host galaxy information on the classification of supernovae, both with and without additional light curve and redshift information. We find that host galaxy information alone can successfully isolate pure samples of Type Ia/Core-collapse SNe; however, classification of rare subclasses of core-collapse supernovae is more challenging, with Type Ia SNe being the major contaminator of CCSN classes. We find that, given redshift information, host galaxy properties do not significantly increase the accuracy of SN classification when paired with complete light curves. In the absence of redshift information, however, galaxy properties significantly increase the accuracy of photometric classification. As a part of this analysis, we present a new objective function, the hierarchical cross-entropy, to the problem of supernova classification. This objective function more naturally accounts for the hierarchical nature of supernova classes and, more broadly, transients. Finally, we present a new set of SN classifications for the Pan-STARRS Medium Deep Survey of supernovae which do not rely on redshift information.

Keywords: Supernovae (1668) — Light curve classification(1954) — Neural networks (1933)

1. INTRODUCTION

Early classification of extragalactic transients, in particular supernovae (SNe), is paramount to enabling multi-wavelength and spectroscopic analysis in real time. Currently, $\simeq 5-10$ percent of SNe receive a spectroscopic classification, the traditional means for understanding the underlying nature of a SN. The fraction of SNe which remain spectroscopically unclassified will significantly grow with the upcoming Legacy Survey of Space and Time (LSST) conducted by Vera C. Rubin Observatory and which is expected to commence in 2025. LSST will discover over one million SNe annually (Kessler et al. 2019); without additional spectroscopic resources, $\lesssim 0.1\%$ of all SNe will have spectroscopic followup. As a result of this growth, the time-domain community has placed significant emphasis on *photometric* classifiers in the years leading up to LSST.

Photometric classifiers aim to classify SNe into their historically spectroscopic classes based on photometric data alone. The taxonomy of SNe has actively evolved as increasingly large samples unveil new diversity in SNe observables, leading to a branching hierarchical structure. Broadly speaking, Type II SNe are those which show spectroscopic signatures of hydrogen near the peak of their optical emission. Within this class, the spectra of Type IIP/L SNe have broad H (typically with P-Cygni profiles), while Type IIn SNe are dominated by narrow H emission. In contrast, Type I SNe lack said features, with Type Ib SNe showing signs of helium and Type Ic SNe lacking signs of either. Type Ibc and II SNe arise from the core-collapse of massive stars. Type Ia SNe, arising from thermonuclear explosions of white dwarfs, lack H and He in their optical spectra but show strong signatures of Si II near peak. A broad overview of the hierarchical SN taxonomy is presented in Gal-Yam (2016).

Some photometric classification methods can be conducted in real time (e.g., Muthukrishna et al. 2019; Möller & de Boissière 2020; Carrasco-Davis et al. 2021; Qu & Sako 2022; Gagliano et al. 2023) while others rely on complete light curves (e.g., Hosseinzadeh et al. 2020;

Boone 2021) for feature extraction. It can be especially challenging to identify and measure useful SN features in the early photospheric phases of the light curves. However, prompt identification of SNe is key to capture early observational phenomena (e.g., flash spectroscopy for young core-collapse SNe; Khazov et al. 2016; Bruch et al. 2021), to guide multiwavelength follow up and to optimize spectroscopic resources. In order to perform this early classification, *all* available information must be utilized, including the *contextual* information provided by the host galaxy of the transient.

Specifically, host galaxy information is known to correlate with transient properties. For example, while thermonuclear (Type Ia) SNe show little preference to host galaxy type (Childress et al. 2013), core-collapse supernovae (CCSNe) preferentially occur in galaxies with recent or ongoing star formation. Some stripped-envelope SNe (SESNe) prefer slightly more massive (i.e., more metal rich) host galaxies compared to their hydrogen-rich counterparts of Type II(b) SNe (Schulze et al. 2021). Similarly, SESNe are over-represented in disturbed galaxies compared to their H-rich counterparts (Habergham et al. 2012). Rare CCSN classes seem to show stronger and more exotic preferences. Both Type Ic-BL (high-energy, "broad-lined" events) and H-poor superluminous SNe (SLSNe) prefer low mass (metal poor) galaxies with high specific star formation rates (Kelly & Kirshner 2012; Schulze et al. 2021). On local scales, SESNe strongly trace Hα (a tracer of ongoing star formation), while Type II SNe tend to show larger spreads in local environment properties Anderson et al. (2012). In contrast the Type IIn SNe, which show signs of interaction between the SN blast wave and dense, pre-existing circumstellar material, show highly heterogeneous local environments (Ransome et al. 2022). Rare Ca-rich SNe of yet unknown progenitor origin show strong preference for high offsets in their host galaxies, although are not preferentially found in star-forming hosts (Dong et al. 2022). These correlations are clues toward the underlying progenitor populations for each of these SN classes. Here, however, we utilize the correlations between global host properties and transients as a tool to help classify the underlying physics of the SNe.

Galaxy-based classifiers for SN identification have been explored in the literature, although they have been primarily limited to binary classification tasks. Foley & Mandel (2013) introduced the first galaxy-based SN classifier based on a Naive Bayes architecture with the LOSS SN sample. In a similar vein, Baldeschi et al. (2020) showed that galaxy morphology and star formation could be used to increase the purity of thermonuclear and CCSNe samples compared to randomly guessing. Gagliano et al. (2021) introduced a random forest classifier based on host galaxy properties from the `GHOST` sample of $\sim 16,000$ SNe associated with host galaxies, finding that such a method could perform thermonuclear vs CCSNe classification with $\simeq 68\%$ accuracy. Gomez et al. (2020) presented one of the first classifiers which extend beyond Ia/CCSNe classification, focusing instead on Type I SLSNe vs non-SLSN classification from a combination of host and light curve properties; a classifier based on host galaxy properties alone is not explored in that work. Recently, Kisley et al. (2022) presented a hierarchical SN classifier based solely on optical and (near) infrared host galaxy photometry from the THEx catalog (Qin et al. 2022). They use a likelihood-based approach reminiscent of a Naive Bayes classifier, using a series of binary classifications to distinguish 11 SN classes (in addition to tidal disruption events). While they were able to classify Type Ia and Type II SNe at an accuracy statistically above random guessing, they were unable to do so for any other SN subclass. Most recently, Gagliano et al. (2023) presented a neural network-based classifier for Zwicky Transient Facility (ZTF) Bright Transient Survey data which utilizes the light curve and host galaxy information for Type Ia, II and Ibc SN classification, achieving an accuracy of $\simeq 82\%$ within 3-days (observer-frame) of discovery.

Furthermore, the hierarchical nature of SN taxonomy is rarely used in photometric classification. Sánchez-Sáez et al. (2021) presents a broad hierarchical classifier for ZTF data. There, multiple"flat" (i.e., those lacking a hierarchical structure) random forest algorithms are trained at each "level" of the hierarchical classifier. For example, all SNe are classified as transient phenomena, then a second flat classifier is trained to classify them as one of four SN types. Kisley et al. (2022) consider the hierarchical nature of SNe in the sense of a prior distribution, which is then used in the classifier's objective function. For example, the rate of Type Ia-91bg events are included when calculating the rate of Type Ia SNe, which is then used to define a Type Ia prior probability within the objective function. However, this prior information is not used in their primary analysis.

Here, we present a broad, hierarchical classifier based on host galaxy properties alone, in addition to how galaxy properties can improve our transient classifications. We specifically test how simple, measurable properties of the host impact our classification accuracy both with and without redshift information. The paper is organized as follows. In Section 2, we review the data used in this work and the features selected for our SN classifiers. In Section 3 we introduce a novel objective function, the hierarchical cross entropy score, to more

naturally account for the hierarchical nature of SN classification. We additionally discuss the architecture of our neural network-based classifier used here for classification. In Section 4, we discuss the results of our classifiers for different subsets of SN classes, feature sets and the in/exclusion of redshift information. In Section 6, we present new classifications of SNe within the Pan-STARRs Medium Deep Survey sample. We conclude in Section 7. Throughout this work, we assume a flat ΛCDM cosmology with $H_0 = 69.3$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.286$, and $\Omega_\Lambda = 0.712$ (Hinshaw et al. 2013).

2. DATA & FEATURE SELECTION

We aim to understand the interplay between host galaxy features and SN light curve features. Therefore, we require a SN sample for which both a light curve *only* classifier and a host galaxy classifier are both available. While there now exist two large catalogs of host galaxies with associated SNe, GHOST Gagliano et al. (2021) and THEx Qin et al. (2022), we instead opt to use a uniform survey in order to directly compare the light curve vs host features. For this reason, we utilize the Pan-STARRS Medium Deep Survey catalog of SN-like light curves recently published by Villar et al. (2020) and Hosseinzadeh et al. (2020). In total, our sample includes 557 spectroscopically identified SNe, in five classes: Type Ia SNe (404 objects), Type II SNe (93 objects), Type IIn SNe (24 objects), Type Ib/c SNe (19 objects), and Type I SLSNe (17 objects). Our Pan-STARRS sample has the additional benefit of being a close analogue to Vera C. Rubin data in terms of cadence, filter selection and depth, offering a realistic view of how our algorithm will perform on the LSST datstream.

For our light curve features, we use the full set available from `SuperRAENN` (44 in total; Villar et al. 2020), which includes:

1-8: Non-linear features extracted from an autoencoder, trained on the PS1-MDS dataset. The correlations between these data-driven features and "typical" observational features are explored in more detail in Villar et al. (2020).

9-21: The *griz* rise times, calculated 1, 2, and 3 magnitudes below peak. Note that these are calculated using light curves which have been interpolated via a 2D Gaussian Process. These are, when redshift is utilized, calculated in the rest frame of the SN.

22-32: The *griz* decline times, calculated for 1, 2 and 3 magnitudes beow peak.

33-36 : The *griz* peak magnitude. When redshift is utilized, these are absolute magnitudes.

37-40: The median slope measured in *griz* between 10 and 30 days post-peak in the observer frame. Note that these features in particular help distinguish between Type II and Type Ibc SNe.

41-44: The integral of the interpolated *griz* light curves.

To associate SNe with host galaxies and extract galaxy features, we use the "Finding Luminous and Exotic Extragalactic Transients" (FLEET) pipeline (Gomez et al. 2020). FLEET is a classification methodology designed to classify rare, extragalactic transients (SLSNe and tidal disruption events) using a combination of host galaxy and light curve information. FLEET queries a 1′ of the Pan-STARRS 3π survey around a given transient to identify the most likely host galaxy using the follow algorithm. First, a probability of being a galaxy (as opposed to a star) is assigned to every object in the field, where 0 means most like a star, and 1 means most likely a galaxy. This probability is estimated using a custom k-nearest-neighbors algorithm trained on data from the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS), which has complete star/galaxy labels to ≈ 26 mag (Hudelot et al. 2012). Then, the probability of chance coincidence P_{cc} of every galaxy is calculated using the Bloom et al. (2002) method described in Berger (2010). We select the galaxy with the lowest P_{cc} as the host galaxy of the transient.

Many host galaxy features are available via FLEET, including multi-survey observables (e.g., magnitudes), derived properties (e.g., photo-z), and inferred properties more directly related to the transient (e.g., offset). We restrict ourselves to data from the Pan-STARRS 3Pi survey, in order to minimize missing features in our dataset. We use the following features from the pipeline:

1-5: *g*-, *r*- *i*-, *z*- and *y*-band Kron magnitudes, estimated using `sextractor`.

6-8: *g*-, *r*- and *i*-band Kron radii, estimated using `sextractor`.

9: Host galaxy separation from the transient in arcseconds.

10: Point source score, a measurement of probability of the object being an extended object (score of 0) or a point source (score of 1). This is a property within the PS1 3Pi Survey.

11: Number of detections in all bands.

12: Host galaxy half-light radius in *i*-band or in *r*-band if *i*-band is not available.

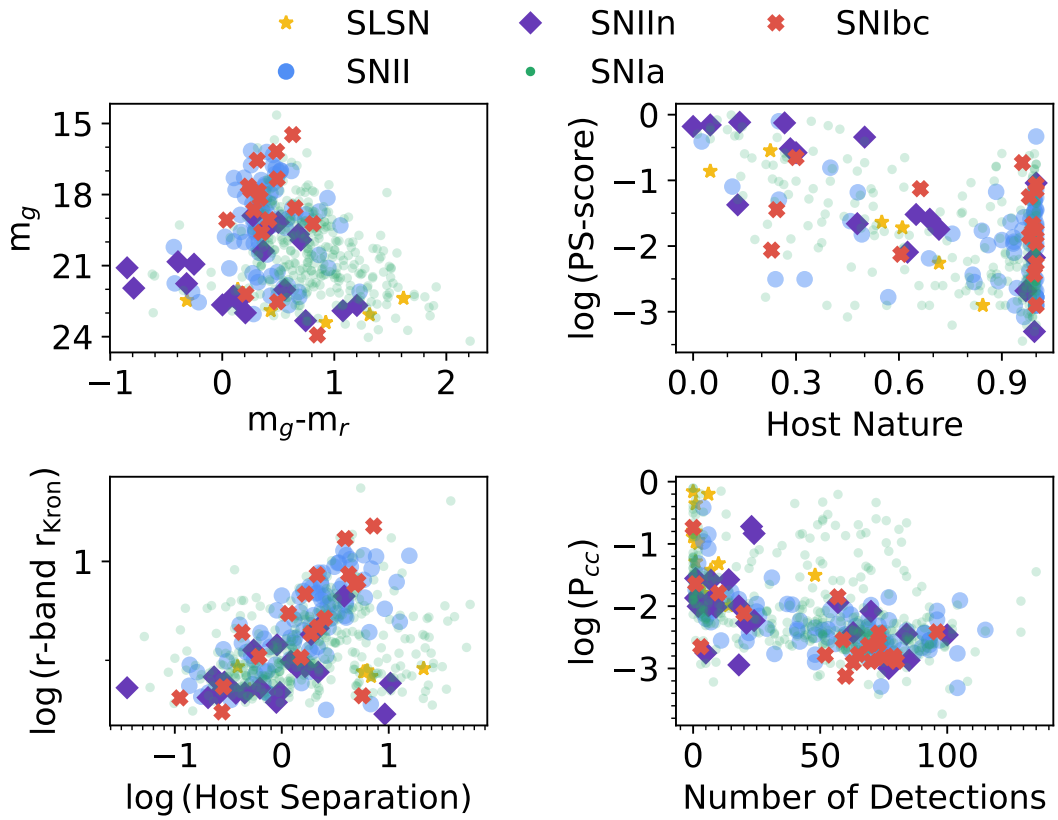


Figure 1: Scatter plot of various host galaxy features to classify SNe in this work (see Sec. 2 for feature definitions). SNe are *spectroscopically* labelled, and for features where the host galaxy features were not observed, they are inferred via K-means imputation. Even in these simple feature spaces, clear clustering is seen.

13: Host galaxy nature, a measurement of probability of the object being a galaxy (score of 1) or a star (score of 0). This is a custom function within FLEET.

14: Probability of chance coincidence, as calculated by FLEET.

All magnitudes are corrected for Milky Way extinction. In some of our pipelines, we additionally include the redshift (spectroscopically measured from the host or transient) as a feature.

Although we attempt to minimize missing data, some features are not observed for every event. We use a K-means imputation method (`KNNImputer` in `scikit-learn`) to fill in the missing data. This method utilizes the information from the $K = 3$ neighbors in the 13-dimensional galaxy feature space, and fills in missing values with an average from these neighbors. This imputation method works as expected with no ad-hoc corrections: For hosts which are not detected due to missing data (e.g., they were never observed in z-band), the nearest neighbor provides a reasonable estimate of the missing properties. For hosts near the survey limit, the method naturally fills in the missing data with dim apparent magnitudes and small observed radii. We note that even hostless transients are included in our classifier; in general, this will impact the classification of SNe with high-z or low-luminosity hosts. In testing, we find that changing the value of K has minimal impact on results. Finally, we normalize our data such that each feature values between 0 and 1 (`MinMaxScaler` in `scikit-learn`).

We note that, even without our nonlinear, neural network classifier, there are clear correlations between SN type and host galaxy properties. We show a number of representative feature spaces in Fig. 1. As expected, brighter objects (SLSNe, Type IIn SNe) tend to occur in dimmer galaxies (i.e., those are higher redshift), while dimmer objects (like normal Type Ibc SNe) tend to lie in brighter host galaxies. However, the *color* of these galaxies also clearly strongly correlates with type. For any given magnitude, CC SNe (especially Type II SNe) are more likely to occur in bluer galaxies (i.e., those with ongoing star formation).

3. CLASSIFICATION METHODS

3.1. *A Novel Hierarchical Loss Function*

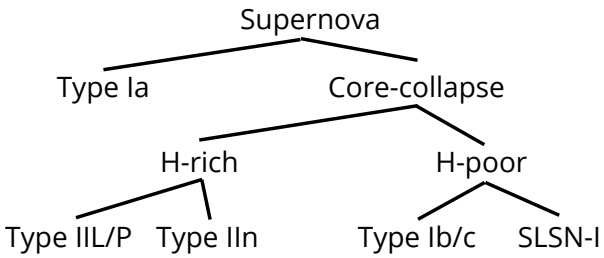


Figure 2: Hierarchical graph structure used in this work.

We explicitly include the hierarchical nature of SN taxonomy within our analysis via a hierarchical cross-entropy (HXE) objective function. Bertinetto et al. (2020) originally introduced the HXE to classify images with a similar hierarchical taxonomy. They compare the performance of HXE compared to standard cross entropy, finding that the two perform similarly in overall accuracy but that HXE "makes better mistakes", i.e., the HXE enforces a graph structure that encourages the classifier to place objects in the correct broader category. In our SN context, the HXE allows us to (1) train a multi-layered classifier that can easily perform binary- and general-classification and (2) encourage transients to be misclassified within their broader class (e.g., CC-SNe are less likely to be classified as Type Ia SNe). In this Section, we define the graph structure used for our SN classification and HXE in detail. We note that recently, Villar et al. (2023) a weighted version of this objective function and applied it to SN and variable star classification; we largely follow the notation in that work.

The graphical nature of our classification hierarchy is shown in Fig. 2. Given the small dataset size, we have not subdivided Type Ia, Type IIP/L and Type Ib/c SNe into additional subclasses. We emphasize, however, that this is not the *only* choice for a classification graph. For example, some bright Type IIn SNe share similar hosts to SLSNe. One could include galaxy information in the graph structure (e.g., preferring to classify bright transients in low-luminosity galaxies together), although the classification bias would be less physically motivated. Alternatively, one could follow the "classic" classification schema, first distinguishing Type I vs Type II SNe (rather than thermonuclear vs core-collapse). This may aid in isolating hydrogen-rich transients, but is not explored in this work.

Following Villar et al. (2023), we now describe the HXE objective function. The classic categorical cross-entropy, most often used in classification tasks, is defined as:

$$\mathcal{L} = -\sum_{i}^{C} t_i \log(p(c_i)), \tag{1}$$

where C is the total number of classes, c_i is a specific SN class, and t_i is an indicator variable for the *true* SN class. This indicator function means that the cross-entropy score *only* rewards the classifier based on assigning a high likelihood to the true class. There is no reward structure for how probabilities are distributed over the incorrect classes.

We contrast this to the HXE, which will utilize the hierarchical graph structure of the SN taxonomy by factorizing each class by its leaf nodes. We will represent the SN class "height" on the tree as $c_i^{(h)}$. The root (the "SN" parent class) is represented as $c_i^{(H)}$ (where $H = 4$ is the height of the tree), and a lowest leaf (e.g., the Type Ib/c class) is represented as $c_i^{(1)}$. For any given class c_i with height h', the classification probability can be written as:

$$p(c_i) = \prod_{h=h'}^{H-1} p(c_j^{(h)} | c_k^{(h+1)}) \tag{2}$$

where c_j and c_k are ancestor nodes of class c_i (i.e., nodes on the path from the leaf to the root). Here, we are assuming that the probability of being SN (the root node) is equal to one. The conditional probabilities can be written more explicitly as:

$$p(c_j^{(h)} | c_k^{(h+1)}) = \frac{\sum_{c_a \in \mathrm{Siblings}(c_j^{(h)})} p(c_a)}{\sum_{c_b \in \mathrm{Siblings}(c_k^{(h+1)})} p(c_b)}, \tag{3}$$

where $\mathrm{Siblings}(c)$ represent the sibling set of node c (i.e., those sharing the same parent). As an example, the probability that a SN is Type Ib/c given it is H-poor is equal to the probability that it is a Type Ib/c (as opposed to SLSN-I) divided by the probability that it is H-poor (as opposed to H-rich).

Analogous to Eqn.1, we can define the hierarchical cross-entropy (HXE) as:

$$\mathcal{L}_{\mathrm{HXE}}(c_i^{(h)}) = -\sum_{h=h'}^{H-1} \lambda(c^{(h)}) \log p(c_j^{(h)} | c_k^{(h+1)}), \tag{4}$$

where $\lambda(c^{(h)})$ is a weighting which emphasizes each level of the taxonomy tree (as a function of class height, h). Bertinetto et al. (2020) suggest the following form for $\lambda(c^{(h)})$:

$$\lambda(c^{(h)}) = \exp(-\alpha h), \tag{5}$$

where α is a free parameter. Larger values of α weight the top of the hierarchy more strongly (i.e., Ia vs CC

classification); lower values of α weight each level of the hierarchy equally, emphasizing fine-grained classifications; α is therefore a hyperparameter of our model which we optimize.

Based on our graph in Figure 2, our objective function requires nine outputs from our classifier, reflecting a probability at each node of the graph. The top node (the "SN" designation) is always equal to one. A softmax function is then applied to each branch of the tree such that the neural network output can be interpreted as a conditional probability. For example, one output represent the probability[1] of being H-rich and another represents the probability of being H-poor. As these are a pair of branches, these outputs will be re-normalized such that their sum is equal to one. The height of each output is also tracked, with an appropriate weighting (a function of α) applied. A working `PyTorch` implementation of the HXE is presented in Villar et al. (2023) and available via GitHub[2].

3.2. *Multi-layer Perceptron Classifier*

Throughout this work, we use a fully connected multi-layer perceptron (MLP, a simple neural network) to classify the SN subclasses. The MLP transforms an input feature vector into an output probability vector which optimizes the HXE objective function. Between these are a series of "hidden" layers with optimizable weights and nonlinear activation functions. Here, we use the standard Rectified Linear Unit. We optimize the MLP using the standard Adam optimizer (Kingma & Ba 2014), a momentum-based gradient descent algorithm. Our model and training procedure is built in `keras` with a Tensorflow backend.

In total, our hyperparameters of the model are: (1) the hierarchy weighting, $\alpha \in 0.0, 0.1, 1.0, 3.0$; (2) the learning rate, $\beta \in 0.001, 0.005, 0.01, 0.05, 0.1$; (3) the batch size, a number from the set $\in 16, 32, 64, 128, 256$; and (4) the number of neurons per layer, a number from the set $\in 3, 5, 10$. Using each combination of hyperparameters, we train a MLP for 300 epochs with early stopping and select the optimal set for each feature set. We use the standard Adam optimizer, a momentum-based gradient descent algorithm Kingma & Ba (2014). In general, the best hierarchy weighting is typically $\alpha \simeq 1$, a learning rate on the higher end, $\beta \simeq 0.05$, a batch size on the higher end $(128 - 256)$ and the higher end of number of neurons $(5 - 10)$. Our model takes minutes on a CPU to complete training. We also note that to account for class imbalances, we weight the importance of each class in our HXE such that minority classes and majority classes are implicitly equally weighted by the classes, using the standard `sklearn` built-in functions. We note that we also attempted to use synthetic minority oversampling techniques to reweight our training set, finding somewhat worse results.

[1] Note: these are not "true" probabilities as they are not properly calibrated. Instead, they share qualitative similarities to a probability vector: they contain non-negative values which sum to one.

[2] https://github.com/villrv/hxe-for-tda/blob/main/supernova-example-24.ipynb

4. CLASSIFICATION RESULTS & DISCUSSION

Our goal is to understand how contextual host galaxy information improves classification performance of SNe both with and without redshift information. Therefore, we train and compare five classifiers in total: (1) one which uses solely the observer-frame host galaxy information; (2) one which uses the rest-frame light curve and redshift information; (3) one which uses solely the observer-frame light curve information; (4) one which uses the galaxy, rest-frame light curve and redshift information; and (5) one which uses the galaxy information and the observer-frame light curve information.

We track the purity, completeness and F1 score of each classifier, defined as:

$$\text{Purity} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Completeness} = \frac{\text{TP}}{\text{TP} + \text{TN}} \tag{7}$$

$$\text{F1} = \frac{2 \times (\text{Purity} \times \text{Competeness})}{\text{Purity} + \text{Completeness}}. \tag{8}$$

, where TP is the "true positive" rate or the fraction of SNe within a given class correctly identified as belonging to said class. TN is the true negative rate or the fraction of SNe *not* in a given class and correctly identified as *not* being a member of said class. Finally, FP is the false positive rate or the fraction of SNe which are identified as belonging to a given class but in fact are *not* of that class. The F1 score is the harmonic mean of the purity and completeness and a commonly used metric for evaluating classifiers.

We report the purity, completeness and F1 for each SN class, as well as the class-averaged and the "weighted" versions of these metrics. Here, the "weighted" averages re-weight each class to represent the total number of objects in each class. In this case, the statistic is dominated by the majority class, Type Ia SNe. Note that we do not use accuracy to evaluate each classifier, which can be a particularly poor metric of success for highly imbalanced training sets; however, when appro-

Table 1: Classification Performance

		Gal. w/o z	Gal. w/ z	LC w/o z	LC w/ z	Gal.+LC w/o z	Gal. + LC w/ z
Five-way Classification							
SLSN	F1	0.19 (0.03)	0.63 (0.11)	0.23 (0.02)	**0.77 (0.05)**	0.29 (0.05)	0.73 (0.05)
	Purity	0.12 (0.02)	0.50 (0.14)	0.20 (0.02)	**0.73 (0.07)**	0.25 (0.06)	**0.73 (0.06)**
	Completeness	0.47 (0.06)	**0.83 (0.09)**	0.27 (0.04)	0.80 (0.05)	0.33 (0.09)	0.73 (0.08)
II	F1	0.38 (0.02)	0.46 (0.06)	0.52 (0.03)	0.71 (0.03)	0.56 (0.04)	**0.73 (0.02)**
	Purity	0.39 (0.03)	0.41 (0.04)	0.55 (0.05)	**0.79 (0.04)**	0.62 (0.04)	**0.79 (0.03)**
	Completeness	0.36 (0.03)	0.53 (0.14)	0.51 (0.05)	0.65 (0.03)	0.51 (0.05)	**0.67 (0.02)**
IIn	F1	0.16 (0.02)	0.16 (0.03)	0.34 (0.04)	0.38 (0.06)	0.37 (0.04)	**0.41 (0.04)**
	Purity	0.10 (0.02)	0.11 (0.02)	0.29 (0.05)	0.32 (0.08)	0.29 (0.04)	**0.41 (0.06)**
	Completeness	0.33 (0.05)	0.27 (0.10)	0.42 (0.09)	0.46 (0.06)	**0.50 (0.07)**	0.42 (0.05)
Ia	F1	0.70 (0.02)	0.73 (0.04)	0.88 (0.01)	**0.94 (0.02)**	0.91 (0.01)	**0.94 (0.01)**
	Purity	0.86 (0.01)	0.91 (0.01)	0.93 (0.01)	**0.95 (0.01)**	0.94 (0.01)	0.93 (0.01)
	Completeness	0.60 (0.03)	0.61 (0.06)	0.84 (0.02)	0.93 (0.04)	0.89 (0.01)	**0.95 (0.01)**
Ibc	F1	0.16 (0.03)	0.13 (0.05)	0.27 (0.04)	0.31 (0.04)	0.29 (0.03)	**0.36 (0.04)**
	Purity	0.12 (0.03)	0.09 (0.04)	0.19 (0.04)	0.24 (0.04)	0.22 (0.03)	**0.32 (0.04)**
	Completeness	0.26 (0.07)	0.29 (0.19)	**0.47 (0.06)**	**0.47 (0.08)**	0.42 (0.08)	0.42 (0.06)
Average	F1	0.36 (0.01)	0.45 (0.03)	0.46 (0.01)	0.63 (0.02)	0.49 (0.02)	**0.64 (0.02)**
	Purity	0.32 (0.01)	0.41 (0.03)	0.43 (0.02)	0.61 (0.02)	0.46 (0.02)	**0.64 (0.02)**
	Completeness	0.40 (0.02)	0.51 (0.06)	0.50 (0.03)	**0.66 (0.02)**	0.53 (0.03)	0.64 (0.02)
Weighted Avg.	F1	0.60 (0.02)	0.65 (0.03)	0.76 (0.01)	**0.85 (0.02)**	0.79 (0.01)	**0.85 (0.01)**
	Purity	0.69 (0.01)	0.75 (0.01)	0.79 (0.01)	**0.86 (0.01)**	0.81 (0.01)	0.85 (0.01)
	Completeness	0.53 (0.03)	0.58 (0.05)	0.73 (0.02)	0.84 (0.03)	0.77 (0.01)	**0.85 (0.01)**
Three-way Classification							
II	F1	0.44 (0.02)	0.48 (0.04)	0.66 (0.02)	**0.75 (0.04)**	0.71 (0.02)	**0.75 (0.02)**
	Purity	0.36 (0.02)	0.37 (0.04)	0.65 (0.03)	**0.79 (0.07)**	0.72 (0.02)	**0.79 (0.03)**
	Completeness	0.58 (0.03)	0.70 (0.10)	0.68 (0.04)	**0.72 (0.03)**	0.70 (0.03)	**0.72 (0.02)**
Ia	F1	0.68 (0.03)	0.70 (0.05)	0.88 (0.01)	**0.94 (0.02)**	0.91 (0.01)	**0.94 (0.01)**
	Purity	0.88 (0.02)	0.92 (0.01)	0.93 (0.01)	**0.95 (0.01)**	0.94 (0.01)	0.93 (0.01)
	Completeness	0.56 (0.04)	0.57 (0.06)	0.83 (0.02)	0.92 (0.04)	0.88 (0.01)	**0.95 (0.01)**
SE	F1	0.20 (0.02)	0.31 (0.06)	0.30 (0.03)	0.48 (0.04)	0.38 (0.03)	**0.52 (0.03)**
	Purity	0.14 (0.02)	0.22 (0.06)	0.22 (0.03)	0.38 (0.04)	0.29 (0.03)	**0.47 (0.04)**
	Completeness	0.41 (0.06)	0.51 (0.10)	0.47 (0.05)	**0.65 (0.05)**	0.53 (0.06)	0.59 (0.05)
Average	F1	0.49 (0.01)	0.54 (0.03)	0.63 (0.01)	0.73 (0.02)	0.68 (0.01)	**0.74 (0.01)**
	Purity	0.46 (0.01)	0.50 (0.02)	0.60 (0.01)	0.71 (0.03)	0.65 (0.01)	**0.73 (0.02)**
	Completeness	0.52 (0.03)	0.59 (0.05)	0.66 (0.02)	**0.76 (0.03)**	0.70 (0.02)	0.75 (0.02)
Weighted Avg.	F1	0.62 (0.02)	0.66 (0.03)	0.80 (0.01)	**0.87 (0.02)**	0.83 (0.01)	**0.87 (0.01)**
	Purity	0.72 (0.01)	0.75 (0.01)	0.82 (0.01)	**0.88 (0.02)**	0.85 (0.01)	0.87 (0.01)
	Completeness	0.55 (0.03)	0.59 (0.05)	0.78 (0.02)	0.86 (0.03)	0.82 (0.01)	**0.88 (0.01)**
Binary Classification							
Ia	F1	0.54 (0.02)	0.62 (0.03)	0.77 (0.02)	**0.85 (0.03)**	0.81 (0.01)	**0.85 (0.02)**
	Purity	0.44 (0.02)	0.49 (0.03)	0.73 (0.03)	0.85 (0.06)	0.78 (0.01)	**0.88 (0.02)**
	Completeness	0.68 (0.04)	0.84 (0.04)	0.82 (0.02)	**0.86 (0.02)**	0.83 (0.01)	0.82 (0.02)
CC	F1	0.74 (0.02)	0.76 (0.04)	0.90 (0.01)	**0.94 (0.02)**	0.92 (0.01)	**0.94 (0.01)**
	Purity	0.84 (0.01)	0.91 (0.01)	0.93 (0.01)	**0.94 (0.01)**	0.93 (0.01)	0.93 (0.01)
	Completeness	0.65 (0.04)	0.65 (0.05)	0.87 (0.02)	0.94 (0.04)	0.91 (0.01)	**0.96 (0.01)**
Average	F1	0.66 (0.01)	0.72 (0.02)	0.84 (0.01)	**0.90 (0.02)**	0.86 (0.01)	**0.90 (0.01)**
	Purity	0.64 (0.01)	0.70 (0.02)	0.83 (0.01)	0.90 (0.03)	0.86 (0.01)	**0.91 (0.01)**
	Completeness	0.67 (0.03)	0.74 (0.03)	0.85 (0.01)	**0.90 (0.02)**	0.87 (0.01)	0.89 (0.01)
Weighted Avg.	F1	0.69 (0.02)	0.74 (0.02)	0.86 (0.01)	**0.92 (0.02)**	0.89 (0.01)	**0.92 (0.01)**
	Purity	0.73 (0.01)	0.79 (0.01)	0.87 (0.01)	**0.92 (0.02)**	0.89 (0.01)	**0.92 (0.01)**
	Completeness	0.66 (0.03)	0.70 (0.04)	0.86 (0.01)	**0.92 (0.03)**	0.88 (0.01)	**0.92 (0.01)**

NOTE—Classification performance (quantified by accuracy, purity and completeness) using a 5-way, 3-way and 2-way split. Optimal feature sets are bolded for each category.

priate (i.e., when comparing to other works), we report accuracy of specific models.

We test our classifier on three classification tasks: (1) five-way classification (SLSN, Type II, Type IIn, Type Ia and Type Ibc); (2) three-way classification (II, Type Ia, and stripped-envelope); and (3) binary classification (core-collapse vs Type Ia). As described above, the classifier does not need to be retrained for each of these tasks; instead, we calculate the conditional probabilities of each, assigning the final label of each object as the category with the highest conditional probability, e.g. for Type IIn SNe:

$$p(\text{IIn}) = p(\text{CC})P(\text{H-rich}|\text{CC})P(\text{IIn}|\text{H-rich}) \quad (9)$$

In total, we compare 18 combinations of feature sets and output classes. Our results are fully summarized in Table 1. Throughout this section, we will exclude uncertainties (which have been calculated using ten random seeds for each MLP) for purity, completeness and F1 scores. However, these values are listed within Table 1 and are typically $\lesssim 0.05$.

4.1. *Classification Performance with Host Galaxy Information Only*

We first explore the performance of SN classification from host galaxy information, with and without redshift information (the first two columns of Table 1). Our results are primarily visualized in the confusion matrices of Figure 3, for the five-way, three-way and binary classification.

In the five-way task without redshift information, we achieve a poor average F1 score of 0.36 across all classes. We find class completeness rate ranging from 0.28 (for Type Ibc SNe) to 0.59 (for Type Ia SNe). We see a much wider spread in purity, from just 0.11 (for Type Ibc SNe) to 0.86 (for Type Ia SNe). Unsurprisingly, the majority of incorrectly labelled SNe are dominated by Type Ia SNe, making up $\gtrsim 50\%$ of each sample. Including redshift information, our average F1 score greatly improves from 0.36 to 0.45. Classification purity and completeness also increase to 0.41 and 0.51, respectively. The class completeness ranges from 0.27 (for Type IIn SNe) to 0.83 (for SLSNe). The improvement in the latter is not surprising, as SLSNe are intrinsically luminous and can therefore be uniquely found at higher redshift for a magnitude-limited survey like PS1-MDS (and will similarly be true in LSST). The spread in purity is still quite high, ranging from 0.09 (for Type Ibc SNe) to 0.91 (for Type Ia SNe). If we impose a confidence cutoff ($p > 0.8$), we find a substantial improvement in overall performance, with the (unweighted) average purity increasing from 0.32 to 0.41, and the completeness from 0.40 to 0.50, suggesting that purer samples of SNe can be collected with these cuts. Type Ia purity, in particular, increases to 0.94, suggesting that a highly pure sample of cosmological Type Ia SNe can be selected from host galaxy properties alone. Unfortunately, the purity of the minority classes remain low $\lesssim 0.2$ even at this high cutoff.

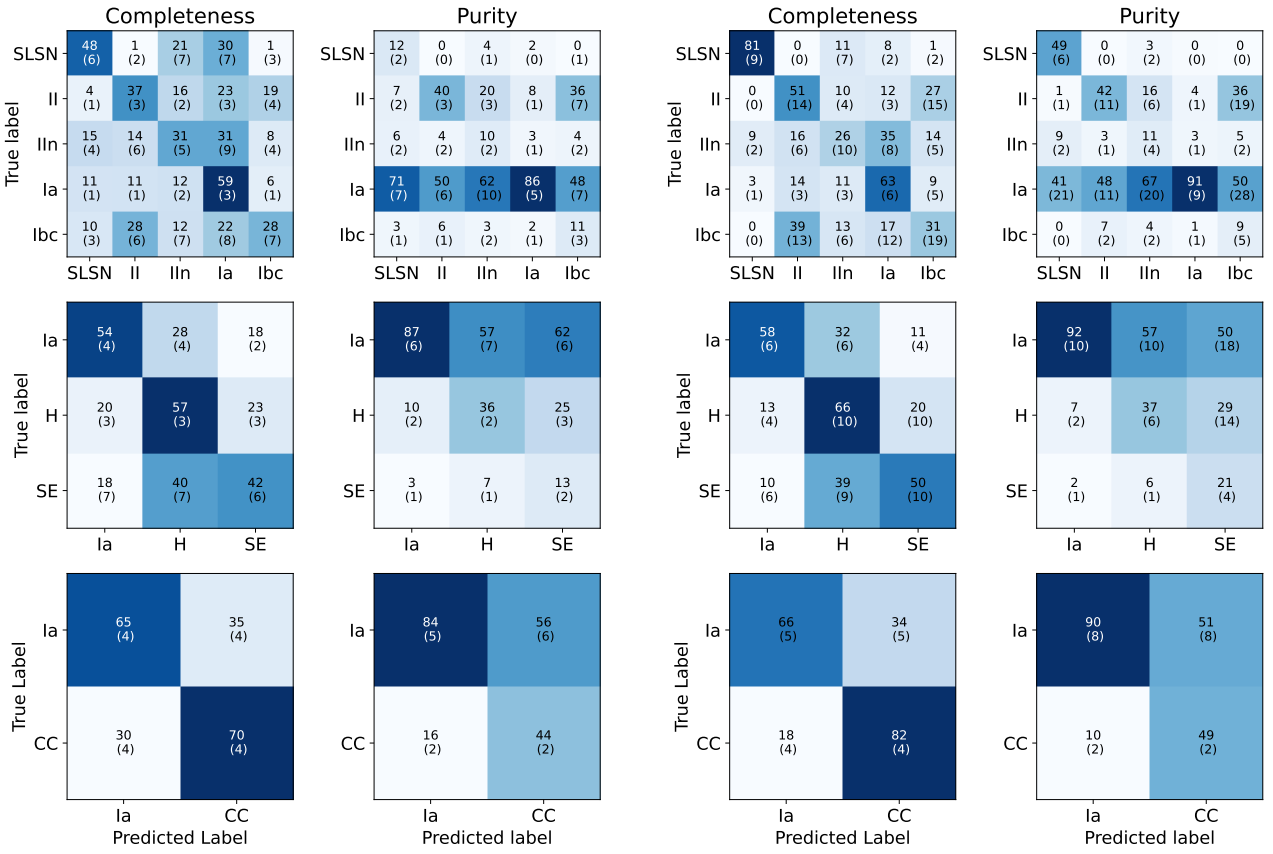
In the three-way classification task, the galaxy-only classifier achieves a notably higher F1 score of 0.49 even when excluding redshift information. The average purity (0.46) and completeness (0.52) also increase. Here, again, the minority classes (stripped-envelope and H-rich SNe) suffer from generally low purity scores (0.14), while Type Ia SNe can be classified with a high purity of 0.88. Including redshift information, the purity and completeness again show positive trends for each class, particularly for stripped-envelope SNe, whose purity increases from 0.14 to 0.22. The F1 score, again significantly improves from 0.49 to 0.54.

In the binary classification task (core-collapse vs. Type Ia), we find average class purity of 0.64 and completeness of 0.67. This final classifier is most readily compared to literature studies. The two-class average accuracy of our classifier (0.66) is very similar to that of GHOST (Gagliano et al. 2021), which reported a class-average accuracy of 0.68. Including redshift information, we increase this accuracy to 0.71, showing state-of-the-art performance in the binary classification task.

The key takeaway from the galaxy-based classifier is twofold. First, we are able to successfully isolate a pure ($\simeq 0.9$) sample of Type Ia SNe with host galaxy information alone at a reasonably high completeness ($\simeq 0.6$) even *without* any redshift information; however, we are unable to isolate other sub-types. Second, we achieve two-class classification performance similar to other methods presented in the literature with relatively simple observational features from the host galaxies. In the absence of any light curve information, redshift does improve the performance of our classifier for all classes, suggesting that photo-z information will be valuable for rapid classification.

4.2. *Classification Performance Using Light Curve and Galaxy Features without Redshift Information*

We next turn our attention to a combined feature set that includes both light curve and host galaxy features. In this Section, we focus on the case where redshift information is not known. Although photometric redshift estimates are often available in current surveys and will be available in the era of LSST, the fraction of catastrophic outlier redshift estimates will be ~ 0.1 in the first half of LSST (Graham et al. 2018). Transients



(a) Galaxy Features, Excluding Redshift

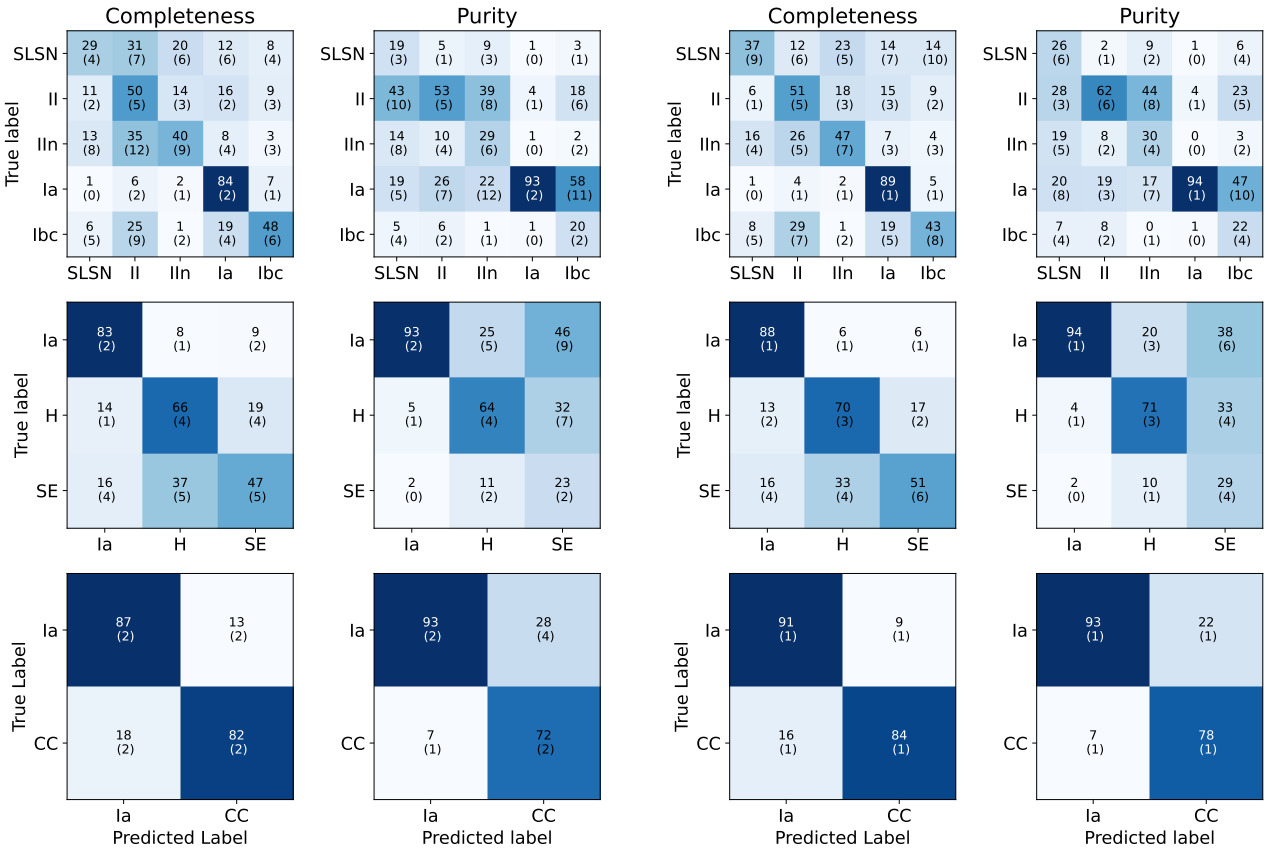
(b) Galaxy Features, Including Redshift

Figure 3: Confusion matrices for the classifier using *only* host galaxy with (*left*) and without (*right*) redshift information. Numbers are overall percent, while uncertainties are given in parentheses (e.g., the completeness of SLSNe in the 5-way classifier is 0.48 ± 0.06). The classifier reaches state-of-the-art results for Type Ia vs. CC classification, but fails to achieve high accuracy in the 5-class split.

which are primarily found in intrinsically low-luminosity galaxies (e.g., SLSNe) will have particularly unreliable redshift information. As a baseline comparison, we will contrast our results with both the galaxy-only classifier and a classifier that relies solely on light curve features. Our results can be primarily visualized by the confusion matrices in Figure 4.

In the five-way classification task, our "baseline" classifier with only light curve information achieves an average F1-score of 0.46. This is only marginally improved with the inclusion of galaxy information, increasing to an average F1 score of 0.49. Although the purity and completeness of all classes improve with the addition of galaxy information, although most are not statistically significant ($> 1\sigma$). Notably, the purity (0.20 to 0.25), completeness (0.27 to 0.33) and F1-score (0.23 to 0.29) substantial improve, with the F1 score having a statistically significant increase. This is not surprising, as SLSNe are known to prefer unusually low-mass galaxy; the contextual information of their hosts is therefore likely a useful feature. However, our galaxy-only classifier with redshift information performs substantially better for SLSNe (with an average F1-score of 0.63). For all other SN classes, light curve features (without redshift information) improve performance when compared to a classifier using only galaxy and redshift information. Imposing a higher confidence threshold ($p > 0.8$) notably improves the purity of Type II SNe (from 0.62 to 0.70), Type IIn SNe (from 0.29 to 0.52), and Type Ibc SNe (from 0.22 to 0.50) when including galaxy information. These improvements are *not* seen for Type IIn and Type Ibc SNe when excluding galaxy information.

In both the three-way and binary classification task, we see even greater performance improvement. In the three-way task, including galaxy information improves our average F1 score to 0.65, compared to the light curve-only baseline of 0.60; this is a $> 3\sigma$ performance improvement. We again see improvement in purity, completeness and F1-scores for all classes. This improvement is substantial for both Type II and stripped-



(a) Light Curve Features, Excluding Redshift

(b) Light Curve and Galaxy Features, Excluding Redshift

Figure 4: Confusion matrices for the classifier using solely light curve features (*left*) and a combination of galaxy plus light curve features (*right*), both excluding redshift information. Inclusion of galaxy features does not improve classification for any one class to a statistically significant degree, and the overall performance only moderately increases.

envelope SNe, with the F1-score improving from 0.66 to 0.71 and from 0.30 to 0.38, respectively. For the binary classification, the average F1 score improves from 0.84 to 0.86 ($> 1\sigma$), with both Type Ia purity and CC completeness most notably improving (both by $> 1\sigma$).

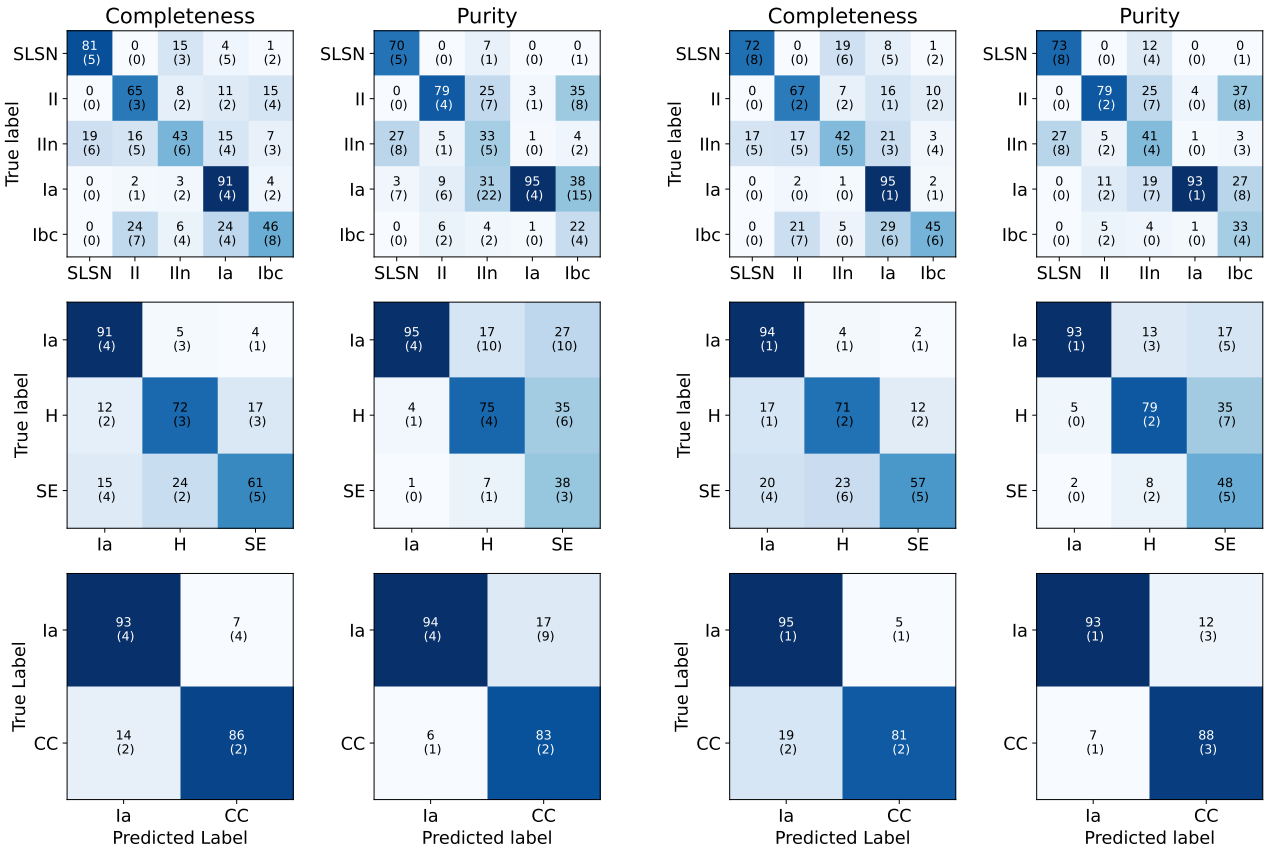
5. CLASSIFICATION PERFORMANCE USING LIGHT CURVE AND GALAXY FEATURES WITH REDSHIFT INFORMATION

Finally, we explore how galaxy information impacts classification performance when redshift is known. Across all three tasks (five-, three- and binary classifications), there is no statistical difference of average F1 scores between classifier with and without galaxy information. The only mildly significant difference ($> 1\sigma$) is an improvement on the purity of Type Ibc SNe (from 0.24 to 0.32) when including galaxy information, although this comes with a decrease in completeness (from 0.47 to 0.42, within 1σ uncertainties). As expected, a high-confidence cutoff ($p > 0.8$), improves accuracy, completeness and purity across all classification tasks with or without galaxy information***. We note that including galaxy information leads to higher Ia completeness (0.95 vs 0.98; a $> 2\sigma$ result). In short, it is clear that contextual host galaxy information does *not* significantly improve classification performance when redshift is known.

6. CLASSIFICATIONS OF THE PS1 SAMPLE

Finally, we use our redshift-independent, five-class classifier which uses both host and light curve information to classify the full set of **5243** SN-like transients from the PS1-MDS originally presented in Villar et al. (2020) and Hosseinzadeh et al. (2020). The expected purity and completeness for each class is, as previously explored, stated in Table 1. The full classifications are provided in Table **??**.

The breakdown of our classifications is shown in Figure 8. Compared to the spectroscopic dataset, we find an underrepresentation of the two majority classes



(a) Light Curve Features, Including Redshift

(b) Light Curve and Galaxy Features, Including Redshift

Figure 5: ignore.

Table 2: Expected and actual agreements between classifiers

Class	SuperRAENN (expected)	SuperRAENN (actual)	SuperPhot (expected)	SuperPhot (actual)
SLSNe	0.14	0.06	0.16	0.04
SN II	0.48	0.43	0.52	0.41
SN IIn	0.15	0.08	0.13	0.09
SN Ia	0.85	0.88	0.78	0.85
SN Ibc	0.07	0.24	0.07	0.19

(Type Ia and Type II SNe) in our photometric classifications, and an overreperesentation of the minority classes (SLSNe and Type Ibc). Hosseinzadeh et al. (2020) showed that the *expected* number of misclassifications could be used to correct the expected purity of each class (by dotting the purity matrix with the final classifications). When we apply this correction to our final class breakdown, we do recover a 5-class breakdown similar to the spectroscopic sample, lending credence to the idea that we correctly capture the biases of our imperfect classifier, even when applied to a new test set.

We directly compare our results to the SNe photometrically classified in Villar et al. (2020) and Hosseinzadeh et al. (2020). In total, we find 67% agreement between the original classifications (which utilized redshift information) and the new classifications for `SuperRAENN`, and 59% agreement between our new classifier and `SuperPhot`. Hosseinzadeh et al. (2020) showed that the *agreement* (A) between two classifiers, assuming independent biases, can be calculated as:

$$A = P^T C \tag{10}$$

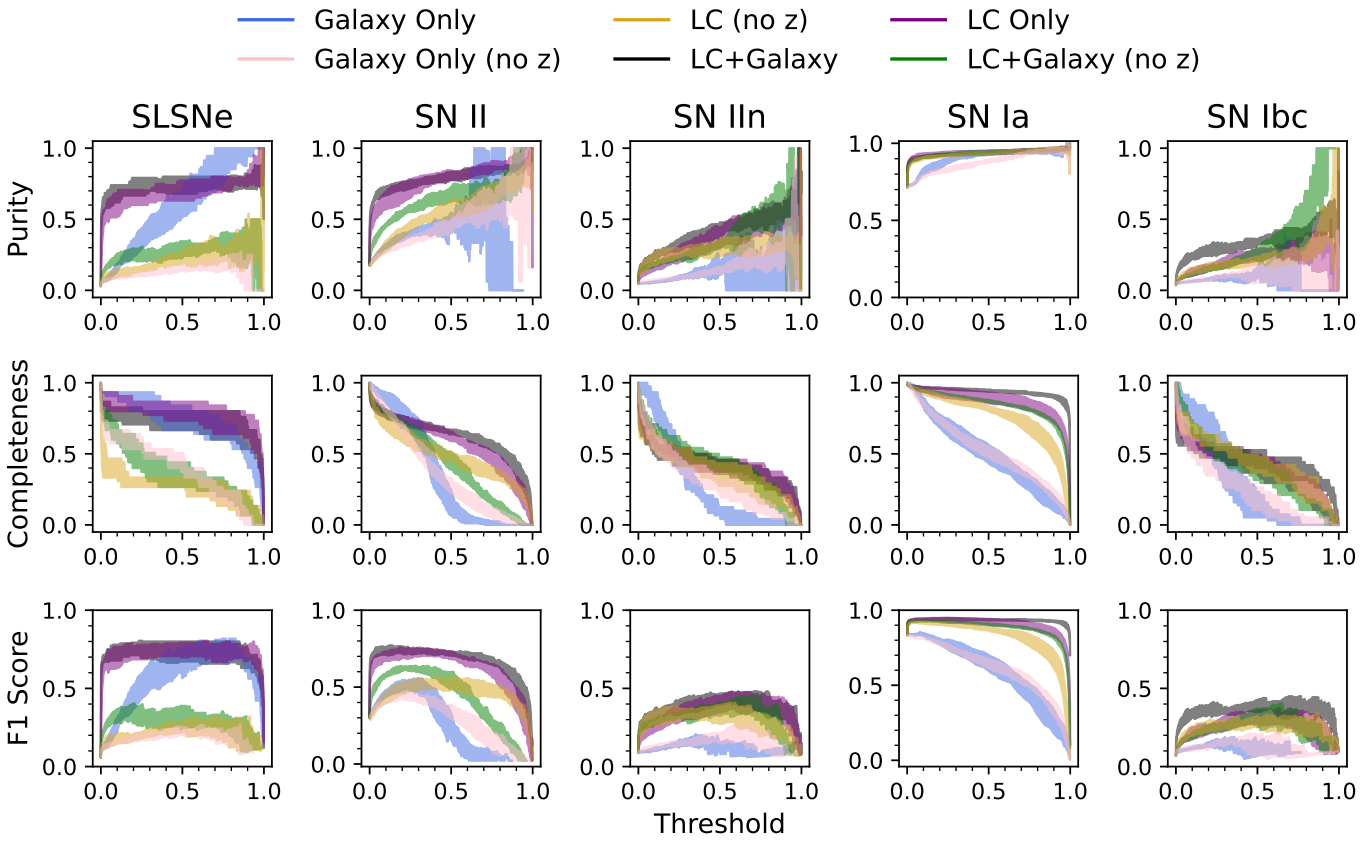


Figure 6: Purity (top row) and completeness (middle row) achieved with the galaxy feature set (blue), redshift-dependent light curve feature set (purple) and the combined feature set (black), as a function of SN type (columns). Short explanation.

where P is the purity matrix of the new classifier and C is the completeness matrix of the old classifiers. Using this, the expected agreements for each class and either classifier are shown in Table 2. "Agreement" here is therefore interpreted as, for a given SN class, the fraction of events which have been identified as the same class by both classifiers divided by the size of the class in the older classifiers. We will note that the definition of "agreement" listed above assumes that the two classifiers are independent, which may not necessarily hold true in our case, as both `SuperRAENN` and our new classifier use the same feature set. This may lead to more agreement than expected between the classifiers.

The agreement between our classifiers, for each class, ranges from $\sim 5-90\%$. Our classifier shows roughly *expected* agreement for both Type Ia and Type II classes. Type Ia SNe, in particular, have strong agreement between the three classifiers ($80-85\%$), giving high confidence in the purity of our Type Ia sample. Type Ibc SNe have notably higher agreement (~ 0.20) than expected; interestingly, this is similar to what was found in Hosseinzadeh et al. (2020). Type IIn and SLSNe, on the other hand, have much *smaller* agreement than expected, by a factor of $2-4$.

For SLSNe, most objects classified by `SuperRAENN` or `SuperPhot` as SLSNe were classified as Type Ia SNe by the new classifier. Interestingly, of the $\sim$ 150 SLSNe identified by the three classifiers, only one object is identified as a SLSN by all three – PSc010186. Examination of the light curve and a cross-match with known active galactic nuclei (Hsu et al. 2022) suggest that object is indeed a $z \sim 1$ active galactic nucleus undergoing regular, long-term variation. Additionally two objects are identified as SLSNe by our new classifier and *one* of the two original classifiers. PSc000553 was classified as a SLSN by both `SuperRAENN` and our new classifier and as a Type IIn SN by `SuperPhot`. Similarly, PSc151382 was classified as a SLSN by both `SuperPhot` and the

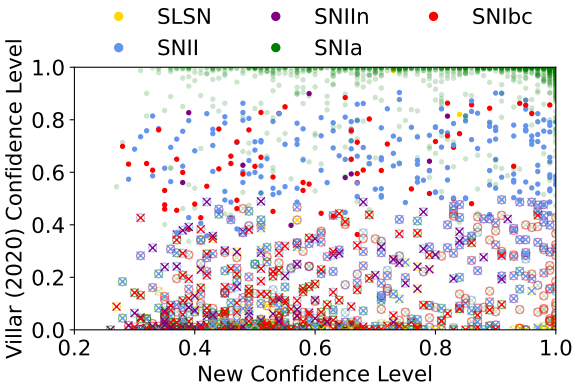


Figure 7: Comparison of SN classifications from Villar et al. (2020) and this work. On the abscissa, we plot the confidence level of the Villar et al. (2018) classifier, while the ordinate shows the confidence of the new classifier for the *same class*. Objects plotted as a point have the same classification with both methods; objects plotted as an 'x' have different classifications with either method. The color of the central point ('x' or 'o') identifies the label from this paper. In the case of mismatches (the 'x' objects), the colored circle represents the label from Villar et al. (2020).

new classified and as a Type IIn by `SuperRAENN`. We take these results as an important warning for the community: without redshift information, our photometric classifiers fail to seemingly produce a reliably pure sample of SLSNe. Given the strong preference for SLSNe to reside in low-luminosity galaxies, which may not even be detected in single LSST exposures, this may severely challenge our ability to conduct large scale population studies on this class of objects, even though tens of thousands are expected to be observable by LSST annually (Villar et al. 2018).

Finally, we visualize our agreement in Figure 7. We find that, in general, Type Ia SNe tend to dominate the sample of objects for which both our new classifier and Villar et al. (2020) confidently agree. Many of the disagreements seem to arise from our new classifier labelling objects as Type Ibc (represented as a red 'x' in the plot) and Type IIn (purple 'x') for a wide range of classes identified in Villar et al. (2018).

A key takeaway of this analysis is that expected classification accuracies and "agreement" between photometric classifiers are highly variable, and any population-level studies of photometrically classified SNe should take care to understand underlying biases and misclassifications in the observed SNe population.

7. CONCLUSIONS

We have presented an analysis on the impact of galaxy properties on the photometric classification of SNe, focusing on CCSN sub-types. We present our key findings and products below:

1. In corroboration with many previous results, we find that the observable properties of host galaxies visibly correlate with SN type. We are able to positively distinguish CC SNe from Type Ia SNe with $\simeq 70\%$ accuracy, matching similar studies in the literature.

2. We present a new objective function which better accounts for the hierarchical nature of our SN classification.

3. We find that contextual information – i.e., the host galaxy observable properties – does not necessarily aid in SN classification when given sufficient broadband light curve observations and redshift information. Host galaxy information is most helpful in improving the classification accuracy of Type Ia and Type II SNe, but does not necessarily increase the accuracy of classification for other subtypes.

4. Given no redshift information, host galaxy features marginally improve the classification accuracy of SNe, in general. However, galaxy information can greatly increase the completeness of Type Ia samples, especially given no redshift information.

5. Finally, we present an updated classification set for the PS1-MDS set of SN-like transients without redshift information. We explore the agreement between these classifications and those originally presented, finding strong agreement in some classes (Type Ia, Type Ibc) and weaker than expected agreement in others (most notably, SLSNe).

1 VAV acknowledges support by the NSF through grant
2 AST-2108676.

Software: FLEET (**?**)

REFERENCES

Anderson, J. P., Habergham, S., James, P., & Hamuy, M. 2012, Monthly Notices of the Royal Astronomical Society, 424, 1372

Baldeschi, A., Miller, A., Stroh, M., Margutti, R., & Coppejans, D. 2020, The Astrophysical Journal, 902, 60

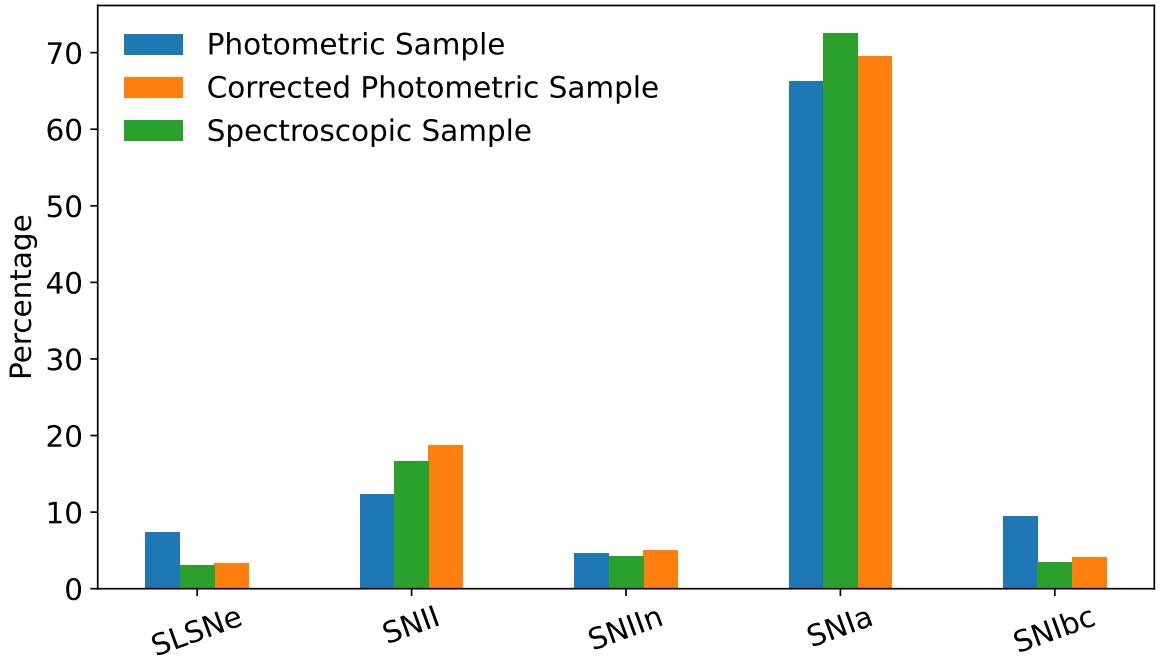


Figure 8: Breakdown of SN subclasses in our new photometrically classified set of SNe without redshift information. Compared to the spectroscopic sample (green), our classifier (blue) predicts a smaller fraction of Type Ia and Type II SNe and a larger fraction of SLSNe and Type Ibc SNe. However, using our purity matrix (see text), we can correct our classification breakdowns accounting for expected mis-identifications. We see that, following this correction, our classes breakdowns are largely consistent with the spectroscopic sample. This indicates that although our classifier is biased, it is biased in a way which can be understand via our training set and can likely be carefully corrected for population-level studies.

Berger, E. 2010, The Astrophysical Journal, 722, 1946

Bertinetto, L., Mueller, R., Tertikas, K., Samangooei, S., & Lord, N. A. 2020, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 12506–12515

Bloom, J. S., Kulkarni, S. R., & Djorgovski, S. G. 2002, The Astronomical Journal, 123, 1111

Boone, K. 2021, The Astronomical Journal, 162, 275

Bruch, R. J., Gal-Yam, A., Schulze, S., et al. 2021, The Astrophysical Journal, 912, 46

Carrasco-Davis, R., Reyes, E., Valenzuela, C., et al. 2021, The Astronomical Journal, 162, 231

Childress, M., Aldering, G., Antilogus, P., et al. 2013, The Astrophysical Journal, 770, 107

Dong, Y., Milisavljevic, D., Leja, J., et al. 2022, The Astrophysical Journal, 927, 199

Foley, R. J., & Mandel, K. 2013, The Astrophysical Journal, 778, 167

Gagliano, A., Contardo, G., Mackey, D. F., Malz, A. I., & Aleo, P. D. 2023, arXiv preprint arXiv:2305.08894

Gagliano, A., Narayan, G., Engel, A., et al. 2021, The Astrophysical Journal, 908, 170

Gal-Yam, A. 2016, arXiv preprint arXiv:1611.09353

Gomez, S., Berger, E., Blanchard, P. K., et al. 2020, The Astrophysical Journal, 904, 74

Graham, M. L., Connolly, A. J., Ivezić, Ž., et al. 2018, AJ, 155, 1, doi: 10.3847/1538-3881/aa99d4

Habergham, S., James, P., & Anderson, J. 2012, Monthly Notices of the Royal Astronomical Society, 424, 2841

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, The Astrophysical Journal Supplement Series, 208, 19

Hosseinzadeh, G., Dauphin, F., Villar, V. A., et al. 2020, The Astrophysical Journal, 905, 93

Hsu, B., Hosseinzadeh, G., Villar, V. A., & Berger, E. 2022, arXiv preprint arXiv:2204.09809

Hudelot, P., Cuillandre, J.-C., Withington, K., et al. 2012, VizieR Online Data Catalog, II

Kelly, P. L., & Kirshner, R. P. 2012, The Astrophysical Journal, 759, 107

Kessler, R., Narayan, G., Avelino, A., et al. 2019, Publications of the Astronomical Society of the Pacific, 131, 094501

Khazov, D., Yaron, O., Gal-Yam, A., et al. 2016, The Astrophysical Journal, 818, 3

Kingma, D. P., & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980, doi: 10.48550/arXiv.1412.6980

Kisley, M., Qin, Y.-J., Zabludoff, A., Barnard, K., & Ko, C.-L. 2022, arXiv e-prints, arXiv:2209.02784. https://arxiv.org/abs/2209.02784

Möller, A., & de Boissière, T. 2020, Monthly Notices of the Royal Astronomical Society, 491, 4277

Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, Publications of the Astronomical Society of the Pacific, 131, 118002

Qin, Y.-J., Zabludoff, A., Kisley, M., et al. 2022, The Astrophysical Journal Supplement Series, 259, 13

Qu, H., & Sako, M. 2022, The Astronomical Journal, 163, 57

Ransome, C. L., Habergham-Mawson, S., Darnley, M. J., James, P. A., & Percival, S. M. 2022, Monthly Notices of the Royal Astronomical Society, 513, 3564

Sánchez-Sáez, P., Reyes, I., Valenzuela, C., et al. 2021, The Astronomical Journal, 161, 141

Schulze, S., Yaron, O., Sollerman, J., et al. 2021, The Astrophysical Journal Supplement Series, 255, 29

Villar, V. A., de Soto, K., & Gagliano, A. 2023, arXiv preprint arXiv:2312.02266

Villar, V. A., Nicholl, M., & Berger, E. 2018, The Astrophysical Journal, 869, 166

Villar, V. A., Hosseinzadeh, G., Berger, E., et al. 2020, The Astrophysical Journal, 905, 94