Prepared for submission to JCAP

arXiv:1608.05423v2 [astro-ph.IM] 24 Nov 2016

Photometric classification of type Ia supernovae in the SuperNova Legacy Survey with supervised learning

A. Möller,[a,b,c,1] **V. Ruhlmann-Kleider,**[c] **C. Leloup,**[c] **J. Neveu,**[c,d] **N. Palanque-Delabrouille,**[c] **J. Rich,**[c] **R. Carlberg,**[e] **C. Lidman,**[f,b] **C. Pritchet.**[g]

[a]Research School of Astronomy and Astrophysics,
Australian National University, Canberra, ACT 2611, Australia.

[b]ARC Centre of Excellence for All-sky Astrophysics (CAASTRO), Australia.

[c]Irfu, SPP, CEA Saclay,
F-91191 Gif sur Yvette Cedex, France.

[d]Université Paris-Sud, LAL UMR 8607, F-91898 Orsay Cedex, France & CNRS/IN2P3,
F-91405 Orsay, France.

[e]Department of Astronomy and Astrophysics, University of Toronto,
50 St. George Street, Toronto, ON M5S 3H8, Canada.

[f]Australian Astronomical Observatory,
North Ryde, NSW 2113, Australia.

[g]Department of Physics and Astronomy, University of Victoria,
P.O. Box 3055, Victoria, BC V8W 3P6, Canada.

[1]Corresponding author.

E-mail: anais.moller@anu.edu.au, vanina.ruhlmann-kleider@cea.fr, clement.leloup@cea.fr, jneveu@lal.in2p3.fr, nathalie.palanque-delabrouille@cea.fr, james.rich@cea.fr, raymond.carlberg@utoronto.ca, chris.lidman@aao.gov.au, pritchet@uvic.ca

Abstract. In the era of large astronomical surveys, photometric classification of supernovae (SNe) has become an important research field due to limited spectroscopic resources for candidate follow-up and classification. In this work, we present a method to photometrically classify type Ia supernovae based on machine learning with redshifts that are derived from the SN light-curves. This method is implemented on real data from the SNLS deferred pipeline, a purely photometric pipeline that identifies SNe Ia at high-redshifts ($0.2 < z < 1.1$).

Our method consists of two stages: feature extraction (obtaining the SN redshift from photometry and estimating light-curve shape parameters) and machine learning classification. We study the performance of different algorithms such as Random Forest and Boosted Decision Trees. We evaluate the performance using SN simulations and real data from the first 3 years of the Supernova Legacy Survey (SNLS), which contains large spectroscopically and photometrically classified type Ia samples. Using the Area Under the Curve (AUC) metric, where perfect classification is given by 1, we find that our best-performing classifier (Extreme Gradient Boosting Decision Tree) has an AUC of 0.98.

We show that it is possible to obtain a large photometrically selected type Ia SN sample with an estimated contamination of less than 5%. When applied to data from the first three years of SNLS, we obtain 529 events. We investigate the differences between classifying simulated SNe, and real SN survey data. In particular, we find that applying a thorough set of selection cuts to the SN sample is essential for good classification. This work demonstrates for the first time the feasibility of machine learning classification in a high-z SN survey with application to real SN data.

Contents

1 Introduction **2**

2 SNLS data and SN simulation **2**
- 2.1 SN simulation 3
- 2.2 SNLS3 data 3
  - 2.2.1 Type Ia SNe 4
  - 2.2.2 Core-collapse SNe 4

3 Light-curve analysis before classification **4**
- 3.1 Photometric SN redshifts 5
- 3.2 Selection cuts 7
  - 3.2.1 Photometric SN redshift quality 7
  - 3.2.2 Light-curve quality 7
- 3.3 Light-curve fitter 7

4 Classification with Machine Learning **8**
- 4.1 Boosted Decision Trees (BDTs) 9
- 4.2 Cross-validation 10
- 4.3 Evaluating the classifier 10
  - 4.3.1 ROC curve 10
  - 4.3.2 Purity and efficiency 10
- 4.4 Parameter setting and feature selection 11

5 Comparison of different classification methods **12**
- 5.1 Application on simulated SNe 12
  - 5.1.1 Efficiency evolution with redshift 13
  - 5.1.2 Evolution of purity with redshift 13
- 5.2 Application on SNLS3 data 16
  - 5.2.1 Comparison with the SNLS3 subluminous and peculiar SNIa samples 16
  - 5.2.2 Effect of spectroscopic confidence index 19

6 Choosing a method: XGB with high purity 98% **19**
- 6.1 Effect of selection cuts and classification 19
- 6.2 Classification and photometric SN redshifts 19
- 6.3 Efficiency evolution: classification, selection and total 20
- 6.4 The effect of the spectroscopic confidence index 21
- 6.5 The effect of light-curve quality 21

7 Conclusions **21**

1 Introduction

Type Ia supernovae are used as standard candles to measure the expansion history of the Universe. Since the discovery of the accelerated expansion of the Universe [1, 2], enormous effort has been dedicated to obtaining larger samples of SNe Ia with high quality light-curves. Second-generation surveys such as SNLS and SDSS-II [3] have obtained such samples using spectroscopic classification of the SN type. However, in the era of large surveys such as DES [4] and LSST [5], spectroscopic resources are insufficient for complete candidate follow-up and classification, making purely photometric classifications necessary.

The SNLS deferred photometric pipeline can be separated in two parts: i) the detection of SN-like events, and ii) their classification as type Ia SNe [6] or core-collapse SNe [7]. In the analysis of the 3-year SNLS data (SNLS3), this pipeline provided a sample of 486 photometrically identified SNe Ia, almost twice the number of spectroscopically identified type Ia SNe found by the SNLS real-time analysis pipeline. Classification in the deferred pipeline included the use of host-galaxy photometric redshifts. Redshifts were assigned by matching SNe to host-galaxies with photometric redshifts in the *Ilbert* catalog [8]. This assignment had an efficiency of 83%. Events without a photometric redshift could not be classified in that analysis and were subsequently removed.

This paper presents a new photometric classification of SNLS SNe based on supervised learning, with redshifts derived directly from SN light-curves. The redshift algorithm was trained on SNLS3 data and has better average precision and fewer catastrophic errors than the host galaxy photometric redshift catalog used in the previous analysis. It provides redshifts for all SNe and is independent of cosmological parameters. The redshift algorithm is described in more detail in [9].

In order to exploit all the available information and to optimize classification, we take advantage of machine learning algorithms, notably Boosted Decision Trees (BDTs). BDTs are supervised learning methods where the algorithm learns from a known "training sample" before classifying unidentified data. Supervised learning methods have been previously used for photometric SN classification, e.g. the SuperNova Photometric Classification Challenge [10], with good results [11–14]. In this work, the performance of the supervised learning classification is estimated not only with simulated SNe but also with the large type Ia photometric sample obtained in the previous SNLS3 analysis [6]. This work therefore sets a precedent for the application of machine learning methods to the classification of real SN data.

The outline of the paper is as follows: the SNLS data and simulations are presented in Section 2. In Section 3, we describe the measurement of light-curve parameters as well as the selection of light-curves of sufficient quality to be treated for classification. In Section 4, we introduce the classification algorithms used in this work and the metrics used for evaluation. We compare different classification algorithms in Section 5 using simulated, photometric and spectroscopic SN samples. Our best performing method is studied in detail in Section 6. We summarize and conclude this work in Section 7.

2 SNLS data and SN simulation

SNLS is part of the Deep Synoptic Survey conducted at the Canada-France-Hawaii Telescope (CFHT). Using a rolling-search strategy, it targeted four one square degree fields during 5 to 7 consecutive lunations per year over a period spanning five years using four different

broadband filters g_M, r_M, i_M and z_M [15]. Two independent analysis pipelines processed SNLS data. The first, which we refer to as the real-time pipeline, relies on the spectroscopic follow-up of detected SN candidates for classification and redshift determination [16]. The second is the deferred photometric pipeline.

This work is based on the deferred photometric pipeline. It is independent of the real-time analysis and requires only photometric information. In this pipeline, transient events are detected in one filter and multi-band light-curves are processed for all detections. Then, a set of cuts described in [6] is applied to reject spurious objects and obtain a sample of events whose light-curves are consistent in shape with that expected from SNe, hereafter referred as SN-like events. Classification of these SN-like events is the subject of this work. We will classify SNe into two types: type Ia SNe with correct redshifts ("the signal") and "the background" which consists of other types of SNe or type Ia SNe with inaccurate redshifts. In this work, redshifts are considered accurate if, when compared to the generated (spectroscopic) redshift for simulations (data), they satisfy $|\Delta z|/(1+z) < 0.1$, inaccurate redshifts exceed 0.15. More details on the deferred photometric pipeline can be found in [6].

2.1 SN simulation

To set up the classification procedure, we use synthetic type Ia and core-collapse SNe that were generated for the SNLS3 analysis in [6]. This allows for a more direct comparison between this work and the method used in [6].

Synthetic light-curves of SNe Ia within the redshift range $(0, 1.2]$ were produced in [6] with SALT2 [17] assuming a flat Universe with $\Omega_m = 0.23$. Simulated SNe Ia were generated assuming a constant co-moving volumetric rate. Values of the SALT2 color and X_1 parameters (X_1 is related to the light-curve width) were randomly selected from Gaussian distributions that match the distributions of the spectroscopically identified SNe Ia. To represent the SNIa population before selection effects, only SNe Ia with $z < 0.7$ were used to parametrize the distributions [18]. For each synthetic SN, a random position in the SNLS fields was assigned, Milky Way dust extinction corrections were applied and detection effects were simulated. From the original list of $20,000$ simulated SNe Ia, 49% were detected and passed the SN-like selection cuts. The majority of the lost supernovae had low signal-to-noise ratios.

To increase the number of events, we chose not to impose the detection criteria in [6]. This allows a larger number of faint events to be used in the classification. We keep selection cuts for consistency. In total $10,522$ SNIa synthetic light-curves are used in this work.

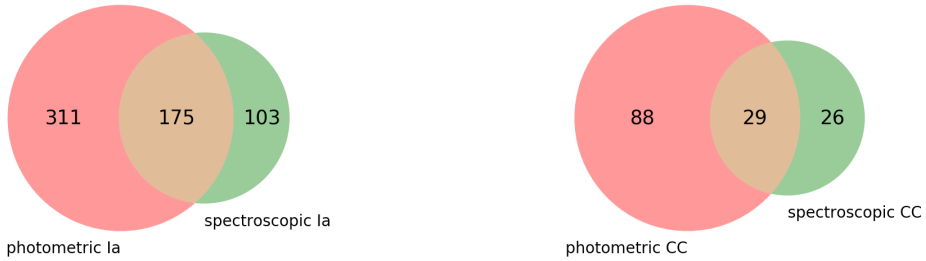
Core-collapse supernovae (CC SNe) can be separated into those that have a plateau in the light-curve after maximum light, and those that do not. Synthetic light-curves were generated for both CC types in the redshift range $0 < z < 1.2$ assuming a constant co-moving volumetric rate. A light-curve model [6] was constructed based on the sample of 117 SNLS CC SNe at $z < 0.4$ [7]. From 40,000 synthetic light curves, 20,000 of each type, only 10.5% were detected and selected as SN-like. The efficiency is very low because of the low luminosity of CC SN compared to SNIa. As for synthetic type Ia SNe, for our classification we chose not to impose the detection criteria used in [6] and obtain a sample of 5000 simulated core-collapse SNe.

2.2 SNLS3 data

To evaluate the performance of our classification, we use not only simulated events but also SNe that have been identified from the first 3 years of SNLS (hereafter referred to as SNLS3).

These SNe were identified photometrically or spectroscopically by independent analyses. The photometric and spectroscopic samples contain a number of events in common as illustrated in Figure 1.

As previously mentioned, in the SNLS deferred photometric pipeline, classification is performed on events that are first detected as transients and then selected as SN-like. This SN-like sample is the starting point of our classification and consists of 1483 events.



(a) SNe Ia: 175 common events.

(b) CC SNe: 29 common events.

Figure 1: Venn diagrams of spectroscopically and photometrically identified type Ia (1a) and CC (1b) from the SNLS3 classification. Supernovae common to both samples are shown in beige.

2.2.1 Type Ia SNe

The spectroscopically and photometrically identified type Ia samples are shown in Table 1. Events common to both samples are illustrated in Figure 1a. In SNLS3, 486 events were photometrically classified as SNe Ia with an estimated purity of $94.4 \pm 0.5\%$ [6].

In this work, we consider two sub-classes of spectroscopic SNe Ia, split according to the confidence index (CI) of the spectroscopic identification: certain SNIa are denoted "SNIa" (corresponding to $CI = 4$ and $CI = 5$ in the classification scheme of [19]) and probable SNIa ($CI = 3$) are labeled "SNIa*". Further details about these indices can be found in [19].

A photometrically classified sample of 18 subluminous SNe Ia at $z < 0.6$ was obtained in [20]. The SN-like sample contains 16 of these 18 subluminous SNe Ia. SNLS also detected 8 SNe Ia that were spectroscopically classified as peculiar. This sample includes super-Chandrasekhar and 1991T-like SN events [21–24].

2.2.2 Core-collapse SNe

In the SNLS3 SN-like sample, 55 events were identified spectroscopically as core-collapse SNe [6]. A photometric classification based on the deferred pipeline identified 117 events as core-collapse SNe with an estimated purity of 97% [7]. Common events between both samples are illustrated in Figure 1b.

3 Light-curve analysis before classification

Before classification, all SN-like events are processed. First, photometric SN redshifts are obtained by using the algorithm described in Section 3.1. Then, selection cuts are applied on

SNLS3	SNe Ia		
	spectroscopic		photometric
SN-like	278		486
CI	SNIa 213	SNIa* 65	

Table 1: Type Ia SNe in SNLS3. The number of events identified spectroscopically (photometrically) in the SN-like sample are shown. For spectroscopic SNe Ia, we further divide the sample using the confidence indices (CI) as defined in [19].

the photometric SN redshift quality and the light-curve quality as determined by a SALT2 fit. Finally, events are fitted with a general light-curve fitter. Only the photometric SN redshift and the general light-curve fitter parameters are used in the classification. The complete procedure is illustrated in Figure 2. In the following we introduce the algorithms and selection cuts used in our analysis.

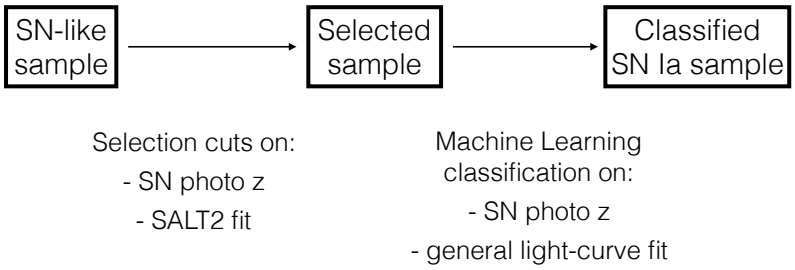


Figure 2: Flow chart showing the selection and classification steps used in this work. The SN samples are shown as solid lines. *SNphotoz* stands for photometric SN redshifts.

3.1 Photometric SN redshifts

We use the algorithm elaborated in *Palanque-Delabrouille et al.* [9], trained with SNLS3 data, to obtain redshifts for each SN-like event. These redshifts will be hereafter called photometric SN redshifts. The algorithm obtains the redshift of type Ia SNe using the SALT2 light-curve fitter in a iterative procedure. In the first iteration, successive values of redshift are fit while color and stretch SALT2 parameters are constrained by priors. When a solution if found, another scan is done around the fitted redshift with free color and stretch parameters.

The precision of the redshifts for SNe in the SNLS3 sample was reported in [9]. The average precision was defined as $\sigma_{\Delta z/(1+z)} \equiv 1.48\times$ median $[|\Delta z|/(1+z)]$, where Δz is the difference between the real and the photometric SN redshift. The rate of catastrophic errors, η, was defined as the proportion of events with $|\Delta z|/(1+z) > 0.15$. For the SNLS3 sample an average precision of $\sigma_{\Delta z/1+z} = 0.022$ up to $z \sim 1$ was found, while for $z < 0.45$ it was 0.006. The precision degrades with redshift due to low flux, first in the g band and then in the r band as redshift increases. This degradation is irregular as seen in Fig. 4 and described in Section 5 in [9]. Catastrophic errors were found to be under 1.4% for type Ia SNe passing color and stretch cuts. When restricting the test sample to spectroscopically confirmed SNe Ia the catastrophic errors fell to 0.4%. The authors found a net bias on the fitted redshift of 0.008 on average. Further details can be found in [9].

The algorithm can be used for obtaining redshifts for all SN-like events. In Figure 3, we plot the light-curve redshift against the real redshift for simulated type Ia and core collapse SNe. Since the algorithm assumes that all events are SNIa, redshifts obtained for core-collapse SNe are usually inaccurate. However, there are some core-collapse SNe that have redshifts that are close to correct. These events have colors that are consistent with type Ia SNe.

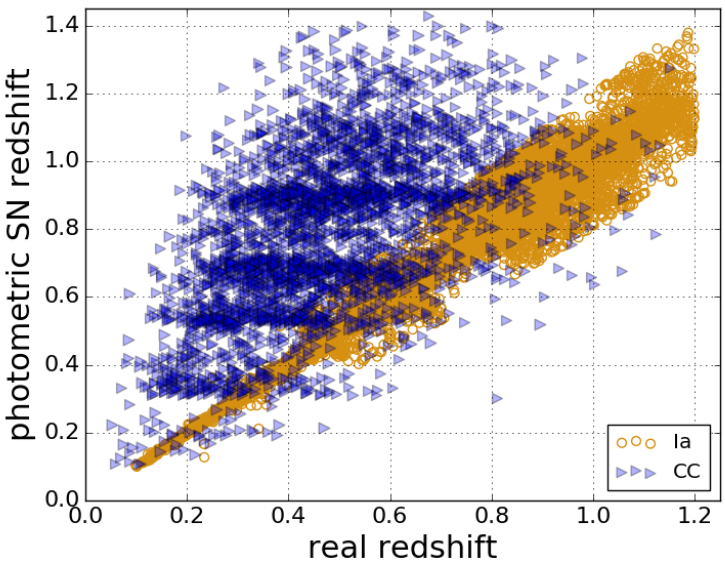


Figure 3: Photometric SN redshift vs. the real redshift for simulated SNe. We plot type Ia SNe with yellow circles and core-collapse SNe with blue triangles.

The photometric SN redshift algorithm provides not only the redshift, but also the reduced χ^2_ν of the agreement in color and width between the processed light-curve and the expected light-curve of an SNIa at the determined redshift. In the following we will refer to the reduced chi-square, which is defined as $\chi^2_\nu = \chi^2/N_{dof}$, there N_{dof} is the number of degrees of freedom.

3.2 Selection cuts

Before classification, we apply selection cuts to ensure meaningful photometric redshifts and reliable light-curves, and to mitigate the number of non-SN events still present in the SN-like sample. Indeed, since the identification of SN-like events in the photometric pipeline was designed to be sensitive to different types of SNe and to faint ones, non-SNe are probably still present and may bias the classification. We thus restrict the sample further by applying cuts more focused on type Ia events.

3.2.1 Photometric SN redshift quality

We assess the quality of photometric SN redshifts through the goodness of the light-curve fit. The photometric SN redshifts algorithm performs iterative fits and in each passing total χ^2 and contributions to the total χ^2 by the priors can be obtained. More details can be found in [9].

To investigate possible cuts, we visually inspected a subset of the light curves that are outliers in a number of diagnostic plots, such as the plot shown in Fig. 4. In particular, the χ^2_ν of the total multi-band fit was found to be sensitive to non SN-like events. We chose to exclude events with a total χ^2_ν greater than four.

Cuts were also derived for other output variables of the photometric SN redshift algorithm such as the χ^2_ν of the redshift, color and stretch.

3.2.2 Light-curve quality

With the redshift fixed to the photometric redshift we now assess the quality of our light-curves. For this, we fit our light-curves with SALT2. The output of this fit is used only to remove events with insufficient light-curve coverage and events with poor fits. It provides a fitted date of maximum light as well as color, stretch and peak magnitudes. It is not used in the classification since SALT2 was already used for obtaining photometric SN redshifts. We require:

- Minimal sampling of the light-curve before and after the SALT2 fitted time of maximum in rest frame, τ:
  - at least one measurement in the range $-10 < \tau < +5$ days,
  - at least one measurement in the range $+5 < \tau < +20$ days for a reasonable shape evaluation,
  - at least one measurement in each band from a pair selected from $(g-i)$, $(r-z)$ or $(i-z)$ must be within the range $-10 < \tau < +35$.
- SALT2 convergence: events for which the SALT2 fit did not converge are discarded.

3.3 Light-curve fitter

To parameterize the light-curve shape we use the functional form [6]:

$$f^k(t) = A^k \frac{\exp -(t-t_0^k)/\tau_{fall}^k}{1+\exp -(t-t_0^k)/\tau_{rise}^k} + c^k \ , \tag{3.1}$$

where A^k sets the normalization, τ_{fall}^k (τ_{rise}^k) defines the fall (rise) time, t_0^k is related to the date of maximum as $t_{max}^k = t_0^k + \tau_{rise}^k \ln(\tau_{fall}^k/\tau_{rise}^k - 1)$ and c^k is a constant.

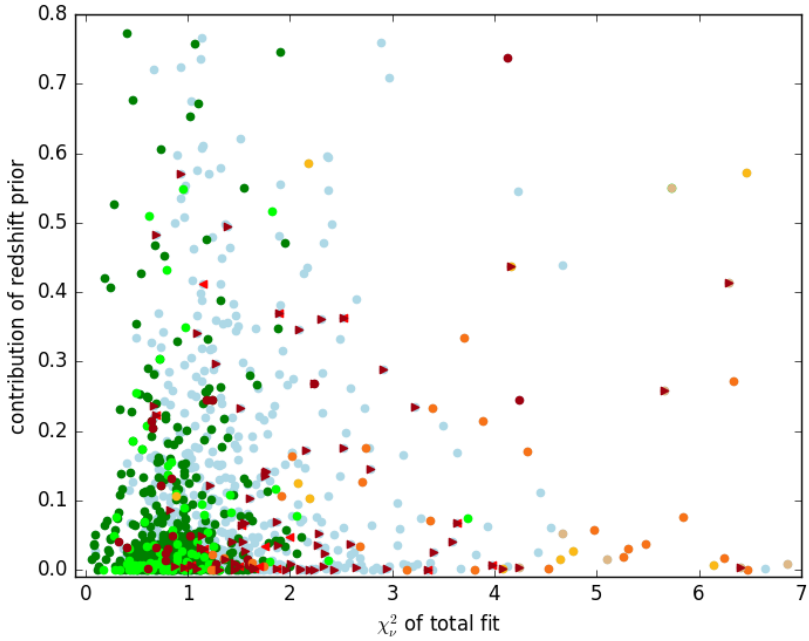


Figure 4: The contribution of the redshift prior to the total reduced χ^2_ν for SNLS3 SN-like events. Events classified as type Ia are in light (dark) green for spectroscopically confirmed (photometrically identified) supernovae. Core-collapse SNe are in bright red left-pointing (dark red right-pointing) triangles for spectroscopic (photometric) sample. Orange (yellow) dots stand for events whose light-curves were visually inspected and found not to be compatible with a SN-like signal (compatible with long declining events similar to SNe II). Beige dots are events with incompatible photometric and spectroscopic redshifts. All other SN-like events of unknown type are in blue. SNe with a total χ^2_ν that is greater than four are excluded from our selection.

First, the flux in each filter, k, is fitted. Then, we impose that all fits share the same t_0 from the i band [1]. A second fit is done using Gaussian priors from the first fit.

The fitting procedure provides the amplitude, rise and fall times for each filter. These are relevant features to characterize a SN light-curve and are the ones used for the following SN classification.

4 Classification with Machine Learning

Our goal is to select type Ia SNe from a SN sample. This can be reduced to a problem of predicting the type for each event. Machine learning algorithms provide an automated way of classifying events. In particular, supervised learning algorithms can learn from data in order to make predictions.

The features available for the classification are: the photometric SN redshift, the color and stretch obtained from the redshift fit and their χ^2_ν, the values for the general light-curve

[1] Some SNe have different maximum dates for different filters. This requirement was set for consistency between fits. It is not expected that this affects the classification.

fit in each band (amplitudes, rising and falling times) and the χ^2_ν of the general light-curve fits.

In this section, we will first introduce the machine learning algorithms used in this work.[2] Then, we will introduce our model validation technique, cross-validation. Performance is evaluated using metrics that will be described in Section 4.3.

4.1 Boosted Decision Trees (BDTs)

BDTs are supervised classification methods that perform well with large data sets and are adapted to classify high-dimensional data. From the *training sample* they learn a mapping function that allows them to classify other data points (*classification sample*). Their output is a *prediction*: the probability of an object to belong to a given class (BDT response).

A decision tree (DT) makes successive rectangular cuts in the parameter space to classify data. Binary splits separate the data into subsamples ("leaf nodes") which at the end of the tree are given a probability to be classified as signal or background (prediction). At each split, the algorithm determines the variable that gives the best separation to discriminate between signal and background (in terms of classification error). Often, trees are too complex and do not generalize to other samples (they "overfit"). To avoid this, trees can be combined to improve generalizability and stability. Two main approaches for combining trees are averaging and boosting methods.

Averaging methods construct several estimators (decision trees). The final prediction is an average of all the DTs prediction. Such methods are:

- Random Forest (RF): n decision trees are constructed from a sample drawn with replacement (an event can be drawn multiple times) from the training set. In the learning process, the feature to be used at each binary split is picked from a random subset of the features (which can include or not the best available feature). This is done to obtain a better model where variance is decreased. The final prediction contains an average of all probabilistic predictions in different trees.

- Bagging: decision trees are built from random subsets of the training set. For each event, the final prediction is the sum of the predictions from all trees.

Boosting methods build a model iteratively. They combine "weak" classifiers as small decision trees on modified versions of the training data. We will use:

- AdaBoost (AB): iteratively constructs an additive model for the data. At each iteration, the data set is classified and each individual event is given a weight which represents its importance in the classification. The weight of individual events on the training sample is modified at each iteration. Those events that were incorrectly predicted at the previous step have their weights increased and those that were correctly predicted have their weights decreased. In this sense, misclassified events are the focus of the next iteration. All predictions are combined by a weighted sum to produce the final prediction.

- XGBoost (XGB): constructs an additive model while optimizing a loss function. The loss function accounts for the inaccuracy of predictions in the classification. The performance is given by an objective function that contains both a loss function and a

[2] We rely on the Python package scikit-learn [25] for implementation.

regularization term (controls complexity of the model). This is a more refined version of GradientBoosting which is accurate and has shown good performance in classification challenges [26] [3].

Although we separate here the different methods for combing trees, from now on AdaBoost, XGBoost and Random Forest will be referred to as BDTs.

4.2 Cross-validation

Cross-validation is a technique that allows us to assess how a classification generalizes to an independent data set. The idea behind cross-validation is to partition the data into independent subsets, training with one set while evaluating with the other. This can be done several times (number of folds) which allows one to train and measure the success rate of the classifier with the different samples ensuring that one is using information that is available in the entire simulation. In this work, we choose to do a 3-fold cross-validation which is enough to avoid over-fitting and to have a robust assessment of our model validity while maintaining a large training sample.

4.3 Evaluating the classifier

The classification results in each SN being classified as a SNIa with the correct redshift or as a SNCC or SNIa with an inaccurate redshift. SNIa in the first group constitute the "signal" and those in the second group the "background".

The performance of a classifier can be evaluated using different metrics introduced in this section (e.g. AUC). Since our goal is to obtain a large and reliable SNIa sample, it is natural to use purity and efficiency as indicators. The latter will also allow us to set a criterion for choosing a probability threshold. For efficiency and purity studies we will use a subset of the simulation to train and another independent subset to classify and therefore estimate the performance of the classifier.

4.3.1 ROC curve

Our problem is a binary one: events are either signal or background and are classified in these categories. A metric that is commonly used as an evaluation method for dichotomic classifications, is the AUC metric. AUC stands for Area Under Curve, where the curve is the ROC curve (Receiver Operating Characteristic). The ROC curve illustrates the performance of a binary classifier by plotting the true positive rate (efficiency) against the false positive rate (contamination).

While the ROC curve represents the performance of a model in two-dimensions, the AUC simplifies this into a number. A perfect model would score an AUC of 1 while a random classification would score 0.5.

4.3.2 Purity and efficiency

For a classified sample, purity and efficiency can be used as metrics. This requires a choice of the BDT response threshold.

Our goal in the classification of SN-like events is to obtain a sample of type Ia SNe with correct redshifts separated from other types of SNe or type Ia SNe with inaccurate redshifts. Redshifts are considered accurate if, when compared to the generated redshift for

[3] http://xgboost.readthedocs.org/

simulations, they satisfy $|\Delta z|/(1+z) < 0.1$, inaccurate redshifts exceed 0.15. SNe Ia with inaccurate redshifts are not suitable for cosmological analysis.

We define our efficiency or true positive rate as:

$$\epsilon_{Ia} = \frac{n_{Ia}^{true}}{N_{Ia}^{total}} \tag{4.1}$$

where n_{Ia}^{true} are type Ia supernovae that have the correct redshift and are correctly classified and N_{Ia}^{total} contain all synthetic SNIa before classification. In order to use the statistics of different cross-validation folds, we define the total efficiency as the weighted sum of each fold efficiency.

We make a distinction between: *total efficiency*, which is includes the detection, selection and classification steps; the *SN-like efficiency* which evaluates only the effect of selection cuts and classification; and the *classification efficiency*, which assesses our machine learning classification methods only.

The purity of the SNIa sample is defined as:

$$P_{Ia} = \frac{n_{Ia}^{true}}{n_{Ia}^{true} + n_{Ia}^{false}} \tag{4.2}$$

where n_{Ia}^{true} is defined above and n_{Ia}^{false} are either core-collapse that were classified as type Ia SNe, or type Ia supernovae with inaccurate redshifts.

Contamination by core collapse or SNe Ia with inaccurate redshifts is defined as:

$$C_i = \frac{n_{Ia}^{false\ i}}{n_{Ia}^{true} + n_{Ia}^{false}} \tag{4.3}$$

where i is the contaminating type.

Since our Ia and CC light-curve simulations are volumetric ones, SN rates are accounted for by weighting events when computing efficiencies and purity. We take volumetric rates for type Ia and core-collapse SNe from [27] and [7], respectively.

4.4 Parameter setting and feature selection

Learning algorithms have a set of chosen parameters, referred to as hyperparameters. In this work we implemented an automated search of hyperparameters using a grid of possible parameter values evaluated on our cross-validated sample with an appropriate score for each classifier (e.g. RF: mean accuracy, XGB: log loss). Using available scikit-learn tools to select hyperparameters and rank feature importance [25].

Inefficient features were identified by measuring the impact of each feature in the model score. They were discarded in our classification. To avoid biasing this selection we used an iterative procedure where the order of evaluation of each feature was permuted at each round.

For each classification method a different ranking of features was obtained. However, some features were selected as efficient for all methods. These were: photometric SN redshift, the χ^2_ν of the agreement in color and width of the redshift fit, the rise time for i, r and g filters, the fall time for i and r filters, the χ^2_ν of the general light-curve fit for i and r filters and the amplitude in g for the general light-curve fitter.

5 Comparison of different classification methods

5.1 Application on simulated SNe

Three classification methods are examined: Random Forest, AdaBoost and XGBoost Decision Trees. We evaluate the performance of the different methods using the AUC metric from the ROC curve shown in Figure 5. All methods obtain an AUC metric above 0.95 and are therefore considered as excellent classifiers. This curve shows the trade-off between having a high positive rate (equivalent to classification efficiency) and a low false positive rate (equivalent to contamination). In the following, we study, the differences between methods and their impact on both simulated and real data.

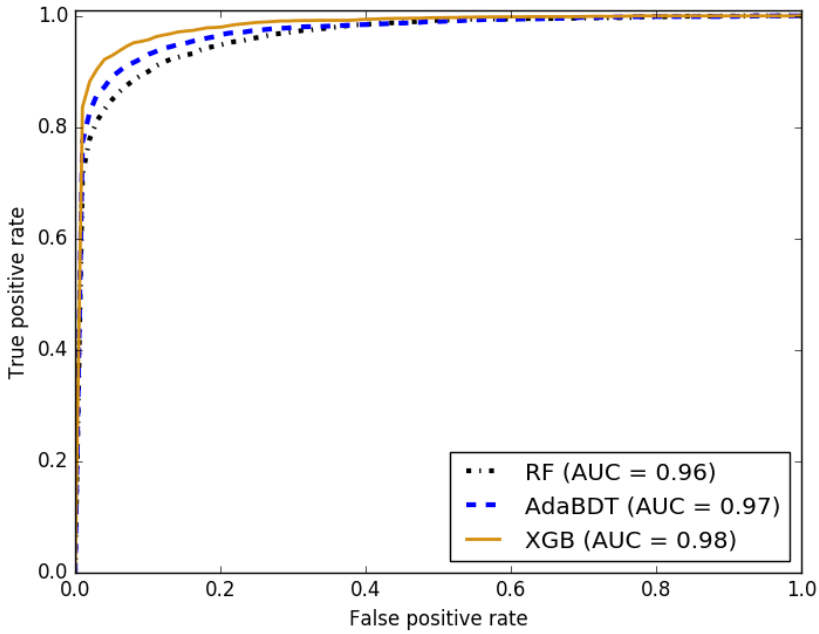


Figure 5: Receiver operating characteristic (ROC) curves for different classification methods applied on synthetic SNe. True and false positive rates are given by Equations 4.1 and 4.3 respectively. Each curve represents a different algorithm: Random Forest RF (dotted black), AdaBoost Decision Tree AdaBDT (dashed blue) and Extreme Gradient Boosting XGB (solid yellow). The AUC score (area under the curve) is shown in the legend.

We classify synthetic SNe to estimate efficiencies and purities. For each classifier, a BDT response threshold must be chosen. This choice results from a trade-off between purity and efficiency of the classified sample. For our three methods, we plot in Figure 6 total efficiency (as defined in Section 4.3.2) against purity of the classified sample for different BDT response thresholds. The performance of each algorithm is in agreement with the AUC metric ranking. It is clear that the trade-off between efficiency and purity is more favorable for the XGB algorithm.

To compare the three algorithms, we set the BDT response threshold such that the estimated purity is 95%. Total efficiencies and purities for these samples can be seen in

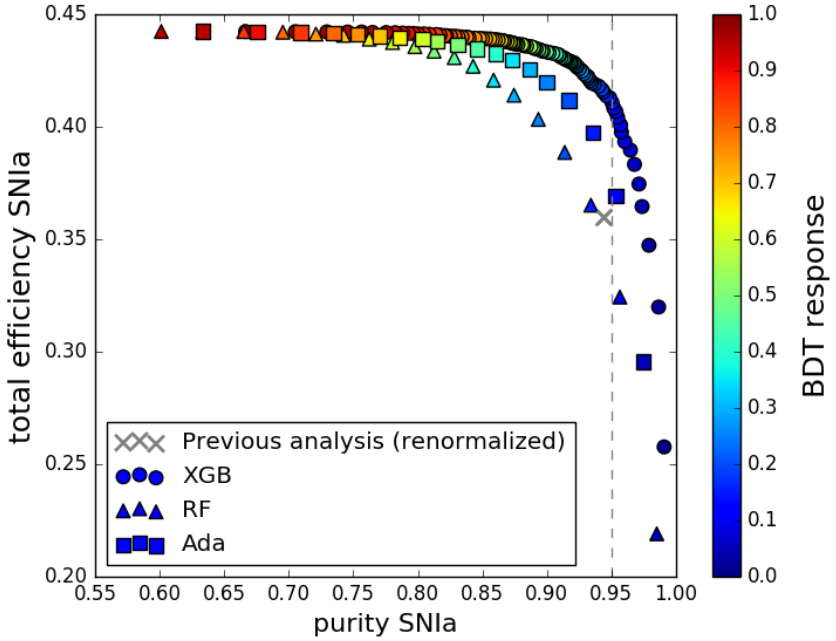


Figure 6: Total efficiency (Equation 4.1) versus purity (Equation 4.3.2) for different choices of the BDT response cut for the 3 classification methods. The plot has been constructed for the simulated sample. The results for the XGB, RF and Ada classifiers are shown as circles, triangles, and squares respectively. The gray dashed line indicates a purity of 95%. The gray cross shows the performance of previous SNLS photometric classification (using host-galaxy redshifts [6]) with the efficiency renormalized upwards to account for the effect of redshift assignment (83%).

Table 2. The algorithm with highest efficiency for our set purity is found to be XGB.

5.1.1 Efficiency evolution with redshift

The total efficiency as a function of redshift is shown in Figure 7 for all classification methods. The higher efficiency at low redshift can be attributed to higher quality light-curves for nearby SNe Ia. The SNIa classification efficiency varies from one algorithm to the other, XGB being the best performing method over the whole redshift range. Interestingly AdaBoost and XGB differences are quite homogeneous which can be attributed to the similarities of their optimization methods.

5.1.2 Evolution of purity with redshift

Figure 8 shows the evolution of purity and contamination as a function of real and photometric SN redshifts for each algorithm. The contamination by type CC SNe is higher at lower real redshift but remains small (below 15%) whatever the method. Comparing the contaminations as a function of real redshift and photometric SN redshift, there is a migration of low-z events towards higher redshifts. This is attributed to inaccurate photometric SN redshifts for some core-collapse events, as illustrated in Figure 3.

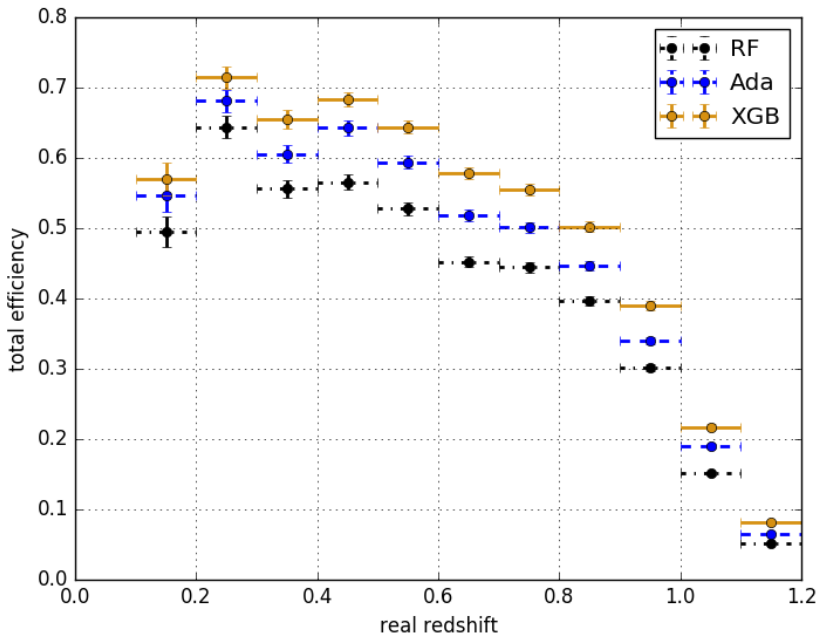
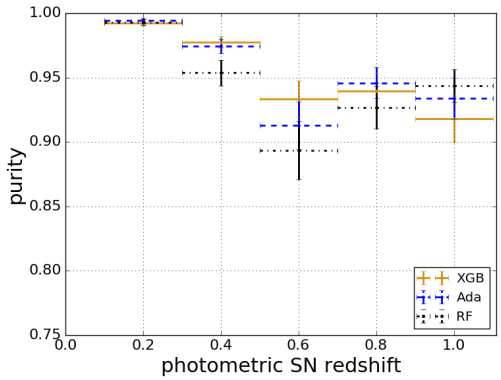
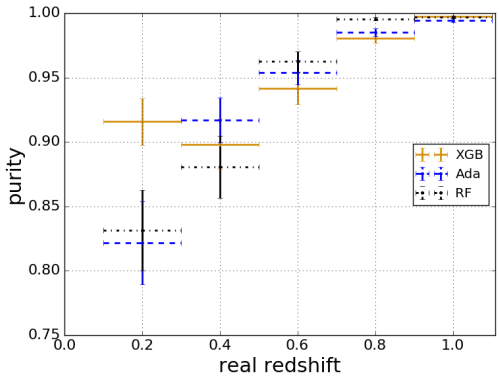


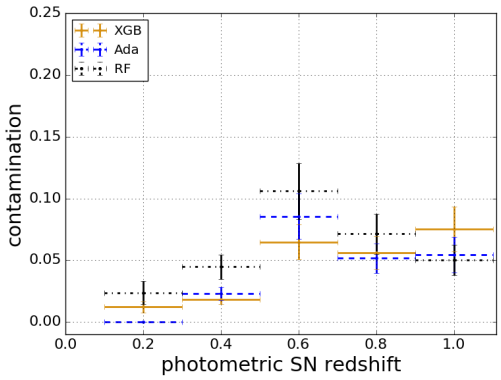
Figure 7: Total efficiency from synthetic SNIa light-curves as a function of the real simulated redshift for different classification methods with the purity set to 95%. Random Forest (RF) points are plotted with black dotted error bars, AdaBoost Decision Tree (Ada) with dashed blue and Extreme Gradient Boosting (XGB) with solid yellow. Note that the total efficiency of XGB is higher than the other two methods at all redshifts.



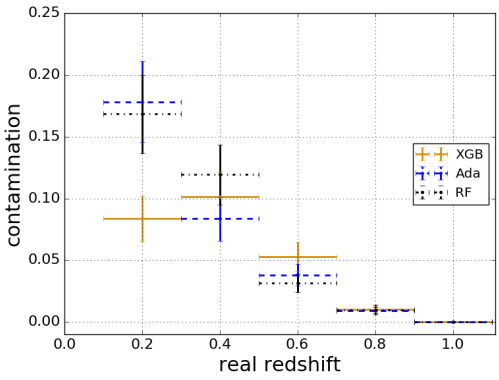
(a) SNIa purity



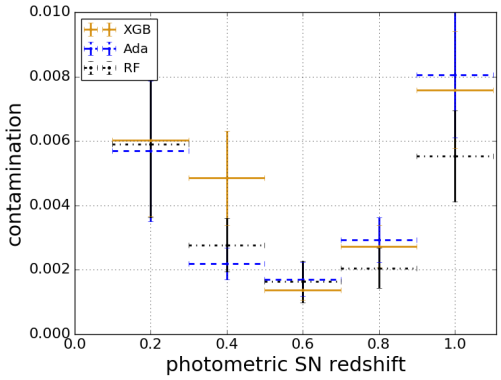
(b) SNIa purity



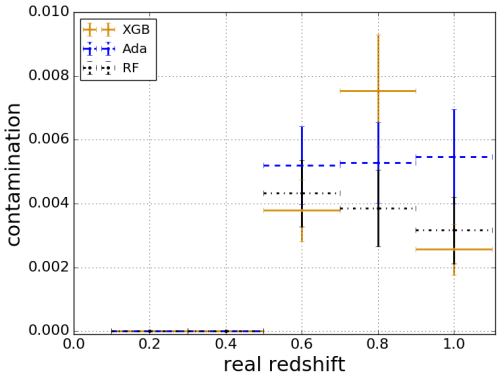
(c) CC contamination



(d) CC contamination



(e) SNIa inaccurate z contamination



(f) SNIa inaccurate z contamination

Figure 8: SNIa purity and background contamination from simulated SN light-curves as a function of redshift for different methods with a given purity of 95%. Left column : photometric SN redshift, right column : real redshift. Same color code as in Figure 7. Note how the CC SNe that are incorrectly classified as SNe Ia tend to have an assigned photometric SN redshift that is much higher than their real redshift.

The contamination by type Ia SNe with inaccurate photometric SN redshift increases with higher redshift, but the overall contamination stays well below 1%.

5.2 Application on SNLS3 data

Classification is also evaluated on SNLS3 selected data with the purity set to 95%. In Figure 9 we show a Venn diagram with the events classified as type Ia SNe by each method. The large number of common events shows the coherence between the three algorithms.

For each method, Table 3 gives the number of events classified as type Ia for the selected sample and the sub-samples of photometrically and spectroscopically identified type Ia and core-collapse SNe. XGB has the largest number of spectroscopically and photometrically classified type Ia supernova. However, since these two samples share common events we visualize the superposition of these samples in Figure 10. XGB continues to have the largest common sample.

For RF, all classified core-collapse events had an inaccurate photometric SN redshift. For AdaBoost, one event was found to have inaccurate photometric SN redshift. The other event, common to both spectroscopic and photometric samples, is a spectroscopically classified type II event whose light-curve is incomplete because it was observed at the end of a season. For XGB classification 4 core-collapse (2 spectroscopic and 2 photometric) events had an inaccurate photometric SN redshift. Four events had correct photometric SN redshifts (1 spectroscopic, 2 photometric and one common to both samples). One last event was classified photometrically, therefore no spectroscopic redshift was available.

The XGB method selects more CC events than the other two algorithms. Given the expected CC contaminations (see Table 2) and photometric sample sizes (see Table 3), we expect XGB to classify 20 to 40% more CC events than the other two methods, less than what we observe in data on the two test-samples of CC events that we have at our disposal. This might be a statistical fluctuation, or a reflection of the incompleteness of the CC test-samples or an indication that our photometric samples are still contaminated by residual non SN-backgrounds that make our expected CC contaminations only indicative.

All classified samples contain the same spectroscopically confirmed SNIa with inaccurate photometric SN redshift when compared to its spectroscopic redshift. Using Tables 2 and 3, the total number of type Ia SNe with inaccurate redshifts is expected to be between 1.5 and 4 events depending on the algorithm. This is in reasonable agreement with what we see in data on the sub-sample of spectroscopically identified SNe Ia for which we have both redshifts.

To check the agreement between expectations and data with the three methods, we compare the SNIa efficiency ratio between any two methods with the ratio of the classified sample sizes for the same two methods. The expected and observed ratios based on XGB and RF compare well. The two ratios defined with respect to AdaBoost are found to be higher in data than expected. This discrepancy remains unexplained. It may indicate that the XGB and RF samples are contaminated, in the same way, by non-SN backgrounds.

5.2.1 Comparison with the SNLS3 subluminous and peculiar SNIa samples

The SN-like sample (the starting point of our classification) contains 11 photometrically identified subluminous events and 5 spectroscopically identified peculiar events.

For all classification methods, the same 3 peculiar events are contained in our photometrically classified sample. None of them exhibit any sign of peculiarity in their light-curves.

		AdaBoost	Random Forest	XGBoost
total efficiency	Ia	36.9 ± 0.6	32.4 ± 0.7	41.1 ± 0.7
purity	Ia	95.6 ± 0.5	95.6 ± 0.4	95.3 ± 0.4
contamination	Ia inaccurate z	0.53 ± 0.09	0.29 ± 0.07	0.60 ± 0.09
contamination	CC	3.9 ± 0.4	4.1 ± 0.4	4.0 ± 0.4

Table 2: Estimated total efficiency, purity and contamination from simulated SNe for different methods with a given purity of 95%.

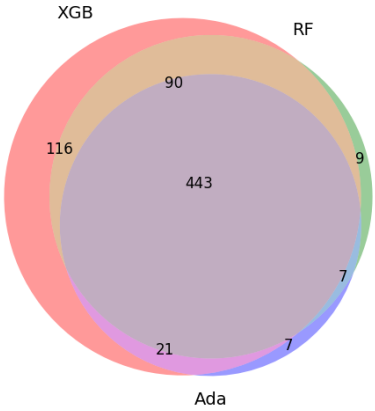


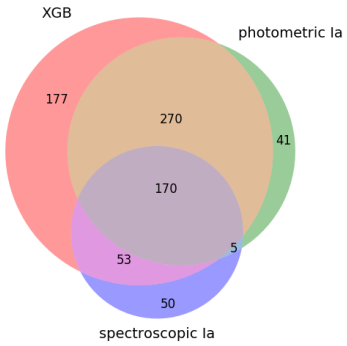
Figure 9: Venn diagram for SNLS3 data photometrically classified by three different algorithms with the purity set to 95%. Color code as follows : overlap between the three methods in mauve, overlap between RF and XGB (Ada) in beige (light blue), between XGB and Ada in purple, pure RF in green, XGB in pink and Ada in dark blue. The total number of classified events for each method are given in Table 3.

	AdaBoost	Random Forest	XGBoost
photometric sample	478	549	670
spectroscopic Ia	166	198	223
photometric Ia	318	364	444
spectroscopic CC	2	2	3
photometric CC	1	1	6

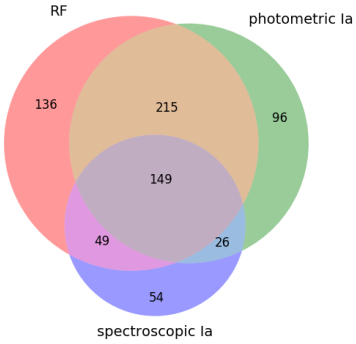
Table 3: Events classified as SNe Ia by the three methods with purity set at 95%. The first line gives the numbers found from SNLS3 and the next four lines give the numbers for the spectroscopically and photometrically identified subsamples [16], [6], [7].

The super-Chandrasekhar type Ia and the 1991T-like object are not classified as type Ia SNe by any of our methods.

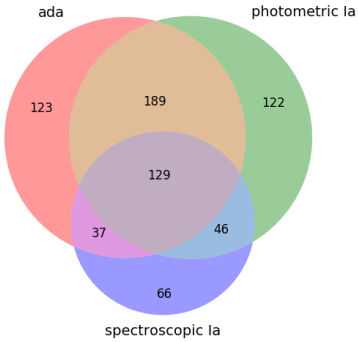
Subluminous supernovae are found in our classified samples. In the case of Random Forest and AdaBoost classifications, 4 events are in the classified sample while 8 are included in the XGB sample. Despite our methods not being trained for disentangling normal type



(a) XGB



(b) Random Forest



(c) AdaBoost

Figure 10: Venn diagrams for each classification algorithm, showing the intersection between events classified by our methods and the photometric and spectroscopic classifications. Dark blue and green regions correspond to events in those samples missed by our classification.

Ia and subluminous SNe, our photometric classification appears to have some efficiency in

	AdaBoost	Random Forest	XGBoost
% SNIa ($CI = 4$ or 5)	74 ± 3	87 ± 3	96 ± 2
% SNIa* ($CI = 3$)	58 ± 6	73 ± 6	88 ± 4

Table 4: Classification efficiency for our three methods when compared with the confidence index of the spectroscopic classification in [19].

cut	SNLS3 events in sample	spectroscopic Ia	spectroscopic CC	photometric Ia	photometric CC	simulated Ia%
SN-like	1483	246	42	486	109	50
selected	1193	238	30	481	77	47
classified	529	205	1	374	1	35

Table 5: Effect of selection cuts and classification (XGB) using SNLS3 data and synthetic type Ia. The classifier threshold is adjusted so that the purity is 98%.

detecting subluminous SNe Ia as well.

5.2.2 Effect of spectroscopic confidence index

In Section 2.2 we split the spectroscopically confirmed type Ia SNe according to the confidence level of the spectroscopic identification. Table 4 shows the percentage of events correctly classified for each method and sub-class. All three methods in this work have a larger classification efficiency for SNe Ia with the highest confidence index.

6 Choosing a method: XGB with high purity 98%

The best performing algorithm was found to be XGB with high achievable purity and efficiency. We chose to select a sample with a purity of $98.0 \pm 0.3\%$. The corresponding total efficiency is $34.7 \pm 0.7\%$. We now study this sample in detail.

6.1 Effect of selection cuts and classification

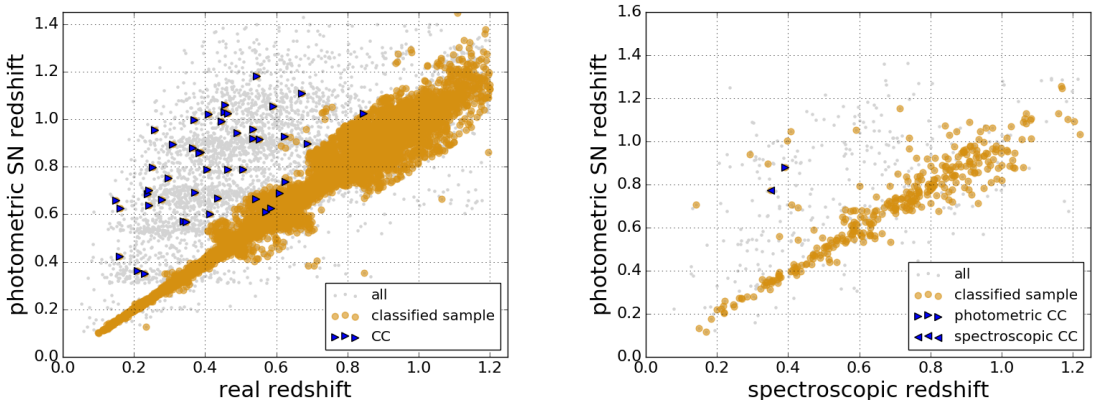
The impact of the selection cuts and the classification is shown for data and synthetic SNe Ia in Table 5. The selection cuts (defined in Section 3.2) are shown to reduce the spectroscopic and photometric type Ia subsamples by 3.3% and 1% respectively. The core-collapse SNe are mainly discarded through classification. The two core-collapse events remaining after classification have inaccurate photometric SN redshifts. The classified sample contains 6 subluminous and 3 peculiar type Ia SNe from samples introduced in Section 5.2.1, and a spectroscopic type Ia that is classified with an inaccurate photometric SN redshift.

6.2 Classification and photometric SN redshifts

We investigate the impact the accuracy of the fitted photometric SN redshifts has on the classification. Figure 11 shows the comparison between spectroscopic and photometric SN redshifts for events in the classified sample when both redshifts are available. Contamination by core-collapse SNe is mostly due to events that have an inaccurate photometric SN redshift.

Interestingly, those core-collapse SNe that were assigned correct photometric SN redshifts were not classified as type Ia SNe by our method. A core-collapse event that has the

correct photometric SN redshift is an event that has colors and photometry consistent with a type Ia supernova (photometric SN redshifts are obtained under the hypothesis that the object is a SNIa, see Section 3.1). We highlight this rejection by our classification of core-collapse SNe with properties similar to SNe Ia and attribute it to the features obtained using the general SN fitter (see Section 3.3).



(a) synthetic SNe in our photometric sample

(b) SNLS3 SNe in our photometric sample

Figure 11: Photometric SN redshift versus real redshift for synthetic SNe (left) and SNLS3 data (right). All SNe in the SNLS3 selected sample are shown as gray dots. Events classified as type Ia with the XGB method at 98% purity are shown as yellow circles. Blue triangles indicate core-collapse SNe classified as type Ia by our method.

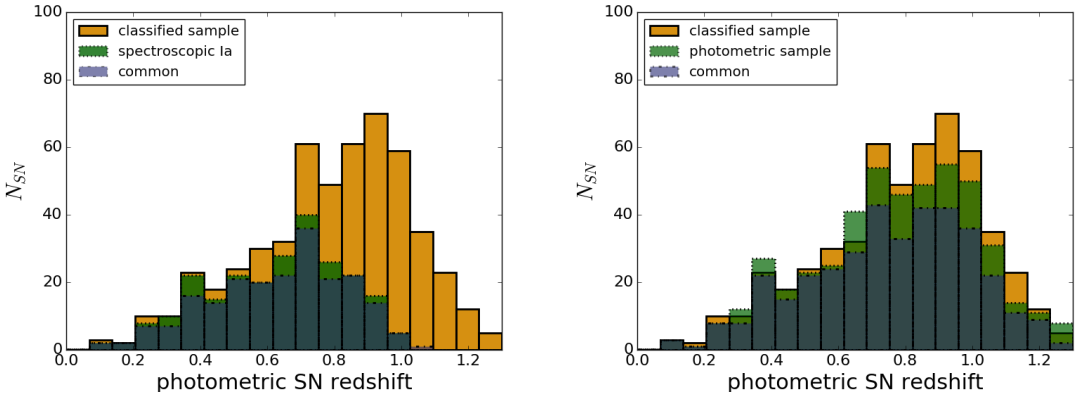
The photometric SN redshift distribution of classified events peaks at higher redshifts when compared to the spectroscopically identified sample (Figure 12a). There is a large overlap between events in both samples and no particular trend over photometric SN redshift is seen.

The distribution of the SN-photometric redshift for the photometric sample classified in [6] and the one of this work are shown in Figure 12b. The new classification provides a larger number of $z > 0.7$ events while maintaining the number of events at lower redshift, and therefore a large fraction of the spectroscopic and photometric samples.

6.3 Efficiency evolution: classification, selection and total

In Figure 13 we evaluate the efficiency-purity diagram taking into account: classification only, classification and selection cuts of Section 3.2 and the complete pipeline. Our machine learning classification can achieve a 100% efficiency at the expense of selecting a photometric sample with purity of at most 70%. The effect of adding selection cuts (necessary to ensure light-curve quality and reduction of non-SN backgrounds) is to reduce the maximum achievable efficiency to 80%. When accounting for the rest of the pipeline (detection and SN-like cuts) the maximum achievable efficiency drops to 45%.

Despite a $\sim 20\%$ reduction in the maximum achievable efficiency between classification only and classification with selection cuts, we emphasize the importance of these cuts that reduce other possible non SN-like events. This is of great importance since our algorithms have been trained to disentangle type Ia from core-collapse SNe only.



(a) comparison with spectroscopic sample

(b) comparison with photometric sample

Figure 12: Photometric SN redshift distribution of the XGB classified SNLS3 sample in yellow (solid line) compared with spectroscopically (12a) and photometrically (12b) identified test samples. Common events between samples are indicated. The photometric test sample is that obtained in [6] using host-galaxy redshifts.

6.4 The effect of the spectroscopic confidence index

Spectroscopically identified type Ia SNe in SNLS3 with high confidence index $CI = 4$ or 5 (as defined in Section 2.2) have a photometric classification efficiency of $90 \pm 2\%$. Those events with a CI of 3 have an efficiency in our analysis of $75 \pm 6\%$. SNe with $CI = 3$, are on average more distant than those with $CI = 4$ and $CI = 5$, so the photometry (and spectroscopy) will be noisier. This leads to a lower classification efficiency.

6.5 The effect of light-curve quality

We studied the performance of our classifier according to the quality of the available light-curves for type Ia SNe. The quality of light-curves was assessed through the number of exposures in the i and r filters before and after maximum light. In Figure 14, we show the percentage of correctly classified type Ia SNe as a function of the number of exposures in the previously mentioned filters.

For type Ia SNe, the larger the number of measurements after maximum, the higher the percentage of correctly classified events. This occurs for all redshift intervals. High redshift events require, as expected, better sampling to be correctly classified.

7 Conclusions

In this paper, we presented a new method for photometrically classifying type Ia supernovae using photometric redshifts derived from SN light-curves and machine learning techniques. This work is the first time that machine learning has been used to classify high redshift supernovae from photometry alone. We show that a sample of SNIa can be photometrically classified with a purity that is greater than 95%. Compared to previous work using external host photometric redshifts and sequential cuts, we obtain a purer sample at an equivalent efficiency.

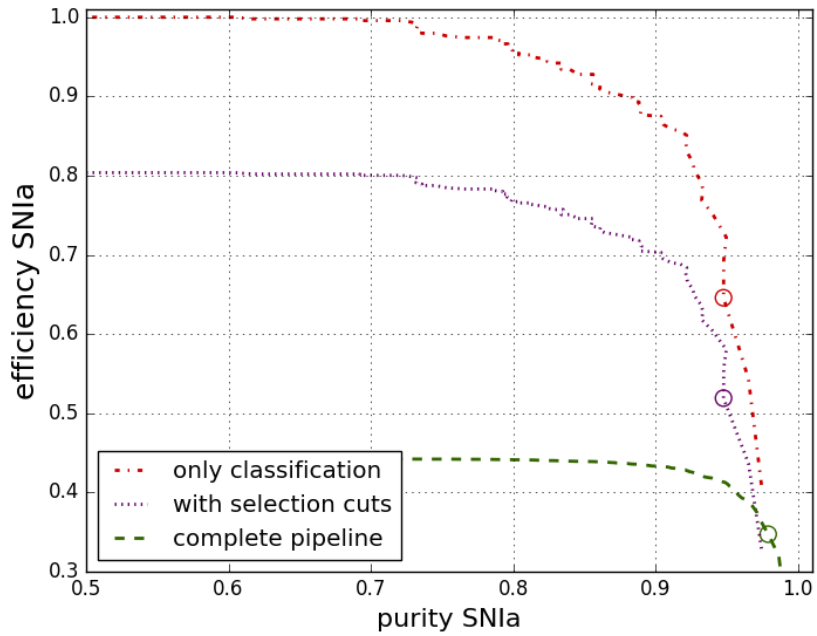
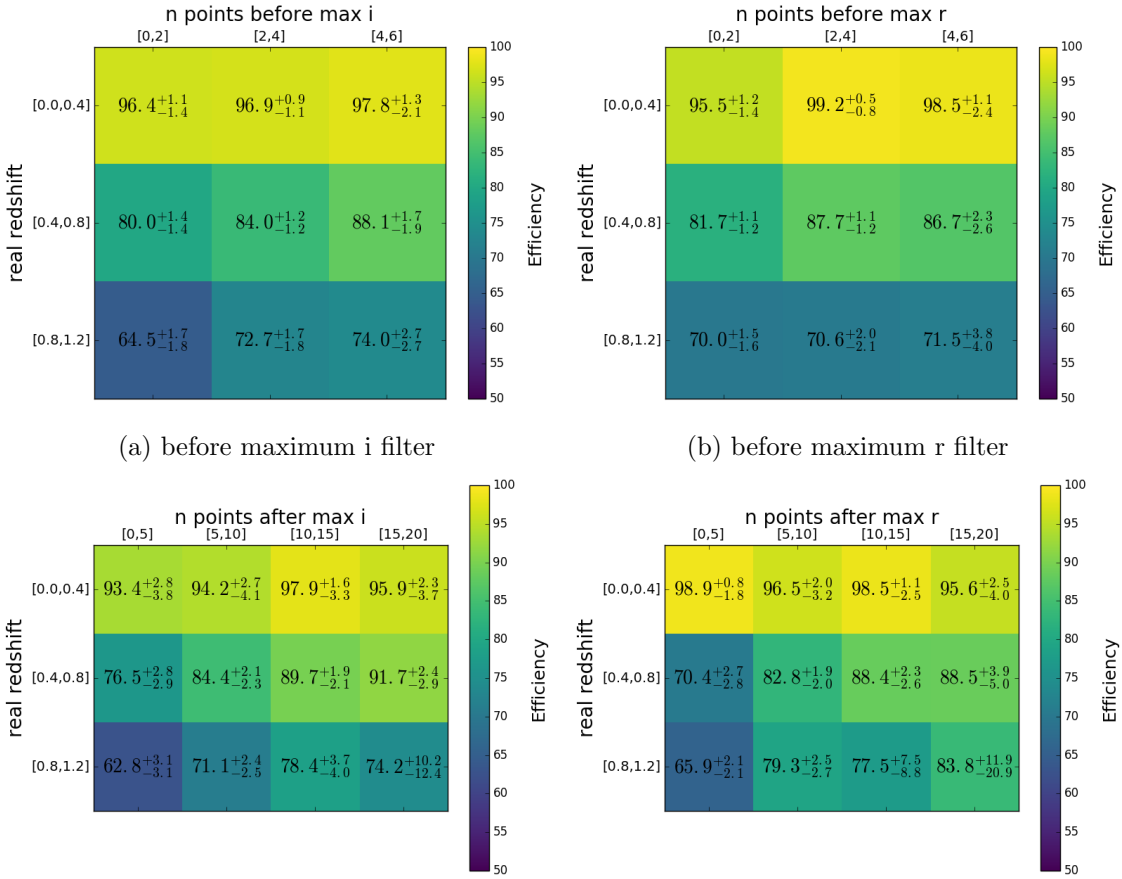


Figure 13: Efficiency versus purity for XGB classification of simulated SNe. We show the trade-off between efficiency and purity when considering only classification effects (red dash-dot line), including selection cuts (purple dotted line) and taking into account the complete pipeline including detection (dashed green line). Circles indicate the evolution of efficiency and purity for the same BDT threshold (chosen such as to obtain a 98% purity for the complete pipeline).

We studied three different supervised learning algorithms for classification: Random Forest, and Decision Trees boosted using AdaBoost and XGBoost algorithms. We compared the results of the three classifiers using both simulated and real SN data. For a purity of 95%, we find that total efficiencies can vary by 10% from one algorithm to the other, which may be linked to their optimization procedures. The XGB algorithm has the best performance both in terms of the AUC score (with a score of 0.98, with 1 being the perfect score) and the estimated efficiency when compared to other methods. When applied to real data, we obtain photometrically classified samples that are double the size of the spectroscopically confirmed sample in SNLS3. The coherence between the three algorithms can be seen from the large number of common classified events.

The best performing classification algorithm was found to be XGBoost. When trained with our synthetic SNe, it is able to provide a sample of 98% purity and satisfactory efficiency. Core-collapse contamination is shown to be dominated by events with inaccurate redshifts. Interestingly, core collapse events with correct redshifts are properly classified as background by our method. These events have colors and photometry consistent with type Ia SNe and therefore should be harder to disentangle. This highlights the performance of our classification using features from a general SN light-curve fitter and the XGB algorithm.

In a real SN survey, efficiency is affected by different stages of the pipeline. In other



(a) before maximum i filter

(b) before maximum r filter

(c) after maximum i filter

(d) after maximum r filter

Figure 14: Percentage of simulated type Ia SNe correctly classified. The data is binned in redshift intervals (y axis) and the number of exposures before and after (top and bottom) maximum (left: i filter, right: r filter). The text indicates the percentage of correctly classified events in each bin with 1σ confidence intervals.

classification studies, efficiencies and purities are computed directly from generated SN light-curves without taking into account selection cuts. These cuts are fundamental for the selection of a sample where non-modeled backgrounds are limited. The impact of selection cuts will vary with the pipeline and must be studied case by case.

We find that selection cuts are fundamental for supernova photometric classification with supervised learning. An algorithm is only as good as its training set. Therefore if other backgrounds are present in the sample, the algorithm will perform less well. We argue that if the goal is SN classification, a substantial study must be done to ensure that non-modeled backgrounds are strongly reduced, and the extracted features are meaningful.

We acknowledge that our study was limited by the number of simulated SNe. Although $20,000$ simulated core-collapse were generated, when applying selection cuts to obtain the SN-like sample (the starting point of our classification) only a small percentage of core-collapse pass them. We expect a low number of core-collapse events at this stage, but it would be advisable to have a larger number to be used as training and for estimating efficiency and

purity. This paper is a first step towards classification of real SN data using supervised learning and we will address this limitation in future work.

This work demonstrates for the first time the feasibility of machine learning classification in a high-redshift SN survey with application to real SN data. We have successfully classified a high-purity type Ia photometric supernova sample in the SNLS survey. An analysis of the impact on cosmology coming from the use of supervised learning techniques to produce SN samples will be subject of a future work. Additionally, this classification will be applied to the SNLS 5-year photometric analysis that will be the subject of a forthcoming paper.

Acknowledgments

Part of this research was conducted by the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020.

AM thanks B. Schmidt, F. Yuan and B. Tucker for useful discussions.

This work was done based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/IRFU, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de lUnivers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at Terapix available at the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

References

[1] **Supernova Search Team** Collaboration, A. G. Riess et al., *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, *Astron.J.* **116** (1998) 1009–1038, [astro-ph/9805201].

[2] **Supernova Cosmology Project** Collaboration, S. Perlmutter et al., *Measurements of Omega and Lambda from 42 high redshift supernovae*, *Astrophys.J.* **517** (1999) 565–586, [astro-ph/9812133].

[3] **SDSS Collaboration** Collaboration, M. Betoule et al., *Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples*, *Astron.Astrophys.* **568** (2014) A22, [arXiv:1401.4064].

[4] J. P. Bernstein et al., *Supernova Simulations and Strategies For the Dark Energy Survey*, *Astrophys. J.* **753** (2012) 152, [arXiv:1111.1969].

[5] **LSST Science Collaborations, LSST Project** Collaboration, P. A. Abell et al., *LSST Science Book, Version 2.0*, arXiv:0912.0201.

[6] G. Bazin, V. Ruhlmann-Kleider, N. Palanque-Delabrouille, J. Rich, E. Aubourg, et al., *Photometric selection of Type Ia supernovae in the Supernova Legacy Survey*, *Astron.Astrophys.* **534** (2011) A43, [arXiv:1109.0948].

[7] **SNLS** Collaboration, G. Bazin et al., *The Core-collapse rate from the Supernova Legacy Survey*, *Astron.Astrophys.* **499** (2009) 653, [arXiv:0904.1066].

[8] O. Ilbert, S. Arnouts, H. McCracken, M. Bolzonella, E. Bertin, et al., *Accurate photometric redshifts for the cfht legacy survey calibrated using the vimos vlt deep survey*, *Astron.Astrophys.* **457** (2006) 841–856, [astro-ph/0603217].

[9] N. Palanque-Delabrouille, V. Ruhlmann-Kleider, S. Pascal, J. Rich, J. Guy, et al., *Photometric redshifts for supernovae Ia in the Supernova Legacy Survey*, *Astron.Astrophys.* **514** (2010) A63, [arXiv:0911.1629].

[10] R. Kessler, A. Conley, S. Jha, and S. Kuhlmann, *Supernova Photometric Classification Challenge*, arXiv:1001.5210.

[11] R. Kessler et al., *Results from the Supernova Photometric Classification Challenge*, *Publ. Astron. Soc. Pac.* **122** (2010) 1415–1431, [arXiv:1008.1024].

[12] N. V. Karpenka, F. Feroz, and M. P. Hobson, *A simple and robust method for automated photometric classification of supernovae using neural networks*, *Mon. Not. Roy. Astron. Soc.* **429** (2013) 1278, [arXiv:1208.1264].

[13] E. E. O. Ishida and R. S. de Souza, *Kernel PCA for type Ia supernovae photometric classification*, *Mon. Not. Roy. Astron. Soc.* **430** (2013) 509, [arXiv:1201.6676].

[14] M. Lochner, J. D. McEwen, H. V. Peiris, O. Lahav, and M. K. Winter, *Photometric Supernova Classification With Machine Learning*, arXiv:1603.0088.

[15] N. Regnault, A. Conley, J. Guy, M. Sullivan, J.-C. Cuillandre, P. Astier, C. Balland, S. Basa, R. G. Carlberg, D. Fouchez, D. Hardin, I. M. Hook, D. A. Howell, R. Pain, K. Perrett, and C. J. Pritchet, *Photometric calibration of the Supernova Legacy Survey fields*, *AAP* **506** (Nov., 2009) 999–1042, [arXiv:0908.3808].

[16] **SNLS Collaboration** Collaboration, J. Guy et al., *The Supernova Legacy Survey 3-year sample: Type Ia Supernovae photometric distances and cosmological constraints*, *Astron.Astrophys.* **523** (2010) A7, [arXiv:1010.4743].

[17] **SNLS Collaboration** Collaboration, J. Guy et al., *SALT2: Using distant supernovae to improve the use of Type Ia supernovae as distance indicators*, *Astron.Astrophys.* **466** (2007) 11–21, [astro-ph/0701828].

[18] K. Perrett, D. Balam, M. Sullivan, C. Pritchet, A. Conley, R. Carlberg, P. Astier, C. Balland, S. Basa, D. Fouchez, J. Guy, D. Hardin, I. M. Hook, D. A. Howell, R. Pain, and N. Regnault, *Real-time Analysis and Selection Biases in the Supernova Legacy Survey*, *AJ* **140** (Aug., 2010) 518–532, [arXiv:1006.2254].

[19] D. A. Howell, M. Sullivan, K. Perrett, T. J. Bronder, I. M. Hook, P. Astier, E. Aubourg, D. Balam, S. Basa, R. G. Carlberg, S. Fabbro, D. Fouchez, J. Guy, H. Lafoux, J. D. Neill, R. Pain, N. Palanque-Delabrouille, C. J. Pritchet, N. Regnault, J. Rich, R. Taillet, R. Knop, R. G. McMahon, S. Perlmutter, and N. A. Walton, *Gemini Spectroscopy of Supernovae from the Supernova Legacy Survey: Improving High-Redshift Supernova Selection and Classification*, *APJ* **634** (Dec., 2005) 1190–1201, [astro-ph/0509195].

[20] S. Gonzalez-Gaitan et al., *Subluminous Type Ia Supernovae at High Redshift from the Supernova Legacy Survey*, *Astrophys. J.* **727** (2011) 107, [arXiv:1011.4531].

[21] **SNLS** Collaboration, C. Balland et al., *The ESO/VLT 3rd year Type Ia supernova data set from the Supernova Legacy Survey*, *Astron. Astrophys.* **507** (2009) 85, [arXiv:0909.3316].

[22] **SNLS** Collaboration, T. J. Bronder et al., *SNLS Spectroscopy: Testing for Evolution in Type Ia Supernovae*, *Astron. Astrophys.* **477** (2008) 717, [arXiv:0709.0859].

[23] **SNLS** Collaboration, R. S. Ellis et al., *Verifying the Cosmological Utility of Type Ia Supernovae: Implications of a Dispersion in the Ultraviolet Spectra*, *Astrophys. J.* **674** (2008) 51–69, [arXiv:0710.3896].

[24] **SNLS** Collaboration, D. A. Howell et al., *The type Ia supernova SNLS-03D3bb from a super-Chandrasekhar-mass white dwarf star*, *Nature* **443** (2006) 308, [astro-ph/0609616].

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825–2830.

[26] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, *ArXiv e-prints* (Mar., 2016) [arXiv:1603.0275].

[27] **SNLS** Collaboration, C. J. Pritchet, D. A. Howell, and M. Sullivan, *The Progenitors of Type Ia Supernovae*, *Astrophys. J.* **683** (2008) L25, [arXiv:0806.3729].