

# The Dark Energy Survey 5-year photometrically identified Type Ia Supernovae

A. Möller<sup>1,2\*</sup> †, M. Smith<sup>3</sup>, M. Sako<sup>4</sup>, M. Sullivan<sup>5</sup>, M. Vincenzi<sup>6</sup>, P. Wiseman<sup>5</sup>, P. Armstrong<sup>7</sup>, J. Asorey<sup>8</sup>, D. Brout<sup>9</sup>, D. Carollo<sup>10</sup>, T. M. Davis<sup>11</sup>, C. Frohmaier<sup>5,12</sup>, L. Galbany<sup>13,14</sup>, K. Glazebrook<sup>1</sup>, L. Kelsey<sup>5,12</sup>, R. Kessler<sup>15,16</sup>, G. F. Lewis<sup>17</sup>, C. Lidman<sup>7,18</sup>, U. Malik<sup>7</sup>, R.C. Nichol<sup>19</sup>, D. Scolnic<sup>6</sup>, B. E. Tucker<sup>7</sup>, T. M. C. Abbott<sup>20</sup>, M. Aguena<sup>21</sup>, S. Allam<sup>22</sup>, J. Annis<sup>22</sup>, E. Bertin<sup>23,24</sup>, S. Bocquet<sup>25</sup>, D. Brooks<sup>26</sup>, D. L. Burke<sup>27,28</sup>, A. Carnero Rosell<sup>21,29,30</sup>, M. Carrasco Kind<sup>31,32</sup>, J. Carretero<sup>33</sup>, F. J. Castander<sup>34,35</sup>, C. Conselice<sup>36,37</sup>, M. Costanzi<sup>38,39,40</sup>, M. Crocce<sup>34,35</sup>, L. N. da Costa<sup>21,41</sup>, J. De Vicente<sup>8</sup>, S. Desai<sup>42</sup>, H. T. Diehl<sup>22</sup>, P. Doel<sup>26</sup>, S. Everett<sup>43</sup>, I. Ferrero<sup>44</sup>, D. A. Finley<sup>22</sup>, B. Flaugher<sup>22</sup>, D. Friedel<sup>31</sup>, J. Frieman<sup>16,22</sup>, J. García-Bellido<sup>45</sup>, D. W. Gerdes<sup>46,47</sup>, D. Gruen<sup>25,48</sup>, R. A. Gruendl<sup>31,32</sup>, J. Gschwend<sup>21,41</sup>, G. Gutierrez<sup>22</sup>, K. Herner<sup>22</sup>, S. R. Hinton<sup>11</sup>, D. L. Hollowood<sup>43</sup>, K. Honscheid<sup>49,50</sup>, D. J. James<sup>9</sup>, K. Kuehn<sup>51,52</sup>, N. Kuropatkin<sup>22</sup>, O. Lahav<sup>26</sup>, M. March<sup>4</sup>, J. L. Marshall<sup>53</sup>, F. Menanteau<sup>31,32</sup>, R. Miquel<sup>33,54</sup>, R. Morgan<sup>55</sup>, A. Palmese<sup>56</sup>, F. Paz-Chinchón<sup>31,57</sup>, A. Pieres<sup>21,41</sup>, A. A. Plazas Malagón<sup>58</sup>, A. K. Romer<sup>59</sup>, A. Roodman<sup>27,28</sup>, E. Sanchez<sup>8</sup>, V. Scarpine<sup>22</sup>, M. Schubnell<sup>47</sup>, S. Serrano<sup>34,35</sup>, I. Sevilla-Noarbe<sup>8</sup>, E. Suchyta<sup>60</sup>, G. Tarle<sup>47</sup>, D. Thomas<sup>12</sup>, C. To<sup>49</sup>, T. N. Varga<sup>61,62</sup>

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

As part of the cosmology analysis using Type Ia Supernovae (SN Ia) in the Dark Energy Survey (DES), we present photometrically identified SN Ia samples using multi-band light-curves and host galaxy redshifts. For this analysis, we use the photometric classification framework SUPERNNova trained on realistic DES-like simulations. For reliable classification, we process the DES SN programme (DES-SN) data and introduce improvements to the classifier architecture, obtaining classification accuracies of more than 98 per cent on simulations. This is the first SN classification to make use of ensemble methods, resulting in more robust samples. Using photometry, host galaxy redshifts, and a classification probability requirement, we identify 1,863 SNe Ia from which we select 1,484 cosmology-grade SNe Ia spanning the redshift range of  $0.07 < z < 1.14$ . We find good agreement between the light-curve properties of the photometrically-selected sample and simulations. Additionally, we create similar SN Ia samples using two types of Bayesian Neural Network classifiers that provide uncertainties on the classification probabilities. We test the feasibility of using these uncertainties as indicators for out-of-distribution candidates and model confidence. Finally, we discuss the implications of photometric samples and classification methods for future surveys such as Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST).

**Key words:** surveys – supernovae: general – cosmology: observations – methods: data analysis

## 1 INTRODUCTION

To fully exploit the power of current and future time-domain surveys, it is necessary to classify astrophysical objects using only photometry. Surveys such as the Supernova Legacy Survey (SNLS), Sloan Digital Sky Survey (SDSS) SN Survey (SDSS-II), Pan-STARRS (PS1), and the Dark Energy Survey (DES) have discovered thousands of supernovae (SNe) but the majority have not been spectroscopically classified (Astier et al. 2006; Frieman et al. 2008; Sako et al. 2018; Rest et al. 2014; Foley et al. 2018; Bernstein et al. 2012; Smith et al.

2020). Photometric classification will be particularly crucial for the upcoming Legacy Survey of Space and Time (LSST) at the Vera C. Rubin Observatory, which is expected to discover up to  $10^7$  SNe over the next decade (LSST Science Collaboration et al. 2009).

The Dark Energy Survey Supernova programme (DES-SN) obtained photometry of more than 30,000 candidate SNe over its five years of operation. These include thousands of high-redshift SNe Ia, of which only several hundred have been spectroscopically classified. The first three years of the DES-SN detected and spectroscopically classified 251 SNe Ia (Smith et al. 2020). Together with low redshift SNe from the Harvard-Smithsonian Center for Astrophysics surveys (CfA3, CfA4; Hicken et al. 2009, 2012) and the Carnegie Supernova Project (CSP; Contreras et al. 2010; Stritzinger et al. 2011), these SNe

\* E-mail: amoller@swin.edu.au

† Author affiliations are shown in Appendix C

were used to constrain cosmological parameters (Dark Energy Survey 2019). The DES-SN candidate sample also contains other types of transients that have been used for astrophysical and cosmological studies: core-collapse SNe (de Jaeger et al. 2020), superluminous SNe (SLSNe; Smith et al. 2018; Angus et al. 2019; Inserra et al. 2021), rapidly evolving transients (Wiseman et al. 2020b; Pursiainen et al. 2018) and ‘peculiar’ events (Gutiérrez et al. 2020; Grayling et al. 2021).

To classify SNe without spectroscopy, a number of methods have been developed to classify them using their light-curves, i.e., their observed brightness evolution in different filters. Due to their cosmological use, much work has focused on disentangling SNe Ia from other SN types. The majority have been trained and tested on simulations, with only a handful applied to large SN surveys (Sako et al. 2011; Möller et al. 2016; Muthukrishna et al. 2019; Möller & de Boissière 2019; Villar et al. 2019, 2020). Several photometric classifiers have been developed and incorporated into the SNIA-cosmology analysis pipeline PIPPIN (Hinton & Brout 2020), including SNIRF (based on the architecture developed by Dai et al. 2018), SUPERNNova (Möller & de Boissière 2019) and SCONE (Qu et al. 2021).

In this work, we use the non-parametric framework SUPERNNova (SNN; Möller & de Boissière 2019) to obtain photometrically classified SN Ia samples from DES-SN. SNN has several strengths it: (i) requires only photometric information (fluxes and time) for classification, (ii) does not rely on the extraction of features, (iii) can be trained to classify any type of transient event, (iv) can use redshifts to improve accuracy, (v) has been thoroughly tested using simulations, (vi) includes algorithms that assign uncertainties to classification probabilities such as Bayesian Neural Networks (BNNs), and (vii) is already being applied to real survey data, including early light-curve classification in alert streams (FINK broker; Möller et al. 2021).

Photometrically classified SN Ia samples have started to be used in cosmology. First constraints on the cosmic expansion using data from SDSS-II and PS1 have shown the feasibility of using these samples for cosmology and their competitive constraining power on the Dark Energy (Sako et al. 2011; Hlozek et al. 2012; Campbell et al. 2013; Jones et al. 2017, 2018). Most of these results use the Bayesian Estimation Applied to Multiple Species method (BEAMS; Kunz et al. 2007) and its extension ‘BEAMS with Bias Corrections’ (BBC; Kessler & Scolnic 2017). These methods incorporate classification probabilities of SNe Ia into the analysis, thus requiring accurate classification probabilities. Recent work estimates the contamination for cosmological constraints in the DES-SN sample using SNN at less than 1.4 per cent (Vincenzi et al. 2021). Aside from cosmology, photometrically classified samples with SNN have also been used to study SN Ia rates (Wiseman et al. 2021).

This paper is organised as follows: We introduce the DES survey and DES-SN candidate sample in Section 2. In Section 3 we present pre-processing needed for accurate classification, SUPERNNova, realistic simulations, training and classification mechanisms and their metrics. In Section 4 we select photometrically classified SNe Ia using host galaxy redshift information together with multi-band photometry. We explore the use of BNNs for classification in Section 5. Finally, in Section 6, we discuss our results and their implications for future surveys such as LSST.

## 2 DES-SN-5-YEAR

The Dark Energy Survey (DES) was a 6-year photometric survey that used the Dark Energy Camera (DECam; Flaugher et al. 2015) on the Victor M. Blanco telescope in Chile to survey  $5000\text{ deg}^2$  of

the southern hemisphere. For time-domain science, DES imaged ten 3-deg<sup>2</sup> in the *griz* filters during the first five years (Abbott et al. 2018). Eight of these ten fields (X1, X2, E1, E2, C1, C2, S1, and S2) were observed to a single-visit depth of  $m \approx 23.5$  mag (‘shallow fields’), and the other two ‘deep fields’ (X3, C3) were observed to a depth of  $m \approx 24.5$  mag.

### 2.1 DES-SN candidate sample

Transients were identified using the DES Difference Imaging Pipeline DIFFIMG (Kessler et al. 2015) coupled with a machine learning algorithm (Goldstein et al. 2015) to reduce artefacts. A candidate SN is defined from the difference image measurements by requiring at least two detections with a signal-to-noise ratio (SNR) larger than five in any filter. This criteria is designed to remove artefacts and asteroids.

Each DES-SN candidate was originally associated with a host galaxy using the shallower SVA survey, created from DES Science Verification data. For the DES-SN analysis, we use deep co-adds in Wiseman et al. (2020a). The major source of host galaxy redshift information was the Australian Dark Energy Survey (OzDES) programme obtaining spectra with the 2dF fibre positioner and AAOmega spectrograph on the 3.9-m Anglo-Australian Telescope (Yuan et al. 2015; Childress et al. 2017; Lidman et al. 2020). SN hosts in OzDES were observed up to a limiting  $r$  magnitude of  $\approx 24$ . Further details on host galaxy association can be found in Gupta et al. (2016); Vincenzi et al. (2020).

For the 31,636 candidates, 29,113 have an identified host and 11,350 have a spectroscopic redshift ( $\sim 30$  per cent of the candidates).

A sub sample of candidates were selected for real-time spectroscopic follow-up observations for classification. For the first 3 years of the survey, the spectroscopically classified sample is presented in Smith et al. (2020). In this work, we use for comparison a preliminary spectroscopic sample containing additional classifications from the full 5 years of DES-SN. This sample contains 415 spectroscopically confirmed SNe Ia (including all 251 spectroscopically classified SNe Ia from the DES-SN 3-year analysis), 84 core-collapse SNe, 2 peculiar SNe Ia, 20 SLSNe, 55 AGN, 1 Tidal disruption event (TDE), and 2 M-stars. We highlight that this spectroscopically classified sample is not complete (Kessler et al. 2019b) and does not represent the true abundances of different transients in nature.

In this work we use the fluxes and uncertainties obtained from DIFFIMG (Kessler et al. 2015) for the DES-SN candidate sample.

### 2.2 Filtering multi-season and other transients

The DES-SN 5 year candidate sample contains not only supernovae but also astrophysical events such as fast transients and AGNs. These events, called out-of-distribution (OOD) or anomalies, can be hard to characterise and thus simulate, therefore photometric classifiers are usually not trained to identify them.

To reject fast, very low SNR transients or transients that have a limited photometric sampling (e.g. transients occurring near the end or beginning of the observing season), we select only transients that have at least 3 nights with a detection that has passed the DES Real/Bogus image classifier (Goldstein et al. 2015).

To reduce the number of slowly-evolving transients that span several observing seasons or multi-season candidates (e.g. AGNs) and spurious detections we make use of two selection criteria. First, we compute the ratio between number of epochs with detections that pass the Real/Bogus classifier, and the total number of epochs with detections. To reject light-curves with long variability periods, we

require this ratio to be the same as in the real-time classification pipeline in Smith et al. (2020). Second, we remove artefacts and transients that have detections in multiple observing seasons. We note that this cut can remove real supernovae, i.e. multiple SNe very close-by in the same galaxy, and is not 100% efficient.

With this filtering, the sample is reduced from 31,636 to 14,070 candidates. This reduces the number of candidates and their contamination; however, some residual AGN and other types of SNe remain. We find that this sample includes 405 spectroscopically classified SNe Ia (247 of which are in the DES-SN 3yr sample), 83 core-collapse SNe, 2 peculiar SNe Ia, 19 SLSNe, 37 AGN, 1 TDE and 1 M-star.

### 2.3 Selection Requirements (cuts)

We apply a series of selection cuts on both the quality of the light-curves and the quality of the redshifts. A thorough review on these cuts and their impact on systematics can be found in Vincenzi et al. (2021).

#### 2.3.1 Loose selection cuts

We select transients that have redshifts obtained from spectra from either the SN or its host galaxy (Lidman et al. 2020) using quality tags in Vincenzi et al. (2020). In this work, we also include lower resolution redshifts from PRIMUS since they are precise enough for photometric classification<sup>1</sup>. After this selection cut we obtain 6,635 SN candidates.

Furthermore, we restrict these redshifts to be within the range of the SNe Ia expected for DES-SN and thus in our simulations,  $z \in [0.05, 1.3]$ . This cut also removes stars in our catalogues.

We fit the light-curves using the SALT2 model (Guy et al. 2007). We require that: (i) at least two filters have at least one observation with SNR larger than 5, (ii) at least one photometric measurement before peak brightness  $t_0$ , and (iii) at least one photometric point ten days after peak brightness.

We select a sample of 2,381 light-curves that satisfy these sampling criteria and have a SALT2 fit that converges and is within SALT2 model boundaries for stretch,  $x_1 \in [-4.9, 4.9]$  and colour  $c \in [-0.49, 0.49]$ . We photometrically classify these candidates in the following. This sample contains a subsample of spectroscopically classified candidates which we will use as a reference: SNe Ia: 366 (DES-SN 3-year: 228), CC 13, SLSN 2, AGN 3. The SALT2 parameters (amplitude, stretch, color) are not used by SNN.

#### 2.3.2 JLA-like cuts

We will consider an additional set of cuts after photometric classification based on the criteria in Vincenzi et al. (2021). They will only be applied when specified.

These cuts are designed to select cosmology-grade SNe Ia and are based on those from the Joint Light-curve Analysis:  $-3.0 < x_1 < 3.0$ ,  $-0.3 < c < 0.3$ , and  $\sigma_{x_1} < 1$  and  $\sigma_{t_0} < 2$  (Betoule et al. 2014). Where  $c$ ,  $x_1$ ,  $\sigma_{t_0}$ ,  $\sigma_{x_1}$  are estimated using SALT2 and represent colour, stretch and uncertainty on  $t_0$  and  $x_1$  respectively. These cuts

<sup>1</sup> The redshifts from the PRISM Multi-object Survey (PRIMUS) were obtained using the Inamori Magellan Areal Camera and Spectrograph camera on the Magellan I Baade 6.5 m telescope (Coil et al. 2011). They are less accurate and they have a higher rate of catastrophic failure, thus not suitable for cosmological constraints.

are implemented in SN Ia cosmology analyses to restrict SNIa parameters to the valid model range, and to reject peculiar SNIa. We also use a SALT2 fit probability  $> 0.001$  selection.

## 3 PHOTOMETRIC CLASSIFICATION

We use the photometric classification algorithm SUPERNNova (SNN) to select SN Ia from the DES-SN 5-year candidate sample that pass loose selection cuts. We introduce pre-processing necessary for accurate photometric classification of our DES-SN 5-year data (Section 3.1). We generate realistic simulations of the DES-SN survey to train and test our photometric classification method (Section 3.2) and the framework SNN (Section 3.3). We evaluate performance and find the best configuration for our framework using small simulations (Section 3.4). We then train optimised models for photometric classification of the DES-SN 5-year sample using larger simulations (Section 3.5).

### 3.1 DES-SN data pre-processing

For accurate photometric classification, the simulations used to train the models and the data to be classified should be similar. While light-curve simulations strive to resemble survey data, pre-processing of the survey data is required to assure this.

First, DES-SN data were taken over five consecutive seasons. Each DES season represented about five months of observations per year. SNe last only for months, thus are only detected in a subset of this photometry. In our simulations (see Section 3.2), supernovae are simulated within a rest-frame time span, e.g. -30 days before to 100 days after peak luminosity. To select an equivalent time window in the DES-SN 5-year data, we first obtain an estimated time of peak brightness ( $t_0$ ) using the SuperNova ANALysis software (SNANA; Kessler et al. 2009). This  $t_0$  estimate is not obtained using SALT2 (Guy et al. 2007), but instead based on max flux in region of dense detections to avoid pathological estimates from a single pathological flux in another season. Once the peak has been determined for each light-curve, we select and classify photometric points within an observed time-window around the light-curve peak of  $[-30, 100]$  days.

Light-curves may contain photometry that has been flagged as flawed. We require that SNN discard photometry that is not reliable using the bitmap flag provided by SOURCE EXTRACTOR (Bertin & Arnouts 1996) and DIFFIMG (Kessler et al. 2015). These photometric outliers are not present in the simulations used to train our photometric classifier. This is in particular important when using normalisation schemes, which will be introduced in Section 3.3.1, since they use maximum fluxes to normalise the light-curves. If that maximum flux comes from a bad photometric point, the light-curve will be distorted and therefore classification will not be accurate. This photometry quality criteria reduces the number of photometric measurements by 6% but keeps the number of transients unchanged.

### 3.2 Simulations of the DES-SN survey

SNN is used with simulations from the supernova analysis software (snana Kessler et al. 2009) and within the PIPPIN orchestration framework (Hinton & Brout 2020). The simulations incorporate information from DES-SN observations (PSF, sky noise, zero point), with detection efficiencies vs. SNR estimated on fake SNe that were overlaid on images and processed with DIFFIMG. Simulations include SNe that have partial light-curves due to season boundaries or observing

**Table 1.** Simulations used for training and testing SNN. Columns indicate simulation name, approximate number of light-curves generated and number of light-curves when balancing simulations to have the same number of normal Type Ia and other SNe.

Simulation name	Number of light-curves ( $10^6$ )	Balanced number of light-curves ( $10^6$ )
TRAIN-SIM	4.5	3.63
S-TRAIN-SIM	2.0	1.4
TEST-SIM	0.8	Not applicable

gaps imitating realistic weather conditions. Detailed information on the inputs necessary to obtain realistic DES-SN simulations can be found in Kessler et al. (2019b). We also make use of recent updates in the library of simulated host galaxies for DES-SN as introduced by Vincenzi et al. (2020). This host galaxy library includes the dependence of SN rates on galaxy properties such as stellar mass and galaxy star formation rate.

We simulate a variety of SNe using volumetric rates and input parameters as described in Vincenzi et al. (2020). Our simulations are performed over a redshift range  $0.05 < z < 1.3$ . These simulations contain normal SNe Ia, peculiar SNe Ia and core-collapse SNe.

Normal SNe Ia are generated using the SALT2 SED model presented in Guy et al. (2007), trained for the JLA sample (Betoule et al. 2014) and extended to UV and IR wavelengths (Pierel et al. 2018) to improve the redshift coverage of our simulated SNe. Volumetric rates from Frohmaier et al. (2019) are used. The intrinsic stretch and colour distributions are taken from Scolnic & Kessler (2016) and we use the G10 intrinsic scatter model from Kessler et al. (2013) based on Guy et al. (2010). Peculiar SNe Ia include SN91bg-like (Kessler et al. 2019a) and SNe Iax (Jha 2017) with models updates in Vincenzi et al. (2021).

We make use of three different core-collapse SN template collections: V19 (Vincenzi et al. 2019), J17 (Jones et al. 2017) and templates used in the Supernova Photometric Classification Challenge (SPCC; Kessler et al. 2010). The main differences between these templates include: the number of SNe used to create them, the rates used, and the interpolation methods and wavelength coverage. Detailed information on these templates can be found in Vincenzi et al. (2019).

Our baseline simulations, and used unless specified, are generated using V19 core-collapse SN templates. Relative core-collapse SN rates are given by Li et al. (2011) updated in Shivvers et al. (2017) and the total rate is assumed to follow the cosmic star formation history presented in Madau et al. (2014) normalised by the local SN rate of Frohmaier et al. (2019).

We generate different simulations to train (TRAIN-SIM and a smaller S-TRAIN-SIM for computing efficiency of certain evaluation tasks) and test (TEST-SIM) SNN as shown in Table 1. For training, after generating the simulation, we randomly trim the simulation to ensure a balanced training sample, with the same number of normal SNe Ia and non-normal Ia (core collapse SNe and peculiar SNe Ia). Volumetric rates guarantee that the mixture of non-Ia SNe is consistent with measured rates. We note that the size of the S-TRAIN-SIM training set is the same as the complete sample used in Möller & de Boissière (2019). Having defined our simulated samples we now turn to methods of classifying them.

### 3.3 SUPERNNova (SNN)

SUPERNNova (Möller & de Boissière 2019) is a deep learning frame-

work for light-curve classification. It makes use of fluxes and their measurement uncertainties over time for accurate classification of time-domain candidates. Additional information such as host galaxy redshifts can be included to improve performance.

SNN includes different classification algorithms, such as LSTM<sup>2</sup> Recurrent Neural Networks (RNNs) and two approximations for Bayesian Neural Networks (BNNs). We show in Fig. 1 the classification probabilities from different methods for a given SN light-curve. These probabilities can be used to select a sample by performing a threshold cut or by weighting the contribution of candidates by their classification score as in the BEAMS and BBC methods (Vincenzi et al. 2021; Kunz et al. 2007; Kessler & Scolnic 2017).

Light-curve simulations are used to train SNN to classify candidates into different classes. For cosmology, it can be trained to accurately classify SNe Ia versus other other kinds of transients. For time-domain astronomy, where brokers are designed to disentangle multiple types of transients, SNN can classify subtypes of SNe or transients simultaneously.

Throughout this work we only perform a binary classification, i.e., a normal SN Ia or a non-Ia SN. Our results are expressed in the form of a prediction of the SN type by using a threshold on the obtained SN Ia probability,  $P$ , larger than 0.5.

#### 3.3.1 SNN normalization schemes: *cosmo* and *cosmo quantile*

Since light-curve fluxes and uncertainties exhibit large variations, SNN supports different input data (e.g., fluxes, flux-uncertainties and time steps) and normalisation schemes (Möller & de Boissière 2019). In previous work, the default was the *global*<sup>3</sup> normalisation. However, to avoid cosmological bias when using redshifts for classification, it is important to avoid using distance information encoded in the apparent magnitudes.

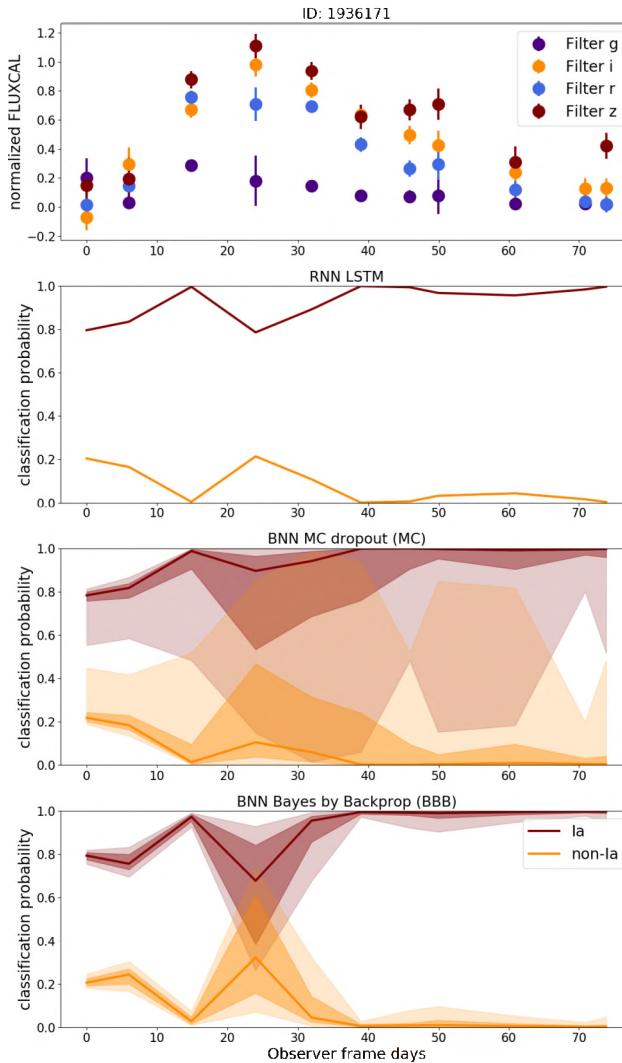
For classification using redshifts, we introduce two new normalisation schemes in SNN that ignore distance information: *cosmo* and *cosmo\_quantile*<sup>4</sup>. In these schemes, for a given light curve, fluxes and their respective uncertainties are normalised by the maximum light-curve flux in any filter (*cosmo*) or the 99th quantile of the flux distribution to avoid normalisation using an outlier (*cosmo\_quantile*). This normalises the fluxes for each light-curve to 1 or near 1, and retains colour and signal-to-noise information for the classification. The normalisation of the time step, given as an input to SNN, remains log transformed and displaced to zero as in the *global* normalisation scheme.

To evaluate these new normalisation schemes, we measure the classification accuracy of SN Ia vs non-SN Ia including redshift as an input using simulations from Möller & de Boissière (2019) since these were the simulations used to benchmark the SNN framework. We find that they slightly improve performance with accuracies of  $99.33 \pm 0.02$  per cent for both *cosmo* and *cosmo\_quantile* as compared to the  $98.43 \pm 0.08$  per cent accuracy of the *global* normalisation scheme using same dataset, redshift information and default settings (seeds and hyper-parameters). In the following analysis, we will use only the *cosmo\_quantile* norm since it has similar accuracy to *cosmo*.

<sup>2</sup> Long short-term memory (LSTM; Hochreiter & Schmidhuber 1997)

<sup>3</sup> Features,  $f$ , are log transformed and scaled. The log transform ( $f_l$ ) uses the minimum value of the feature in all band-passes  $\min(f)$  and a constant ( $\epsilon$ ) to centre the distribution at zero as follows:  $f_l = \log(-\min(f) + f + \epsilon)$ . Using the mean and standard deviation of the log transform ( $\mu, \sigma(f_l)$ ), standard scaling is applied:  $\hat{f} = (f_l - \mu(f_l)) / \sigma(f_l)$ .

<sup>4</sup> Both normalisation schemes are available at: <https://github.com/supernnova/SuperNNova>



**Figure 1.** SUPERNNova (SNN) classification for the DES-SN candidate DES17C2hqm at redshift  $0.473 \pm 0.001$  using three different neural networks: baseline RNN, BNN MC dropout (MC), and BNN Bayes by Backprop (BBB). All methods were trained with the TRAIN-SIM simulation. Top row shows the SN candidate light-curve from DES (normalised flux with *cosmo\_quantile* method in band-passes *g,r,i,z*; time in Observer Frame days). Bottom rows shows the classification scores for each method (SN Ia: maroon, non SN Ia: orange). Classification scores use all the data before a given date. The BNN methods provide classification uncertainties (shadowed regions show 68 per cent and 95 per cent contours). Each BNN method provides different estimations, this is explored in Section 5.2. The large uncertainties in the classification probability represent the lack of confidence in this classification. For this example, uncertainties around days 20–30 are correlated with the lower SNR, while around days 50–60 that correlation is less straight forward to interpret and could be linked to the secondary peak visible in most filters.

for the simulations but is more robust against photometry outliers in real data.

### 3.4 SNN configuration for performance and robustness

We next study the performance of SNN when classifying SNe using photometry and host galaxy redshifts. We also characterise the classification robustness with respect to the training templates, and find the best set of hyper-parameters for our DES-SNIa sample. We use

**Table 2.** Classification accuracies for models trained by replacing a subset of templates from the original configuration in Section 3.2.

Changed template	accuracy
JLA instead of extended SNIa model	$97.96 \pm 0.05$
without peculiar SNe Ia	$98.21 \pm 0.01$
J17 instead of V19 core-collapse model	$98.06 \pm 0.07$
SPCC templates instead of V19 core-collapse model	$98.59 \pm 0.02$

the S-TRAIN-SIM simulations introduced in section 3.2, for computational efficiency and to compare results with those of Möller & de Boissière (2019), to train a classification model. Our simulation was class-balanced (half normal SNe I and half non-Ia SNe) and randomly split in 80 per cent for training, 10 per cent for validation and 10 per cent for metrics evaluation. Uncertainties in the accuracy represent the standard deviation of predictions from five models obtained with different seeds.

Using the default configuration of SNN we obtain a classification accuracy of  $97.73 \pm 0.04\%$  for the *cosmo\_quantile* norm. While this accuracy is high, it is  $\sim 1\%$  lower than the benchmark in Möller & de Boissière (2019) for a similar training set size. Since the SNN architecture has not been changed, we investigate if this can be attributed to the more complex and realistic DES-SN 5-year simulations in Section 3.4.1. We then investigate whether a modified architecture can improve the classification model and thus its accuracy in Section 3.4.2. We highlight that SNN does not reach its peak performance when trained using the smaller S-TRAIN-SIMS. Thus, larger simulations are needed to improve the model performance.

#### 3.4.1 Templates impact on performance

Here, we study how the set of templates used to generate the training simulation impacts the metrics of our classification algorithm. We train different models using simulations that are similar in size (equivalent to S-TRAIN-SIM) but are generated by replacing a subset of templates from the original configuration. Obtained accuracies are shown in Table 2.

Models trained with SPCC and J17 templates obtain higher accuracies than those trained with V19 templates. This is consistent with the accuracy decrease of our present model when compared to that of Möller & de Boissière (2019). This is evidence of the more complex classification task with the updated simulations. We highlight that V19 uses a large variety of core-collapse templates with greater diversity than previous core-collapse models, J17 and SPCC. From these, SPCC has the fewest number of non-Ia templates and thus less diversity. SPCC templates were used in Möller & de Boissière (2019) simulations. The impact of changes like using the JLA SALT2 model is less. This shows that the complexity of the classification task increases largely with the updated and more diverse core-collapse SN population in the V19 templates and the inclusion of peculiar SNe Ia.

We thus attribute the decrease on accuracy to the more complex task of disentangling SNe Ia from core-collapse and peculiar SNe Ia generated with updated templates.

#### 3.4.2 Hyper-parameters

We investigate whether network hyper-parameters could be modified to improve performance (for a list of available hyper-parameters,

see Möller & de Boissière 2019). We train our models using using 20 per cent of the S-TRAIN-SIM simulations (280,033 light-curves). We modify: batch size (128, 512), dropout (0.05, 0.1, 0.2), bidirectional (True, False), hidden dimensions (32, 64, 128), number of layers (2, 3, 4), two learning policies (cyclic and non-cyclic) and different cyclic phases when using cyclic ([5, 10, 15], [20, 40, 60]). We find that the accuracy in different configurations varies up to  $\approx 2\%$ . We find that deeper (3 or 4 layers) and wider networks (up to 64 hidden dimensions) result in the biggest changes to the accuracy. This reflects the increasing complexity of the classification task with updated SN templates. Our chosen configuration for S-TRAIN-SIM is: batch size 512, dropout 0.05, bidirectional network, 64 hidden dimensions, 4 layers, and non-cyclic learning policy. Using the whole S-TRAIN-SIM dataset with this new configuration, the classification accuracy rises to  $98.10 \pm 0.06$  per cent.

### 3.5 SNN trained models for DES-SN 5-year analysis

In the following we use SNN models trained with a larger dataset to improve classification accuracy, TRAIN-SIM, and the best configuration of SNN found in the previous section. We increase the batch size to 1024 for efficient resource allocation. The larger simulation and optimised hyper-parameters provide a better classification accuracy with accuracies above 98% as shown in Table 3. Accuracies are computed with a balanced test set, where half of the candidates are SNe Ia and half are non-Ia SNe.

To evaluate the accuracy, efficiency and purity of our photometric samples, we estimate the performance of our models in the independent TEST-SIM. This simulation is not balanced and thus reflect the relative rate between SN types. We present performance metrics for different levels of selection cuts in Table 4. We highlight that we provide the balanced accuracy which shows that after the JLA-like cuts, the remaining non-Ia SNe are harder to disentangle. A thorough analysis on systematics linked to this classification method can be found in Vincenzi et al. (2021).

In this work, the traditional classification method is named "single model". This method represents classifications done using probabilities obtained from one SNN trained model with a single seed. In the following, we provide a mean value and uncertainty on the metric or classified sample of the "single model" method by taking the probabilities obtained with 5 models trained with different seeds. These probabilities are then used to compute the mean and standard deviation of the metrics listed in Table 3.

#### 3.5.1 Ensemble methods

For cosmology, we aim to have a classification method that is not highly sensitive to statistical fluctuations in the model and training dataset. In ML, ensemble methods have been shown obtain more robust predictions (Dietterich 2000; Lakshminarayanan et al. 2016) and have been introduced for regression in astronomy (Kim et al. 2015; Carrasco Kind & Brunner 2014). To produce ensemble classifications, predictions from multiple models are combined. This can be viewed as a mechanism of Bayesian marginalisation (Wilson & Izmailov 2020; Izmailov et al. 2021) and an alternative to Bayesian Neural Networks using Variational Inference explored in Section 5.

We explore two possible ensemble methods: "probability averaging" and "target averaging". Probability averaging uses the probability scores and averages them to select light-curves that are above the 0.5 probability threshold of being SN Ia. The "target average" method averages the predictions and selects the most common one.

**Table 3.** SNN Baseline Performance vs. method on TRAIN-SIM without cuts. The chosen method in this work is the Ensemble (probability average) and is highlighted in bold.

method	balanced accuracy	efficiency	purity
<i>cosmo</i>			
single model	$98.33 \pm 0.01$	$98.65 \pm 0.05$	$98.03 \pm 0.06$
ensemble (target av.)	$98.43 \pm 0.02$	$98.81 \pm 0.02$	$98.08 \pm 0.02$
ensemble (prob. av.)	$98.45 \pm 0.01$	$98.80 \pm 0.02$	$98.11 \pm 0.02$
<i>cosmo_quantile</i>			
single model	$98.35 \pm 0.01$	$98.68 \pm 0.07$	$98.03 \pm 0.05$
ensemble (target av.)	$98.45 \pm 0.00^5$	$98.84 \pm 0.02$	$98.09 \pm 0.01$
<b>Ensemble (prob. av.)</b>	<b><math>98.46 \pm 0.01</math></b>	<b><math>98.83 \pm 0.03</math></b>	<b><math>98.10 \pm 0.03</math></b>

**Table 4.** SNN Baseline Performance vs. method on TEST-SIM with loose selection and JLA-like cuts.

method	balanced accuracy	efficiency	purity
with loose selection cuts			
single model	$98.61 \pm 0.03$	$99.61 \pm 0.02$	$99.43 \pm 0.02$
ensemble (prob. av.)	$98.69 \pm 0.01$	$99.68 \pm 0.01$	$99.45 \pm 0.00^5$
+ JLA-like cuts			
single model	$98.26 \pm 0.06$	$99.81 \pm 0.01$	$99.7 \pm 0.01$
<b>Ensemble (prob. av.)</b>	<b><math>98.36 \pm 0.01</math></b>	<b><math>99.86 \pm 0.01</math></b>	<b><math>99.71 \pm 0.00^5</math></b>

Uncertainties are computed using the standard deviation of the metric for three different sets of five models with different seeds.

We find that ensemble methods increase the accuracy and purity  $\approx 0.1\%$  from just using one model prediction, or "single model", as can be seen in Table 3. We find a 99.4% overlap between photometrically selected Type Ia SNe using both the ensemble and single model methods. In the following, we will use the "probability average" from different models as our ensemble method.

Each ensemble in this work is obtained using the predictions of 5 models trained with different seeds, also called an "ensemble set". To study the performance of ensemble methods, we compute metrics using the output of 3 ensemble sets, quoting their mean and standard deviation.

#### 3.5.2 Generalisation

In this Section we verify the ability of our trained models to classify data generated using different simulation templates. This is called generalisation and showcases the adaptation of our SNN models to new unseen data.

We evaluate the accuracy of our models when trained with simulations generated using SNe Ia, peculiar SNe and the V19 core-collapse templates but applied to simulations generated using other core-collapse templates such as J17 or SPCC. We observe a decrease of  $< 0.5\%$  in accuracy, which shows that our V19 trained models generalise well to other templates of core-collapse SNe.

We find that ensemble methods such as *probability average* reduces the loss in accuracy due to changes in the data by 0.2% relative

<sup>5</sup> We provide only two-significant figures. The uncertainties are negligible and less than 0.005.

to the single model. This is expected as ensemble methods are usually more robust and thus generalise better than single models.

### 3.6 Bayesian Neural Networks (BNNs)

In scientific analyses using machine-learning outputs, it is important to evaluate the reliability of a model's predictions, expressed through uncertainties. Uncertainties can be divided into: *Aleatoric*, usually linked to measurement uncertainties (e.g. noise or other effects of data acquisition); *Epistemic* or model uncertainty, which encompasses uncertainties in the training set and NN architecture.

In this section we introduce Bayesian Neural Networks (BNNs) which are a promising method to provide uncertainties reflecting the model's confidence on the prediction.

To compute uncertainties, we obtain different classification probabilities for a given input and evaluate their variance. In NNs this is equivalent to finding a posterior distribution of weights. Typically, this posterior distribution is intractable for deep neural networks, thus different methods have been developed to approximate it. A review on BNNs, approximation methods and their use in astronomy can be found in Charnock et al. (2020).

In this Section we use two BNN implementations approximating the posterior distribution of weights: MC dropout (Gal & Ghahramani 2015) and Bayes by Backprop (Fortunato et al. 2017). MC dropout (MC in the following) provides a Bayesian interpretation by using the same dropout mask at the different NN layers including the recurrent ones (each time step). Bayes by Backprop (BBB in the following) learns a posterior distribution of weights which can then be sampled. Both methods have been previously implemented and tested on simulations in SNN (Möller & de Boissière 2019).

#### 3.6.1 BNN classification probabilities and uncertainties

For both methods, to obtain the classification probability distribution, we sample the predictions from our BNN 50 times. This sampling number is also known as the number of *inference samples*,  $n_s$ . In the following we compute the classification probability,  $P_i$  for a given light-curve,  $\mathbf{x}_i$  as the mean of sampled probabilities:

$$P_i = \frac{1}{n_s} \left( \sum_{j=1}^{n_s} p_j(\mathbf{x}_i) \right) \quad (1)$$

where  $j \in [1, n_s]$  is the index of inference samples,  $p_j(\mathbf{x}_i)$  is the  $j^{th}$  sample of the classification probability distribution for the light-curve  $\mathbf{x}_i$ .

We compute the classification probability uncertainty for a given light-curve  $\mathbf{x}_i$  as the standard deviation of sampled probabilities:

$$\hat{\sigma}_i = \frac{1}{n_s} \sqrt{\sum_{j=1}^{n_s} (p_j(\mathbf{x}_i) - P_i)^2} \quad (2)$$

where  $j \in [1, n_s]$  is the index of inference samples,  $p_j(\mathbf{x}_i)$  is a classification probability for the given light-curve  $\mathbf{x}_i$  for each inference sample  $j$ , and  $P_i$  is given by Equation 1.

#### 3.6.2 BNN trained models

Using the TRAIN-SIM simulations we train the two Bayesian models, MC and BBB, for light-curve classification with host galaxy redshifts. Both methods obtain high classification accuracies for the ensemble probability average method,  $98.33 \pm 0.01$  and  $98.11 \pm 0.01$  for

**Table 5.** Performance metrics of BNNs evaluated using TEST-SIM simulations with JLA-like cuts. These simulations are indicative of the expected purity and efficiency of our photometrically classified samples.

method	accuracy	efficiency	purity
MC with JLA-like cuts			
single model	$98.01 \pm 0.03$	$98.41 \pm 0.03$	$97.63 \pm 0.07$
ensemble (prob. av.)	$98.11 \pm 0.01$	$98.51 \pm 0.06$	$97.73 \pm 0.05$
BBB with JLA-like cuts			
single model	$98.01 \pm 0.03$	$98.41 \pm 0.03$	$97.63 \pm 0.07$
ensemble (prob. av.)	$98.11 \pm 0.01$	$98.51 \pm 0.06$	$97.73 \pm 0.05$

MC dropout and BBB respectively. Balanced accuracies are slightly lower than the ensemble method in Table 3. These may be improved by adjusting of the hyper-parameters. We choose to keep the current configuration and focus on the behaviour of the classification uncertainties.

Traditionally, BNNs are not used in ensembles, combining predictions by different models. To do so, ideally, the probability distributions for each model in the ensemble set should be concatenated into a "joint probability distribution". Then, the ensemble classification probability would be computed using Equation 1 sampling  $n_s$  times the "joint probability distribution". However, this can be computationally expensive. Using TEST-SIM simulations, we find that averaging the mean probability obtained for each model in the ensemble set is a close approximation of the one obtained using "joint probability distribution". We find that the differences between probabilities using the approximation and the "joint distribution" are centred at  $0.00 \pm 0.01$  and accuracies change by less than 0.1%. We use this approximation in the following for computational efficiency.

We also test approximating ensemble uncertainties as the sum of uncertainties from each model in the ensemble set assuming the covariance between models is zero. We find on average that the uncertainties obtained with this approximation and from the "joint probability distribution" are similar. However, we note that the approximation for the BBB method has a larger dispersion than the one for the MC method. We will evaluate the potential use of BNN classification uncertainties in Section 6.2.

We use TEST-SIM to evaluate the expected metrics for our photometrically classified samples with JLA-like cuts in Table 5. The samples obtained with BNNs have less than 3% contamination but that is higher than our Baseline DES-SNIa samples with JLA-like cuts. BNN performance could be eventually be improved with a different network configuration and initialisation. However, for comparison we keep this architecture for the analysis in Section 5.

## 4 DES-SN 5-YEAR PHOTOMETRICALLY CLASSIFIED SNE IA

In this Section, we photometrically classify DES-SN 5-year candidates with host spectroscopic redshifts using our baseline RNN trained in Section 3.5.

First, we classify candidates that pass loose cuts using SNN trained with host galaxy redshifts in Section 4.1. We further constrain the sample using JLA-like cuts and visual inspection in Section 4.2. We discuss possible contamination of this sample in Section 4.3 and its classification efficiency in Section 4.4. We summarise the properties of the baseline photometrically classified SNIa sample with JLA-like cuts in Section 4.5.

**Table 6.** DES-SNIa photometric samples with different selection cuts. In the last row, we define the **Baseline DES-SNIa sample** using a single ensemble set probability threshold. Columns indicate: the number of photometrically selected SNe Ia and the number of spectroscopically classified SNe Ia contained in that sample.

method	loose selection cuts		+JLA-like cuts	
	photo Ia	spec Ia	photo Ia	spec Ia
single model	1861 <sup>+17</sup> <sub>-13</sub>	353 <sup>+3</sup> <sub>-3</sub>	1478 <sup>+10</sup> <sub>-11</sub>	320 <sup>+3</sup> <sub>-2</sub>
ensemble (prob. av.)	1867 <sup>+3</sup> <sub>-4</sub>	354 <sup>+1</sup> <sub>-0</sub>	1482 <sup>+2</sup> <sub>-2</sub>	321 <sup>+1</sup> <sub>-0</sub>
<b>Baseline DES-SNIa sample</b>	<b>1863</b>	<b>354</b>	<b>1484</b>	<b>321</b>

#### 4.1 Photometric classification

We use our baseline RNN model to select photometrically classified SNe Ia. We show the number of selected light-curves in Table 6 and their overlap with spectroscopic SN samples defined in Section 2.

As shown in Sections 3.5.1 and 3.5.2, ensemble methods provide more robust predictions than single model methods. We select our Baseline SNe Ia sample using the “probability average” method and the *cosmo\_quantile* norm. This normalisation is more robust towards photometry outliers present in our analysis. We note that the overlap between *cosmo* and *cosmo\_quantile* probability average sample is larger than 98% and between *cosmo\_quantile* probability average and single model samples is larger than 99%.

Our Baseline DES-SNIa sample contains 1,863 photometrically identified SNe Ia passing loose selection cuts. In this sample, twelve spectroscopically classified SNe Ia are not selected, representing less than 1% of the photometric sample. We do not find a particular redshift or SALT2 parameter preference for these lost SNe Ia. Visual inspection reveals some light-curves have variable quality photometry which could contribute to the mis-classification.

The baseline sample with loose selection cuts can be used to study astrophysical properties of SNe Ia like correlations with their host galaxies, diversity and rates. In the following, we further constrain this sample with cosmology-grade cuts as in Vincenzi et al. (2021).

#### 4.2 Cuts towards a cosmology sample (JLA-like)

We further constrain our sample by applying selection cuts based on SALT2 light-curve fits and redshift quality.

First, we implement additional requirements on the fitted SALT2 parameters of the photometrically selected SNe Ia. As in Vincenzi et al. (2021), we implement the JLA-like SALT2 cuts from the Joint Light-curve Analysis (Betoule et al. 2014) introduced in Section 2.3.2. Second, we select only candidates which have a high-precision spectroscopic redshift. We eliminate those candidates that have redshifts provided by PRIMUS since the spectra are of lower-resolution, more prone to catastrophic failures and not high-quality enough for cosmology analysis.

The results of these cuts in the photometrically selected samples are shown in Table 6. We highlight that the JLA-like cuts reduce the scatter in the number of SNe, as can be seen by the reduced standard deviation in the Table when compared to the sample without JLA-like cuts. We obtain a Baseline DES-SNIa sample with JLA-like cuts of 1,484 photometrically classified SNe Ia. The missing spectroscopic SNe Ia are found to be redder in average and at all redshifts with a median around 0.5.

A summary of the selection criteria used to obtain this sample can

be found in Table 7. General properties of these samples are further studied in Section 4.5.

#### 4.3 Contamination

As shown in Vincenzi et al. (2021) and in Table 4 contamination from core-collapse and peculiar SNe in a SNN classified sample with quality cuts is expected to be less than 1%. This estimate was obtained using SN simulations containing various types of core-collapse and peculiar SNe. We inspect the Baseline DES-SNIa sample with JLA-like cuts obtained in the previous section and do not find any spectroscopically identified core-collapse or peculiar SNe. We note that spectroscopic samples are not complete and DES-SN follow-up preferentially targeted suspected Type Ia SNe.

In this section, we explore a different type of potential contaminant, “out-of-distribution” candidates such as AGNs and other unknown transients. These candidates can be erroneously classified since they are not present in the simulated training sample and thus we do not know how SNN classifies them.

We find no spectroscopically identified AGN, SLSNe or other SN spectral types in our Baseline DES-SNIa sample but 5 candidates with host spectra showing AGN features. We find that DES16E2nb, DES16X1ext, DES13X3dbe are displaced by more than 1'' from the centre of the galaxy (additionally DES16E2nb is a spectroscopic Type Ia SN) and the other two candidates are displaced between 0.5'' and 1''. At these separations, the light-curves from these candidates are not dominated by the AGN which we confirm by inspection of the light-curves. Therefore we keep these photometrically selected SNe Ia in our Baseline DES-SNIa sample.

We also perform visual inspection of the light-curves in the Baseline DES sample. We find 3 candidates that can be visually tagged as multi-season visually: DES16E2nb a spectroscopic SN Ia with close by AGN, DES16C3nd two SN Ia in a galaxy (Scolnic et al. 2020), DES14E2rpm a spectroscopic SN Ia with a fake SN inserted at the same coordinates (fakes were inserted to evaluate the detection efficiency in DES-SN images, see Brout et al. 2019). We keep all these candidates since they are real supernovae with fake or other SN light-curves that do not overlap.

Photometrically classified Type Ia SNe samples are expected to have some level of contamination from core-collapse and peculiar SNe and possibly by other transients. For the Baseline DES-SNIa sample in this work we find no clear evidence of contamination from core-collapse and peculiar SNe or long-term variables such as AGNs.

#### 4.4 Classification efficiency

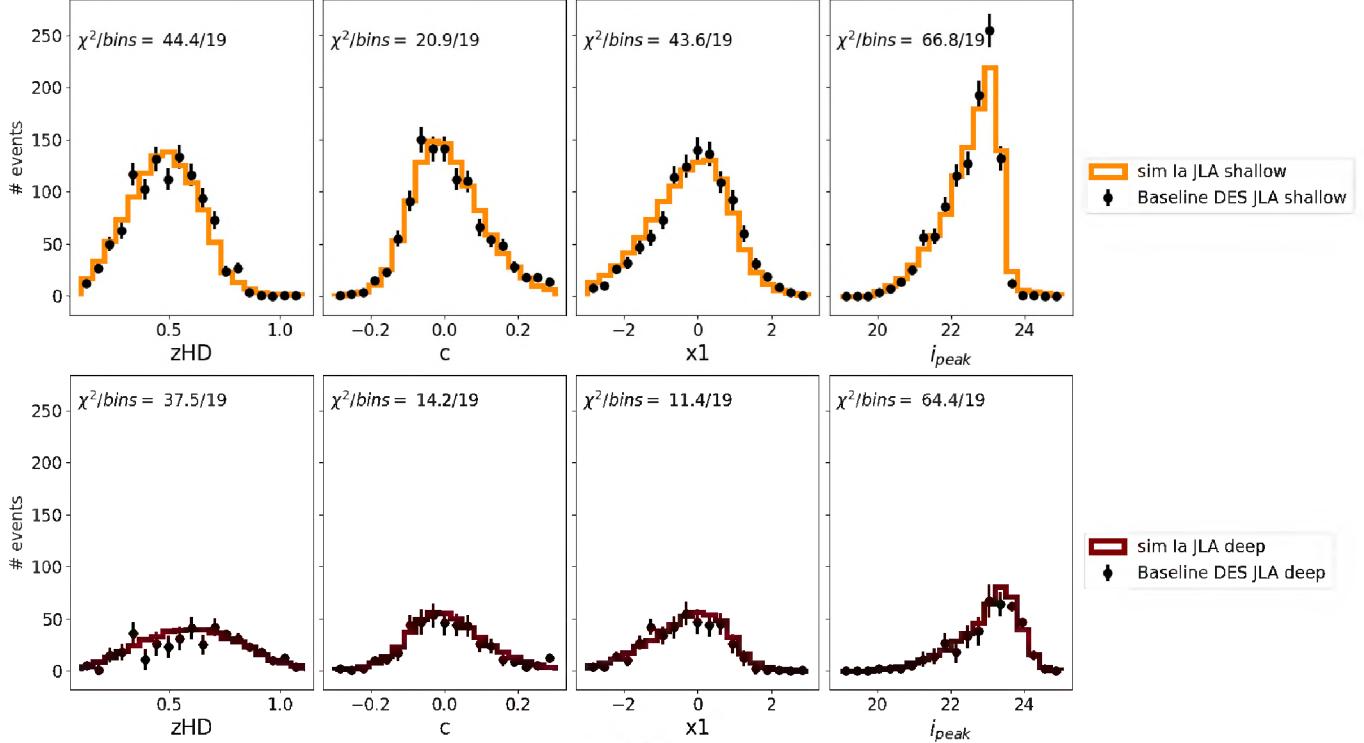
Traditionally, in cosmology analyses using spectroscopically classified SNe samples, modelling selection effects is crucial to estimate biases and systematic uncertainties.

Selection effects arise from a combination of SN detection and other effects. They are usually modelled as an efficiency with respect of an observed magnitude. For host galaxy selection, Vincenzi et al. (2020) uses the host galaxy  $r$  band magnitude,  $m_r^{\text{host}}$ . For spectroscopic classification, Smith et al. (2020); Kessler et al. (2019b) use the modelled supernova peak magnitude in the  $i$  band,  $i_{\text{peak}}$  computed from the best-fit SALT2.

To determine if there is a selection efficiency decrease due to photometric classification, we inspect the differences between the peak observed magnitude in the  $i$  band of our Baseline DES-SNIa sample compared to simulated SNe Ia in DES-SN 5-year in Figure 2. Our Baseline DES-SNIa photometric sample follows the expected

**Table 7.** Effect of the selection cuts on the candidate sample. We show results for the shallow and deep fields, as well as the total number. Note that some events belong to both shallow and deep fields due to field overlap. Columns show the cut, the number of selected candidates, the number of spectroscopic SN Ia in the sample and the Section where the sample is described.

cut	shallow		deep		total		section
	selected	spec Ia	selected	spec Ia	selected	spec Ia	
DES-SN 5-year candidate sample	29203	415	7500	93	31636	415	2.1
Multi-season	13868	405	4428	88	14070	405	2.2
Redshifts in $0.05 < z < 1.3$	6556	401	1812	85	6590	401	2.3.1
SALT2 loose selection	2380	366	698	77	2381	366	2.3.1
RNN>0.5 ( <b>Baseline DES-SNIa</b> )	1863	354	502	76	1863	354	4.1
JLA-like ( <b>Baseline DES-SNIa JLA</b> )	1484	321	408	73	1484	321	4.2



**Figure 2.** Distributions of redshift, SALT2  $x1$ , SALT2  $c$  and peak magnitude in  $i$ -band  $i_{\text{peak}}$  for our Baseline DES-SNIa sample from Section 4 for the shallow (yellow) and deep (maroon) fields

We show one simulated realisation of DES-SN 5-year sample. Poisson uncertainties are assumed. Both the simulation and data pass JLA-like cuts. The goodness-of-fit for each histogram is shown as the  $\chi^2/\text{number of bins}$  on each plot.

SN Ia peak magnitude distribution from simulations but we find an excess on the maximum magnitude with a reduced  $\chi^2_{\nu} = 2.1$ . We do not find evidence for additional selection efficiency effects from the photometric classification procedure.

#### 4.5 Colour and stretch evolution

We study the properties of the Baseline DES-SNIa sample with JLA-like cuts and compare it to that expected from realistic simulations. In Section 4.4 we found that the effects of classification efficiency are negligible, thus we don't correct for this efficiency and use simulations including only detection and host galaxy redshift efficiency introduced in Section 3.2.

Figure 2 shows the redshift  $z_{\text{HD}}$  and SALT2 fitted colour  $c$ , stretch  $x_1$  and  $i_{\text{peak}}$  distributions for the DES-SNIa 5-year photometric sample classified using host galaxy redshifts. Figure 2 also shows one realisation of a DES-SN 5-year simulated SNe Ia. Uncertainties are

calculated as the square-root of the number of candidates per bin. There is decent agreement between the simulation and data, although the reduced  $\chi^2_{\nu}$  are somewhat larger than expected from statistical fluctuations.

In Figure 3 we show the redshift evolution of our sample's colour and stretch. Our baseline sample matches the trends expected from the simulation. Although there are some slight differences outside the 68% simulation contour (equivalent to  $1\sigma$  for a Gaussian distribution) in particular for the shallow fields.

These differences might result from the small number of candidates (the last two redshift bins have only 24 and 16 SNe Ia), unaccounted classification contamination, unaccounted selection effects or whether there is redshift evolution in the intrinsic SN population (Scolnic & Kessler 2016; Popovic et al. 2021; Nicolas et al. 2021) or the effect of dust needs to be introduced (Jha et al. 2007; Mandel et al. 2011, 2017; Brout & Scolnic 2021). The optimisation of the simulation and systematics studies is outside the scope of this work.

**Table 8.** Photometric classification of light-curves with Bayesian Neural Networks. Columns indicate: the number of photometrically selected events and the number of spectroscopic SNe Ia contained in that sample. We show these samples with JLA-like SALT2 cuts as in Section 4.2 and when adding a cut in the BNN classification uncertainty.

method	+JLA-like		+JLA-like +unc	
	photo Ia	spec Ia	photo Ia	spec Ia
MC dropout				
single model	1532 <sup>+7</sup> <sub>-4</sub>	335 <sup>+1</sup> <sub>-1</sub>	1513 <sup>+6</sup> <sub>-3</sub>	333 <sup>+0</sup> <sub>-0</sub>
ensemble (prob. av.)	1535 <sup>+3</sup> <sub>-2</sub>	336 <sup>+0</sup> <sub>-0</sub>	1520 <sup>+2</sup> <sub>-1</sub>	333 <sup>+0</sup> <sub>-0</sub>
<b>Baseline MC sample</b>	1535	336	1520	333
BBB				
single model	1526 <sup>+8</sup> <sub>-6</sub>	334 <sup>+1</sup> <sub>-0</sub>	1487 <sup>+5</sup> <sub>-2</sub>	328 <sup>+2</sup> <sub>-0</sub>
ensemble (prob. av.)	1528 <sup>+1</sup> <sub>-1</sub>	335 <sup>+1</sup> <sub>-0</sub>	1483 <sup>+0</sup> <sub>-0</sub>	324 <sup>+1</sup> <sub>-0</sub>
<b>Baseline BBB sample</b>	1529	336	1483	324

We now turn to select other photometric samples using the novel Bayesian Neural Networks and explore their possible use.

## 5 PHOTOMETRICALLY CLASSIFIED SNE IA WITH BAYESIAN NEURAL NETWORKS

In this section we explore the use of Bayesian Neural Networks (BNNs) for classification. While the accuracy of these Networks is equivalent to the baseline RNN used in Section 4, BNNs also provide classification uncertainties.

We first obtain photometric samples using two BNN schemes (MC and BBB, Section 5.1). We then evaluate the classification uncertainties from BNNs (Section 5.2), and summarise our findings (Section 5.3).

### 5.1 BNN photometric sample

We apply our BNN trained models to candidates passing loose and JLA-like cuts introduced Sections 2.3.1 and 2.3.2. This candidate sample contains 1,701 light-curves that are then photometrically classified.

Using BNN probabilities, the average probability ensemble method and a threshold of  $P$  larger than 0.5, we obtain about 3% more candidates than our Baseline DES-SNIa sample with JLA-like cuts in Table 6 for both BNN methods. The additional BNN selected supernovae, 52 MC and 51 BBB, have distributions of colour, stretch and redshifts that are representative of the Baseline DES-SNIa sample selected using the RNN models (Section 4). We find that 1 and 6 SNe Ia in the Baseline DES-SNIa sample are not selected by MC and BBB methods. These missing SNe Ia have red colours and are at median redshifts close to 0.5. The BNN samples are thus probing a similar parameter space to the Baseline DES-SNIa sample.

As in the previous sample, we find no spectroscopically identified AGN, SLSNe or other SN spectral types in our BNN photometric sample. We find the same 5 candidates with nearby spectra showing AGN features which are kept due to their large enough separation  $> 0.5''$ , with the AGN. In a cosmological sample however, these candidates will be eliminated due to possible issues with the measured photometry.

### 5.2 BNN uncertainties

In this Section we try to interpret which types of uncertainties are captured in the outputs of the BNN model: *aleatoric* or *epistemic*. BNNs provide classification probability distributions that a priori indicate a confidence level on the prediction. These uncertainties are shown in Figure 1 for each classification step. Here we only evaluate the final uncertainty (final time step) for each event.

In Figure 4 we show the distribution of classification uncertainties for different samples. We compare the uncertainties derived from the data and from simulations. For most samples, the simulation and data uncertainty distributions are similar. This indicates that the simulations and data resemble closely after JLA-like cuts. However, a large difference is found where there is no selection cut which is further explored in Section 6.2.

Both BNN methods provide different order of magnitude of uncertainties estimates and distribution of mean uncertainties (e.g. BBB is more clustered in low uncertainty regions), possibly due to initialisation parameters or intrinsic properties of the method. Accounting for those differences is not straight-forward, see Möller & de Boissière (2019) for a discussion on this topic.

We compare BNN uncertainties as a function of light-curves properties in Figure 5. We find that MC dropout and BBB exhibit different behaviours for both data and simulations.

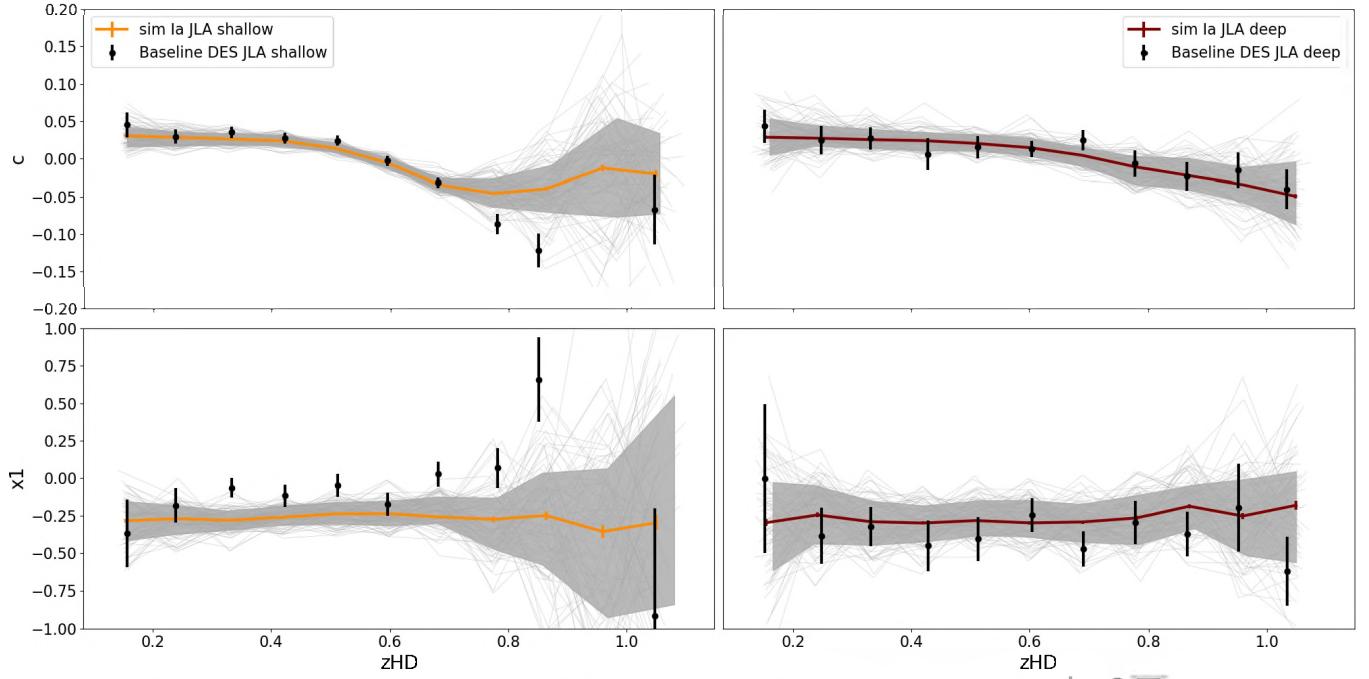
We find both indications in favour (+) and against (-) interpretation of classification uncertainties as a particular type:

a. *aleatoric uncertainty*: linked to measurement uncertainties  
 (+) classification uncertainties are correlated to SNR in data. Bright candidates and those with higher quality light-curves have on average smaller classification uncertainties for both BNNs.  
 (-) this correlation is not seen in the simulations for any of the BNNs.

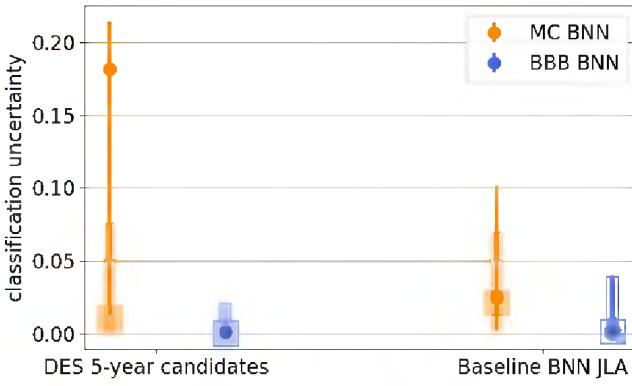
b. *epistemic uncertainty*: linked to training sets or model  
 (+) Large uncertainties are more prevalent in classification probabilities far from 1 (high probability of being a SN Ia) and 0 (low probability of being SN Ia) for both simulations and DES-SN 5-year data.  
 (-) candidates that fulfil selection cuts should more closely resemble simulated SNe Ia, thus it is puzzling the increase on median uncertainty when applying cuts in particular for the MC method (see Figure 4).

These various behaviours highlights the challenges on quantifying uncertainties in complex problems such as astronomical data classification. In Appendix A we explore further correlations between classification uncertainties and SALT2 fit light-curve properties.

We continue exploring the interpretability of the BNNs uncertainties by adding a threshold on the uncertainties for SNIa sample selection, as in Möller & de Boissière (2019) and more recently in Butter et al. (2021). We note that establishing a threshold for uncertainties is not straight-forward. While the whole probability distribution has a calibration that can be verified using diagnostic as reliability diagrams (DeGroot & Fienberg 1983; Möller & de Boissière 2019), the probability uncertainties do not. We chose to eliminate candidates with the highest uncertainties (eliminating candidates that are outside of 99 percentile of the uncertainty distribution). This cut rejects candidates that were in the RNN sample: 12 for the MC model and 45 for BBB. These candidates are not found to be distributed preferentially in a  $c$ ,  $x_1$  or redshift. We visually inspect these light-curves and found that a large proportion have photometry that are outliers.



**Figure 3.** Redshift dependence of SALT2  $c$  and  $x_1$  for the Baseline DES-SNIa photometric sample and simulated SNe Ia for shallow (yellow, left) and deep (right, maroon) fields using the DES-SN host galaxy spectroscopic efficiency (Vincenzi et al. 2020) both with JLA-like cuts. For the simulation, orange lines are rolling averages of the measured parameters, in grey 150 realisations of SNe Ia in the DES-SN 5-year survey and in solid grey the area covered by the 68% of these realisations. The mean and the standard deviation are shown for data using black markers.



**Figure 4.** Classification uncertainties obtained for BNN ensemble models. Columns indicate which sample is used. For each event in a given sample, we obtain their classification uncertainties from the two BNN methods, MC and BBB (orange and blue respectively). We show median uncertainties for data in circles for: all DES-SN 5-year data (no selection cuts), and Baseline BNN SNIa samples with JLA-like cuts. For comparison, we show in squares the median uncertainties obtained for the whole simulation (first column) and simulated photometric samples with JLA-like cuts (second column). For both the data and simulations, we show as errorbars the extent of the 68% of the distribution. The different behaviour of simulated MC uncertainties and that of DES-SN 5-year candidate sample is further studied in Figure 7.

### 5.3 BNN photometric sample contribution

The SNIa samples obtained using BNN methods are found to be similar to the one provided by our Baseline DES-SNIa sample in Section 4. We evaluate BNN uncertainties and show that they are

consistent between simulations and data in average after JLA-like cuts, showing a good agreement between data and simulation predictions. However, BNN uncertainties are difficult to interpret and assess quantitatively (e.g. assigning an uncertainty threshold).

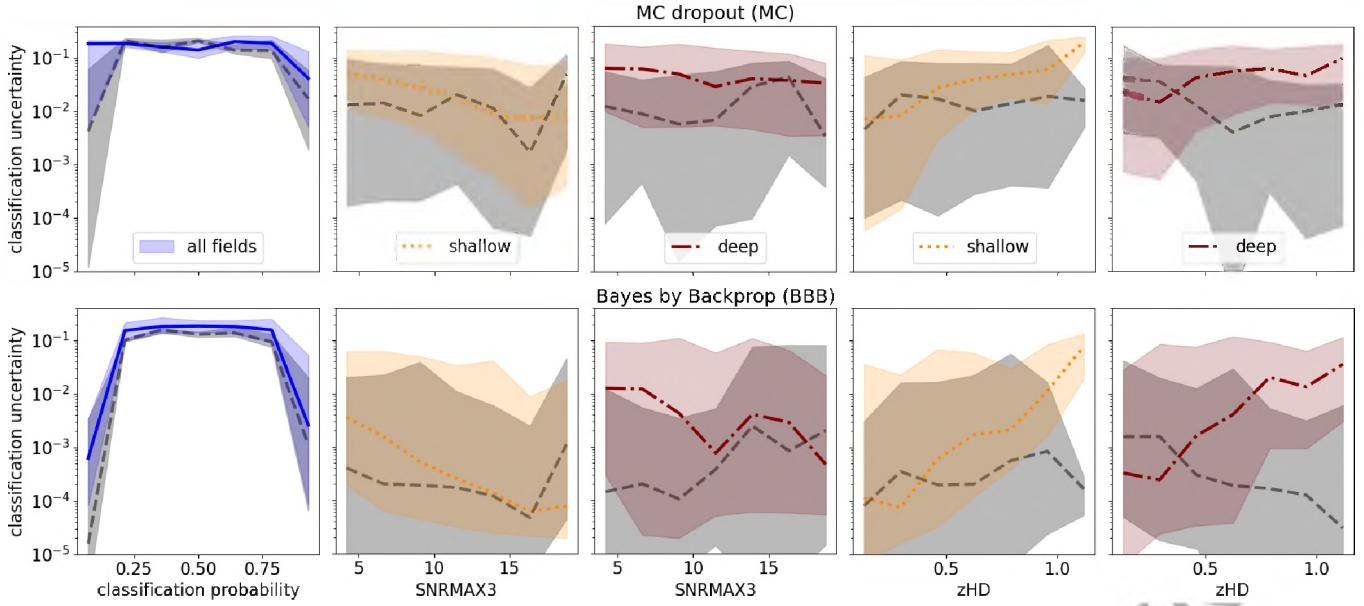
We find that uncertainties exhibit different behaviours in the two BNN methods and between data and simulations. While the higher uncertainties in the MC BNN method for the data could point towards the presence of out-of-distribution candidates, the evidence is not conclusive and is not seen in the BBB method. We will further explore the possible contribution of BNNs in photometric classification without any selection cuts in Section 6.2.

Cuts on uncertainty values potentially improve our photometric SNIa samples by rejecting candidates with photometry that contains outliers. This is a promising avenue shown to improve the quality of samples, both in quality of the data and rejection of out-of-distribution events, in previous work using simulations Möller & de Boissière (2019) and more recently with astronomical data in Butter et al. (2021).

## 6 FROM DES TO RUBIN OBSERVATORY LSST

For the LSST survey, where up  $10^7$  SNe will be detected over 10 years, photometric classification will become increasingly important.

In this work, we have presented different methods for photometric classification with redshift information. We compare the samples obtained with these different methods in Section 6.1 and explore possible applications of Bayesian Neural Networks in future surveys, such as LSST, in Section 6.2.



**Figure 5.** Distribution of classification uncertainty for Baseline MC dropout (upper row) and Bayes by Backprop technique (lower row). We show uncertainties as a function of classification probability for all fields (left), SNR of the third brightest point in the light-curve (SNRMAX3, columns 2 and 3), and redshift (zHD, columns 4 and 5). Coloured lines show the median of the data with solid blue representing all fields, dotted yellow representing shallow fields, and dot-dash red representing the deep fields. Simulations are shown by the grey dashed lines. Shaded regions show the 68% percentile.

## 6.1 DES-SNIa photometric samples

The DES-SN 5-year data contains thousands of potential SNe Ia. We show in Table 7 the different steps used in this work to obtain our Baseline DES-SNIa JLA sample from the DES-SN 5-year candidate sample. Cuts applied before photometric classification reduce the candidate sample by 90%. Photometric classification and JLA-like cuts refine the sample with a small 20% reduction. While this reduction is small, it reduces contamination from  $\sim 10\%$  to below 1.4%, as shown in (Vincenzi et al. 2021) and in Section 4.

In addition to our Baseline DES-SNIa sample classified using RNN probabilities, we have explored identifying samples with Bayesian Neural Networks. We compare these samples with the preliminary DES-SN 5-year spectroscopically classified SNe Ia sample in Figure 6. As expected, we find that photometric samples using RNNs or BNNs provide larger numbers of SNe Ia than the spectroscopic sample, probing a larger parameter space. We do not find a substantial difference in the parameter distributions between different photometric classification methods.

We highlight that the photometric samples peak at fainter magnitudes and higher redshifts than the preliminary DES-SN 5-year spectroscopic SNe Ia sample. This has the potential to reduce selection biases and opens the possibility of stronger statistical analyses with the large numbers of SNe Ia. This will also be true for the immense SN samples obtained with LSST.

## 6.2 Bayesian Neural Networks as a proxy

Introduced as a promising method to quantify model uncertainties, BNNs have not yet been widely used in classification tasks. In Section 5, we have shown the difficulties for uncertainty interpretation given the different uncertainty values for the BNN methods. However, a potential use could be rejecting candidates with large uncertainties, as they sometimes have light-curves with photometry outliers.

Here, we explore other possible uses of BNN uncertainties, using

samples that have not been constrained with selection cuts. We aim to answer two questions: (i) can BNN uncertainties be used as an indicator of the representativity of the training set for a given dataset? (ii) can BNN uncertainties replace selection cuts? We address these questions in Sections 6.2.1 and 6.2.2 respectively. The former could be useful to choose the set of SED templates to simulate a survey. As some selection cuts require feature extraction, the latter could be valuable to avoid this time-consuming process by using instead classification uncertainties from non-parametric classifiers as SNN.

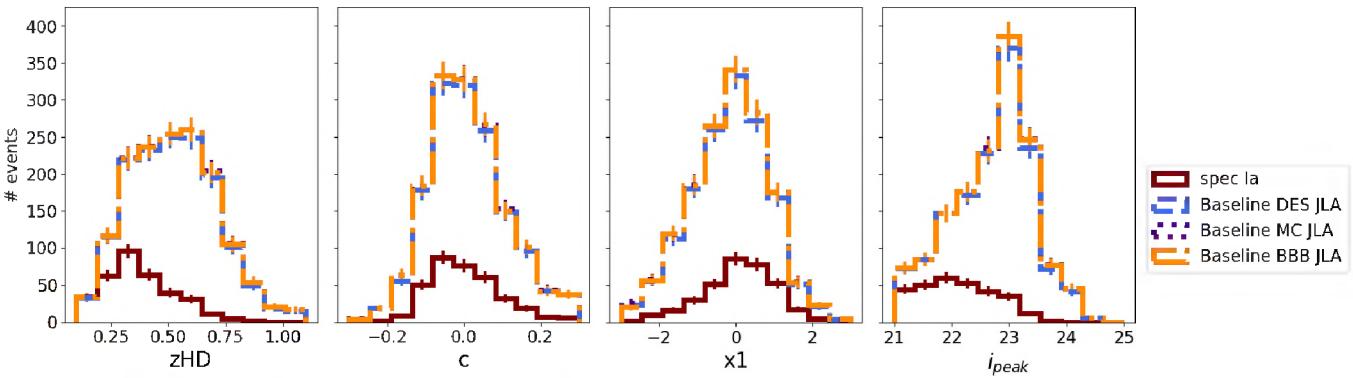
### 6.2.1 BNNs uncertainties vs. simulation representativity

First, we use simulations to assess the expected behaviour of uncertainties when training sets are not representative of the testing data.

We examine how the uncertainties change when using the trained model in Section 3.6.2 and applied to individual simulations with normal Type Ia supernovae and core-collapse SNe generated with the V19, SPCC and J17 templates. We expect that the trained model is representative of the V19 simulation. This will not be true for J17 and SPCC.

We find that both the single seed and ensemble methods have accuracies which decrease for J17 and SPCC simulations by  $\approx 0.5\%$  for both types of BNNs. We see an increase in the *mean model uncertainty* on classified light-curves generated with J17 and SPCC, however this change is within uncertainties. For both BNNs we find a longer and more significant tail for the uncertainty distributions when classifying J17 and SPCC simulations (ending at  $\sim 0.4 - 0.43$  compared to  $\sim 0.35$  for V19).

Next, we compare uncertainties when classifying DES-SN 5-year data with independent BNN models trained with the V19, J17 and SPCC simulations. We find that the *mean model uncertainty* increases for SPCC and J17 classification models for MC dropout but not for BBB SPCC model but again within uncertainties. The tail of the uncertainties varies between  $\sim 0.40 - 0.47$  for all classification



**Figure 6.** Distributions of redshift, SALT2  $c$ ,  $x1$  and peak magnitude in the  $i$  band  $i_{\text{peak}}$ , for the samples with JLA cuts: preliminary DES-SN 5-year SNIa spectroscopically classified (maroon), Baseline DES-SNIa (RNN) (blue) and Baseline MC and BBB SNIa samples, purple and orange, respectively. We note that the MC and BBB samples distributions almost completely overlap one other.

models. We see a longer tail for the uncertainty distributions for BBB but not for MC SPCC classification.

In summary, we do not find strong evidence of BNN uncertainties being sensitive to models trained with different core-collapse templates. There is a small but inconclusive tendency to increase uncertainties for J17 and SPCC in simulations. While these templates are different, the changes may be too small to be captured by BNN uncertainties.

#### 6.2.2 BNN uncertainties as a proxy for selection cuts?

We further study the distribution of classification uncertainties for samples selected with different cuts.

First, we check the behaviour of uncertainties with simulations. Uncertainties are distributed with a peak at low values and a decreasing long tail. We find that as the sample is refined through cuts in redshift, SALT2 convergence, and others, the maximum uncertainty is reduced. For example, if the simulated sample passes loose selection cuts and then a JLA-like cut is applied, the maximum uncertainty in the distribution reduces from 0.37 to 0.26 in MC dropout and from 0.34 to 0.25 in BBB. We do not find a significant change in the median distribution since it is dominated by small uncertainty values.

For the DES-SN 5-year data we show the distribution of classification uncertainties in Figure 7 with different selection cuts (see Section 2.3.1). As selection cuts are applied, the maximum uncertainties reduces for both methods as in simulations.

We highlight an interesting behaviour seen for MC dropout classification uncertainties. We find that this method assigns high uncertainties to candidates that do not have a secured redshift and candidates that are filtered with the multi-season cut. While the model was trained to use host galaxy redshifts, it can provide a classification for objects using a default value provided, here an assigned redshift of  $-9$ . While these candidates are clearly outliers (the redshift provided for classification is  $-9$ ) and can be eliminated using simple cuts, this could indicate that MC dropout uncertainties are indicative of out-of-distribution candidates. Importantly, many of these high-uncertainty candidates are classified with probability larger than 0.5 which, without selection cuts, would end up in our photometric sample if no selection cuts were applied. We do not see this behaviour in the BBB model.

The multi-season veto and redshift availability cut effectively eliminates the light-curves producing the high-uncertainty peak for MC

dropout. After these cuts, the most impactful cut for higher uncertainties is linked to the SALT and JLA quality cuts. This is not surprising since these cuts restrict the SN properties range to the ones for normal SNe Ia.

In summary, we find that BNN methods behave differently when classifying out-of-distribution candidates defined as light-curves without redshift. Interestingly, the high-uncertainty peak found for the MC dropout method in Figure 7 reflects a possible interpretability of these uncertainties. This interpretability could help to quickly identify the presence of anomalies in the dataset which were not in the training sets of the model.

For current surveys, our candidate samples are small enough to easily identify out-of-distribution events using feature distributions. However, for future surveys such as Rubin LSST this may prove difficult given the expected detection of 10 million transient candidates per night. Here we find that BNN uncertainties from MC dropout scheme can provide an indication whether there are out-of-distribution events in a given candidate sample and further selection cuts may be required.

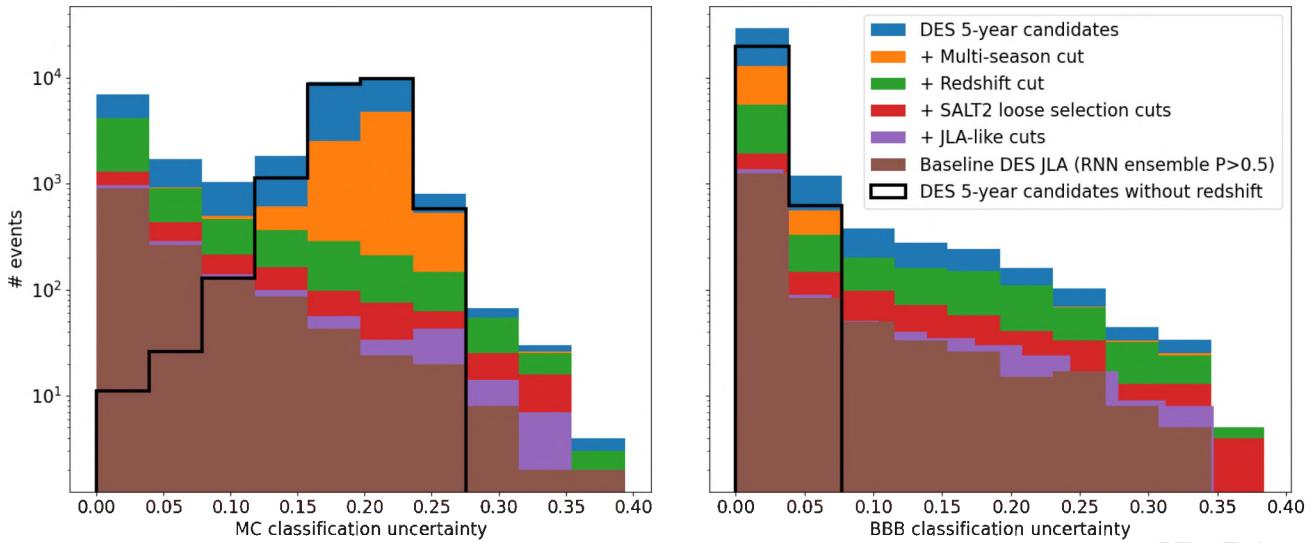
## 7 CONCLUSIONS

In this work we train Type Ia vs. non Ia classification models using large realistic DES-like simulations and apply them to DES-SN 5-year data.

We introduce pre-processing of DES-SN light-curves for accurate photometric classification. This includes selection of light-curve time-span, photometry quality cuts and selection cuts to limit out-of-distribution candidates that are not included in the training set (e.g. AGNs).

We present samples classified with host galaxy redshifts using SNN Recurrent Neural Networks and explore the use of Bayesian Neural Networks. We introduce the use of ensemble predictions for SN classification. We find that selecting SNe using an ensemble of models is more robust and stable than any single model.

Using host galaxy spectroscopic redshifts, we select a Baseline DES-SNIa sample of 1,863 photometrically identified Type Ia SNe. This sample can be used for astrophysical studies of the properties of SNe Ia and their environments. For cosmology, we apply JLA-like cuts and select 1,484 photometrically classified SNe Ia. This sample is more than three times larger than the DES-SN 5-year spectroscopically confirmed SNIa sample and covers a larger redshift range. Most of the spectroscopically identified SNe Ia in DES-SN



**Figure 7.** Uncertainties obtained with the two BNN methods (MC and BBB) for the DES-SN 5-year candidate sample, through different selection cuts: multi-season filtering, redshifts, SALT2 convergence and JLA-like cuts; and our photometrically identified sample (filled histograms). We show the number of events in the y-axis in log scale. MC dropout uncertainties seem to identify those out-of-distribution candidates that have no redshift information (black line) or are filtered multi-season events. This secondary peak drives the mean uncertainty behaviour for MC dropout in Figure 4.

are included in this photometric sample. These 1,484 photometrically identified SNe Ia are currently the largest single-survey high-quality SN Ia sample and is being used for studies such as rates and SNe Ia host-galaxy properties.

We find that the properties of the SNe Ia in our Baseline DES-Ia sample are reproduced in the simulations. We anticipate that with further refinements (improved host galaxy libraries and more accurate dust models), the agreement between the simulations and the data will improve.

Additionally, we explore the use of uncertainties provided by Bayesian Neural Networks for identifying out-of-distribution candidates and defining representative training sets. We highlight some of the BNN pitfalls and the difficulty of comparing classification uncertainties between variational inference methods. We find that the MC dropout BNN provides potentially interpretable uncertainties for out-of-distribution event detection and improving the photometric sample. This work is the first known application of two BNN methods on real astrophysical data for classification tasks.

This work is part of the DES-SN 5-year cosmology analysis. We have optimised simulations, the SNN architecture, as well as developed data pre-processing methods. These methods are a revision from those presented in Vincenzi et al. (2021) where contamination is found to be less than 1.4% for photometrically classified samples. We find that photometric quality is key for robust classification, and an improved sample can be expected from using high-quality Scene Modelling Photometry (Brout et al. 2019).

For future surveys such as LSST, photometric classification will be key to fully harness the power of these surveys. Photometric classification with host redshift information will enable using large, low-contamination, high-quality samples for measuring cosmological parameters. Potentially, MC BNN could provide useful information to filter transient samples in large surveys. Extensions to this work include photometric classification without redshift, which will assist in the allocation of follow-up resources for host galaxy redshift acquisition (such as Time-Domain Extragalactic Survey TiDES; Frohmaier et al. in prep, Swann et al. 2019) and for other astrophysical studies.

## ACKNOWLEDGEMENTS

AM thanks of the University of Queensland for hospitality during early work on this project; parts of this research were supported by the Australian Research Council Australian Laureate Fellowship FL180100168.

This paper has gone through internal review by the DES collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NFS’s NOIRLab, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the Uni-

versity of Sussex, Texas A&M University, and the OzDES Member-  
ship Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory at NSF's NOIRLab (NOIRLab Prop. ID 2012B-0001; PI: J. Frieman), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grants ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciéncia e Tecnologia (INCT) do e-Universo (CNPq grant 465376/2014-2).

This work was completed in part with Midway resources provided by the University of Chicago's Research Computing Center.

This work makes use of data acquired at the Anglo-Australian Telescope, under program A/2013B/012. We acknowledge the traditional owners of the land on which the AAT stands, the Gamilaraay people, and pay our respects to elders past and present.

MS is funded by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation program (grant agreement no 759194 - USNAC). L.G. acknowledges financial support from the Spanish Ministerio de Ciencia e Innovación (MCIN), the Agencia Estatal de Investigación (AEI) 10.13039/501100011033, and the European Social Fund (ESF) "Investing in your future" under the 2019 Ramón y Cajal program RYC2019-027683-I and the PID2020-115253GA-I00 HOST-FLOWS project, and from Centro Superior de Investigaciones Científicas (CSIC) under the PIE project 20215AT016. LK thanks the UKRI Future Leaders Fellowship for support through the grant MR/T01881X/1.

## DATA AVAILABILITY

We provide in [https://github.com/analismoller/DES5YR\\_SNIA\\_host](https://github.com/analismoller/DES5YR_SNIA_host): (i) the SNANA and/or Pippin configuration files to reproduce simulations in this paper, (ii) configuration files and scripts to re-train SNN classification models (SNN is an open source framework available in GitHub), and (iii) analysis code in python to reproduce plots and results. Sample classification probabilities are available in Zenodo <https://doi.org/10.5281/zenodo.5904368>.

## REFERENCES

Abbott T. M. C., et al., 2018, *ApJS*, 239, 18  
 Angus C. R., et al., 2019, *MNRAS*, 487, 2215  
 Astier P., et al., 2006, *Astron. Astrophys.*, 447, 31  
 Bernstein J. P., et al., 2012, *Astrophys. J.*, 753, 152  
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393  
 Betoule M., et al., 2014, *A&A*, 568, A22  
 Brout D., Scolnic D., 2021, *ApJ*, 909, 26  
 Brout D., et al., 2019, *ApJ*, 874, 106  
 Butter A., Finke T., Keil F., Krämer M., Manconi S., 2021, arXiv e-prints, p. [arXiv:2112.01403](https://arxiv.org/abs/2112.01403)  
 Caldeira J., Nord B., 2020, arXiv e-prints, p. [arXiv:2004.10710](https://arxiv.org/abs/2004.10710)  
 Campbell H., et al., 2013, *The Astrophysical Journal*, 763, 88  
 Carrasco Kind M., Brunner R. J., 2014, *MNRAS*, 442, 3380  
 Charnock T., Perreault-Levasseur L., Lanusse F., 2020, arXiv e-prints, p. [arXiv:2006.01490](https://arxiv.org/abs/2006.01490)  
 Childress M. J., et al., 2017, *MNRAS*, 472, 273  
 Coil A. L., et al., 2011, *ApJ*, 741, 8  
 Contreras C., et al., 2010, *AJ*, 139, 519  
 Dai M., Kuhlmann S., Wang Y., Kovacs E., 2018, *MNRAS*, 477, 4142  
 Dark Energy Survey 2019, *ApJ*, 872, L30  
 DeGroot M. H., Fienberg S. E., 1983, *Journal of the Royal Statistical Society. Series D (The Statistician)*, 32, 12  
 Dietterich T. G., 2000, in *Multiple Classifier Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 1–15  
 Flaugher B., et al., 2015, *AJ*, 150, 150  
 Foley R. J., et al., 2018, *MNRAS*, 475, 193  
 Fortunato M., Blundell C., Vinyals O., 2017, arXiv e-prints, p. [arXiv:1704.02798](https://arxiv.org/abs/1704.02798)  
 Frieman J. A., et al., 2008, *AJ*, 135, 338  
 Frohmaier C., et al., 2019, *MNRAS*, 486, 2308  
 Gal Y., Ghahramani Z., 2015, arXiv e-prints, p. [arXiv:1506.02142](https://arxiv.org/abs/1506.02142)  
 Goldstein D. A., et al., 2015, *AJ*, 150, 82  
 Grayling M., et al., 2021, *MNRAS*, 505, 3950  
 Gupta R. R., et al., 2016, *AJ*, 152, 154  
 Gutiérrez C. P., et al., 2020, *MNRAS*, 496, 95  
 Guy J., et al., 2007, *A&A*, 466, 11  
 Guy J., et al., 2010, *A&A*, 523, A7  
 Hicken M., et al., 2009, *ApJ*, 700, 551  
 Hicken M., et al., 2012, *ApJS*, 200, 12  
 Hinton S., Brou D., 2020, *The Journal of Open Source Software*, 5, 2122  
 Hlozek R., et al., 2012, *ApJ*, 752, 79  
 Hochreiter S., Schmidhuber J., 1997, *Neural Comput.*, 9, 1735  
 Inserra C., et al., 2021, *MNRAS*, 504, 2535  
 Izmailov P., Vikram S., Hoffman M. D., Wilson A. G., 2021, arXiv e-prints, p. [arXiv:2104.14421](https://arxiv.org/abs/2104.14421)  
 Jha S. W., 2017, in Alsabti A. W., Murdin P., eds., *Handbook of Supernovae*, p. 375. doi:10.1007/978-3-319-21846-5\_42  
 Jha S., Riess A. G., Kirshner R. P., 2007, *ApJ*, 659, 122  
 Jones D. O., et al., 2017, *ApJ*, 843, 6  
 Jones D. O., et al., 2018, *ApJ*, 857, 51  
 Kelsey L., et al., 2021, *MNRAS*, 501, 4861  
 Kessler R., Scolnic D., 2017, *ApJ*, 836, 56  
 Kessler R., et al., 2009, *PASP*, 121, 1028  
 Kessler R., et al., 2010, *PASP*, 122, 1415  
 Kessler R., et al., 2013, *ApJ*, 764, 48  
 Kessler R., et al., 2015, *AJ*, 150, 172  
 Kessler R., et al., 2019a, *PASP*, 131, 094501  
 Kessler R., et al., 2019b, *MNRAS*, 485, 1171  
 Kim E. J., Brunner R. J., Carrasco Kind M., 2015, *MNRAS*, 453, 507  
 Kunz M., Bassett B. A., Hlozek R. A., 2007, *Phys. Rev. D*, 75, 103508  
 LSST Science Collaboration Abell P. A., Allison J., Anderson 2009, preprint, p. [arXiv:0912.0201](https://arxiv.org/abs/0912.0201)  
 Lakshminarayanan B., Pritzel A., Blundell C., 2016, *NIPS* 6393-6395, p. [arXiv:1612.01474](https://arxiv.org/abs/1612.01474)  
 Li W., et al., 2011, *MNRAS*, 412, 1441  
 Lidman C., et al., 2020, *MNRAS*, 496, 19  
 Madau P., Weisz D. R., Conroy C., 2014, *ApJ*, 790, L17  
 Mandel K. S., Narayan G., Kirshner R. P., 2011, *ApJ*, 731, 120  
 Mandel K. S., Scolnic D. M., Shariff H., Foley R. J., Kirshner R. P., 2017, *ApJ*, 842, 93  
 Möller A., de Boissière T., 2019, *Monthly Notices of the Royal Astronomical Society*, 491, 4277  
 Möller A., et al., 2016, *Journal of Cosmology and Astro-Particle Physics*, 2016, 008  
 Möller A., et al., 2021, *MNRAS*, 501, 3272

Muthukrishna D., Narayan G., Mandel K. S., Biswas R., Hložek R., 2019, arXiv e-prints, p. arXiv:1904.00014

Nicolas N., et al., 2021, *A&A*, 649, A74

Pierel J. D. R., et al., 2018, *PASP*, 130, 114504

Popovic B., Brout D., Kessler R., Scolnic D., Lu L., 2021, *ApJ*, 913, 49

Pursiainen M., et al., 2018, *MNRAS*, 481, 894

Qu H., Sako M., Möller A., Doux C., 2021, *AJ*, 162, 67

Rest A., et al., 2014, *ApJ*, 795, 44

Sako M., et al., 2011, *The Astrophysical Journal*, 738, 162

Sako M., et al., 2018, *PASP*, 130, 064002

Scolnic D., Kessler R., 2016, *ApJ*, 822, L35

Scolnic D., et al., 2020, *ApJ*, 896, L13

Shivvers I., et al., 2017, *PASP*, 129, 054201

Smith M., et al., 2018, *ApJ*, 854, 37

Smith M., et al., 2020, *The Astronomical Journal*, 160, 267

Stritzinger M. D., et al., 2011, *AJ*, 142, 156

Swann E., et al., 2019, *The Messenger*, 175, 58

Villar V. A., et al., 2019, *ApJ*, 884, 83

Villar V. A., et al., 2020, *ApJ*, 905, 94

Vincenzi M., Sullivan M., Firth R. E., Gutiérrez C. P., Frohmaier C., Smith M., Angus C., Nichol R. C., 2019, *MNRAS*, 489, 5802

Vincenzi M., et al., 2020, arXiv e-prints, p. arXiv:2012.07180

Vincenzi M., et al., 2021, arXiv e-prints, p. arXiv:2111.10382

Wilson A. G., Izmailov P., 2020, arXiv preprint arXiv:2002.08791

Wiseman P., et al., 2020a, *MNRAS*, 495, 4040

Wiseman P., et al., 2020b, *MNRAS*, 498, 2575

Wiseman P., et al., 2021, *MNRAS*, 506, 3330

Yuan F., et al., 2015, *MNRAS*, 452, 3047

de Jaeger T., et al., 2020, *MNRAS*, 495, 4860

## APPENDIX A: UNCERTAINTIES AND FITTED PARAMETERS

In Section 5.2 we explored the interpretability of BNN uncertainties. We concluded that this interpretation was not straight forward from our results. Here we extend this discussion by exploring possible correlations with other light-curve properties derived from a SALT2 fit in Figure A1.

In general, we find that uncertainties tend to be larger for the data when compared with simulations. The uncertainties in BBB method varies more with the parameters.

We note that the classification uncertainties are large for red and high stretch SNe in the DES 5-year sample. The median classification probability is also lower for these candidates. If the uncertainties are epistemic due to a smaller training set, then they would be large for the ends of the normal SNe Ia SALT2 parameter distributions since training sets have fewer such candidates. However, we do not find this behaviour. Another possible effect could be that bluer SNe Ia are more easily standardisable as previous literature suggests and thus their classification is more robust (Kelsey et al. 2021; Brout & Scolnic 2021). However, as this tendency is only observed in data and not simulations, no conclusion can be confidently drawn.

The peak magnitude in i-band behaviour in data agrees with that of the SNR of the light-curve. Brighter candidates are classified with higher confidence than fainter ones. However, as in the previous Section we do not see such a behaviour in the simulation.

While the correlation between supernova properties and classification uncertainties are interesting to explore, they are difficult to interpret since multiple effects could be contributing to the uncertainties. Tests based on simple physical systems could provide hints towards further interpretability, such as recent work by Caldeira & Nord (2020).

## APPENDIX B: DES 5-YEAR PHOTOMETRICALLY SELECTED SNE IA

A Table with photometrically classified SNe Ia from all selection methods with their respective probabilities for a subsample of DES 5-year data is provided at <https://doi.org/10.5281/zenodo.5904368>. Samples are selected using  $P$  larger than 0.5 for each method plus selection cuts.

## APPENDIX C: AUTHOR AFFILIATIONS

<sup>1</sup> Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Victoria 3122, Australia

<sup>2</sup> LPC, Université Clermont Auvergne, CNRS/IN2P3, F-63000 Clermont-Ferrand, France.

<sup>3</sup> Université de Lyon, Université Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, F-69622, Villeurbanne, France.

<sup>4</sup> Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>5</sup> School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK

<sup>6</sup> Department of Physics, Duke University Durham, NC 27708, USA

<sup>7</sup> The Research School of Astronomy and Astrophysics, Australian National University, ACT 2601, Australia

<sup>8</sup> Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain

<sup>9</sup> Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

<sup>10</sup> INAF-Osservatorio Astronomico di Trieste via Tiepolo 11, I-34143 Trieste, Italy.

<sup>11</sup> School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia

<sup>12</sup> Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK

<sup>13</sup> Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain.

<sup>14</sup> Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain.

<sup>15</sup> Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA

<sup>16</sup> Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

<sup>17</sup> Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW 2006, Australia

<sup>18</sup> Centre for Gravitational Astrophysics, College of Science, The Australian National University, ACT 2601, Australia

<sup>19</sup> Department of Physics, University of Surrey, Guildford, Surrey, UK, GU2 7XH

<sup>20</sup> Cerro Tololo Inter-American Observatory, NSF's National Optical-Infrared Astronomy Research Laboratory, Casilla 603, La Serena, Chile

<sup>21</sup> Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristina 77, Rio de Janeiro, RJ - 20921-400, Brazil

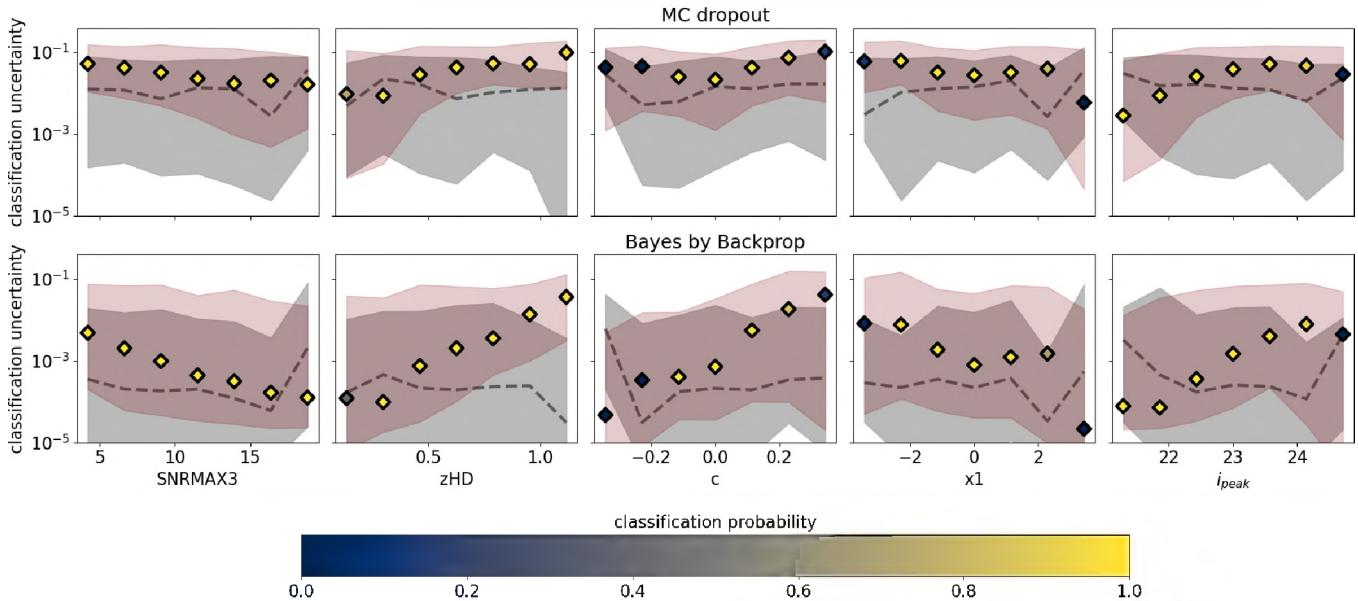
<sup>22</sup> Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL 60510, USA

<sup>23</sup> CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France

<sup>24</sup> Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France

<sup>25</sup> Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, 81679 Munich, Germany

**Table B1.** Example of SNN classification probabilities for DES 5-year candidates. A full list of classification probabilities for all DES 5-year candidates that are selected for any of the samples in this paper is found in <https://doi.org/10.5281/zenodo.5904368>. We show probabilities for baseline Recurrent Neural Network (RNN) in Section 4) and Bayesian Neural Networks methods MC dropout (MC) and Bayes by Backprop (BBB) in Section 5. For each method we provide classification probabilities rounded in two decimals for five different SNN initialisation seeds,  $s_0 = 0$ ,  $s_1 = 55$ ,  $s_2 = 100$ ,  $s_3 = 1000$ ,  $s_4 = 30469$  and the ensemble average probability of these five seeds  $set0$ .



**Figure A1.** Distribution of classification uncertainties for DES 5-year data (maroon) and simulation (grey) using the two BNNs, MC dropout and Bayes by Backprop. We show uncertainties as a function of SNR of the third brightest point in the light-curve, redshift, colour, stretch and peak  $i$ -band magnitude. The median and 68 percentile are shown as a dashed line and filled coloured area. Data is shown as diamonds coloured by the median probability in that parameter bin and a maroon coloured area representing the 68% percentile of the distribution.

<sup>26</sup> Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK

<sup>27</sup> Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA

<sup>28</sup> SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

<sup>29</sup> Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain

<sup>30</sup> Universidad de La Laguna, Dpto. Astrofísica, E-38206 La Laguna, Tenerife, Spain

<sup>31</sup> Center for Astrophysical Surveys, National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA

<sup>32</sup> Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA

<sup>33</sup> Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain

<sup>34</sup> Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain

<sup>35</sup> Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain

<sup>36</sup> Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester, M13 9PL, UK

<sup>37</sup> University of Nottingham, School of Physics and Astronomy, Nottingham NG7 2RD, UK

<sup>38</sup> Astronomy Unit, Department of Physics, University of Trieste, via Tiepolo 11, I-34131 Trieste, Italy

<sup>39</sup> INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy

<sup>40</sup> Institute for Fundamental Physics of the Universe, Via Beirut 2, 34014 Trieste, Italy

<sup>41</sup> Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil

<sup>42</sup> Department of Physics, IIT Hyderabad, Kandi, Telangana 502285,

India

<sup>43</sup> Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA

<sup>44</sup> Institute of Theoretical Astrophysics, University of Oslo. P.O. Box 1029 Blindern, NO-0315 Oslo, Norway

<sup>45</sup> Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, 28049 Madrid, Spain

<sup>46</sup> Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA

<sup>47</sup> Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

<sup>48</sup> Excellence Cluster Origins, Boltzmannstr. 2, 85748 Garching, Germany

<sup>49</sup> Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA

<sup>50</sup> Department of Physics, The Ohio State University, Columbus, OH 43210, USA

<sup>51</sup> Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia

<sup>52</sup> Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA

<sup>53</sup> George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA

<sup>54</sup> Institut Català de Recerca i Estudis Avançats, E-08010 Barcelona, Spain

<sup>55</sup> Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706-1390

<sup>56</sup> Department of Astronomy, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA

<sup>57</sup> Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

<sup>58</sup> Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA

<sup>59</sup> Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton, BN1 9QH, UK

<sup>60</sup> Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

<sup>61</sup> Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85748 Garching, Germany

<sup>62</sup> Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, 81679 München, Germany