The Astronomical Journal, 163:57 (14pp), 2022 February https://doi.org/10.3847/1538-3881/ac39a1


# Photometric Classification of Early-time Supernova Light Curves with SCONE

Helen Qu and Masao Sako
Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA; helenqu@sas.upenn.edu
*Received 2021 September 22; revised 2021 November 10; accepted 2021 November 12; published 2022 January 11*

## Abstract

In this work, we present classification results on early supernova light curves from SCONE, a photometric classifier that uses convolutional neural networks to categorize supernovae (SNe) by type using light-curve data. SCONE is able to identify SN types from light curves at any stage, from the night of initial alert to the end of their lifetimes. Simulated LSST SNe light curves were truncated at 0, 5, 15, 25, and 50 days after the trigger date and used to train Gaussian processes in wavelength and time space to produce wavelength–time heatmaps. SCONE uses these heatmaps to perform six-way classification between SN types Ia, II, Ibc, Ia-91bg, Iax, and SLSN-I. SCONE is able to perform classification with or without redshift, but we show that incorporating redshift information improves performance at each epoch. SCONE achieved 75% overall accuracy at the date of trigger (60% without redshift), and 89% accuracy 50 days after trigger (82% without redshift). SCONE was also tested on bright subsets of SNe ($r < 20$ mag) and produced 91% accuracy at the date of trigger (83% without redshift) and 95% five days after trigger (94.7% without redshift). SCONE is the first application of convolutional neural networks to the early-time photometric transient classification problem. All of the data processing and model code developed for this paper can be found in the SCONE software package[1] located at github.com/helenqu/scone (Qu 2021).

*Unified Astronomy Thesaurus concepts:* Photometry (1234); Light curves (918); Supernovae (1668); Classification (1907); Gaussian Processes regression (1930); Neural networks (1933)

## 1. Introduction

Observations of transient and supernova phenomena have informed fundamental discoveries about our universe, ranging from its expansion history and current expansion rate (Riess 1998; Perlmutter et al. 1999; Freedman et al. 2019; Riess et al. 2019) to the progenitor physics of rare and interesting events (Pursiainen et al. 2018; Armstrong et al. 2021). In the near future, next-generation wide-field sky surveys such as the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST; Ivezić 2019) will have the ability to observe larger swaths of sky with higher resolution and certainly uncover even more new and exciting astrophysical phenomena.

These surveys promise to generate ever-larger volumes of photometric data at unprecedented rates. However, the availability of spectroscopic resources is not expected to scale nearly as quickly. Thus, the challenge of effectively allocating these limited resources is more important than ever. For type Ia SN cosmology, spectroscopic information is used to minimize contamination in constructing pure and representative samples of SNe Ia to continue to constrain the dark energy equation of state. For supernova physicists, spectra uncover important information about an event's potential progenitor processes (Filippenko 2005; Perets et al. 2010; Modjaz et al. 2014; Sollerman et al. 2021). Spectra taken near peak brightness of an event are optimal because they include mostly transient information and are not dominated by host galaxy features.

With millions of alerts each night, fast and accurate automatic classification mechanisms will be needed to replace the time-consuming process of manual inspection. More specifically, the ability to perform classification early on in the lifetime of a transient would allow for ample time to take spectra at the peak luminosity of the event or at multiple points over the course of the event's lifetime.

### 1.1. Photometric Supernova Classification

An impressive body of work has emerged over the past decade on photometric classification of supernovae. Since only a small percentage of discovered supernovae have ever been followed up spectroscopically, a reliable photometric classifier is indispensable to the advancement of supernova science.

The Supernova Photometric Classification Challenge (SNPhotCC; Kessler et al. 2010a, 2010) created not only an incentive to invest in photometric SN classification, but also a data set that would be used to train and evaluate classifiers for years to come. Successful approaches range from empirical template-fitting (Sako et al. 2008) to making classification decisions based on manually extracted features (Richards et al. 2012; Karpenka et al. 2013). The more recent Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC; The PLAsTiCC team et al. 2018) diversified the data set by asking participants to differentiate between 14 different transient and variable object classes, including the six common supernova types included in this work. The top entries made use of feature extraction paired with various machine-learning classification methods, such as boosted decision trees and neural networks (Hložek et al. 2020). Ensemble methods, in which the results of multiple classifiers are combined to create the final classification probability, were widely used as well.

Deep learning is a branch of machine learning that seeks to eliminate the necessity of human-designed features, decreasing the computational cost as well as avoiding the introduction of potential biases (Charnock & Moss 2017; Moss 2018; Naul et al. 2018; Aguirre et al. 2019). In recent years, many deep learning techniques have been applied to the challenge of photometric SN classification.

1 github.com/helenqu/scone

Recurrent neural networks (RNNs) are designed to learn from sequential information, such as time-series data, and have been used with great success on this problem. Charnock & Moss (2017) applies a variant of RNNs known as Long Short Term Memory networks (LSTMs; Hochreiter & Schmidhuber 1997) to achieve impressive performance distinguishing SNIa from core collapse (CC) SNe. Muthukrishna et al. (2019) use a gated recurrent unit (GRU) RNN architecture to be able to perform real-time and early light-curve classification. Möller & de Boissière (2020) perform both binary classification and classification by type with full and partial light curves using Bayesian RNNs. Villar et al. (2020b) use a GRU RNN as an autoencoder to smooth out irregularities in light-curve data that are then fed into a random forest classifier.

Convolutional neural networks (CNNs), which are used in this work, are a state-of-the-art image recognition architecture (LeCun et al. 1989, 1998; Krizhevsky et al. 2012; Zeiler & Fergus 2014). Pasquet et al. (2019) address the issue of nonrepresentative training sets by using a CNN as an autoencoder to learn from unlabeled test data. Carrasco-Davis et al. (2021) developed an image time-series classifier as part of the ALeRCE alert broker, using a CNN to differentiate between various transient types as well as bogus alerts.

Outside of these traditional models, deep learning is still providing new and creative solutions to the photometric transient classification problem. Convolutional recurrent neural networks are used to classify a time series of image stamps by Kodi Ramanah et al. (2021) to detect gravitationally lensed supernovae. A newer type of deep learning architecture, known as a transformer, achieves a very impressive result when applied to the PLAsTiCC data set by Allam & McEwen (2021). A variational autoencoder is used by ParSNIP (Boone 2021) to develop a low-dimensional representation of transient light curves that uses redshift-annotated photometric data to perform full light-curve photometric classification and generate time-varying spectra, among other tasks.

### 1.2. Early Photometric Supernova Classification

Though much progress has been made on the photometric supernova classification problem, most of the solutions tackle classification of full supernova light curves retrospectively. However, the earlier an object can be classified, the more opportunities there are for the community to perform follow-up observation. Spectroscopic or photometric follow-up at early stages not only reveals insights into progenitor physics, but can also serve as a benchmark for further observations at later epochs. SN type IIb, for example, exhibit hydrogen features in early spectra that quickly disappear over time (Woosley et al. 1987). Shock breakout physics is another use case of follow-up observation. Armstrong et al. (2021) were the first to report capturing the complete evolution of a shock cooling light curve, a short-lived event preceding peak luminosity that reveals properties of the shock breakout and progenitor star for stripped-envelope supernovae such as the IIb.

Despite the general focus on full light-curve classification, several notable works have addressed the challenge of early photometric classification. Sullivan et al. (2006) were able to not only differentiate between SNIa and CC SN, but also predict redshift, phase, and light-curve parameters for SNIa using only two or three epochs of multiband photometry data. Poznanski et al. (2007) also performed binary Ia versus CC SNe classification, but using a Bayesian template-fitting technique on only single-epoch photometry and photometric redshift estimates. PSNID (Sako et al. 2008, 2011), the algorithm that produced the highest overall figure of merit in SNPhotCC, was used by the Sloan Digital Sky Survey (Frieman et al. 2008) and the Dark Energy Survey (Smith et al. 2020) to classify early-time and full supernova light curves.

The work of Muthukrishna et al. (2019) is a recent application of deep learning techniques specifically to early-time transient classification. A GRU RNN is trained and tested on a PLAsTiCC-derived data set of 12 transients, including seven supernova types, that are labeled at each epoch with "pre-explosion" prior to the date of explosion and the correct transient type after explosion. Thus, the model is able to produce a classification at each epoch of observation. Möller & de Boissière (2020) have also produced an RNN-based photometric classifier that is capable of classifying partial supernova light curves, but primarily achieves good results for Ia versus CC SN classification. Villar et al. (2020a) use a recurrent variational autoencoder architecture to perform early-time anomaly detection for exotic astrophysical events within the PLAsTiCC data set, such as active galactic nuclei and superluminous SNe. Finally, LSST alert brokers such as ALeRCE (Sánchez-Sáez et al. 2021) specialize in accurate early-time classification of transient alerts.

### 1.3. Overview

Originally introduced in Qu et al. (2021), hereafter Q21, as a full light-curve photometric classification algorithm, SCONE was able to retrospectively differentiate Ia versus CC SN with $>99\%$ accuracy and categorize SNe into six types with $>98\%$ accuracy without redshift information. Our approach centers on producing heatmaps from two-dimensional Gaussian processes fit on each light curve in both wavelength and time dimensions. These flux heatmaps of each supernova detection, along with "uncertainty heatmaps" of the Gaussian process uncertainty, constitute the data set for our model. This preprocessing step smooths over irregular sampling rates between filters, mitigates the effect of flux outliers, and allows the CNN to learn from information in all filters simultaneously.

Section 2 outlines the details of the data sets and models used in this work, and we discuss the classifier's performance on the various data set types in Section 3, including a comparison with existing literature. We state our conclusions and goals for future work in Section 4.

## 2. Methods

### 2.1. Simulations

For this work, SCONE was trained and tested on a set of LSST deep drilling field (DDF) simulations. The data set was created with SNANA (Kessler et al. 2009) using the PLAsTiCC transient class models for supernovae types Ia, II, Ibc, Ia-91bg, Iax, and SLSN (Guy et al. 2010; Kessler et al. 2010; Kasen & Bildsten 2010; Kessler et al. 2013; Jha 2017; Nicholl et al. 2017; Guillochon et al. 2018; Villar et al. 2017, 2017; Pierel et al. 2018; The PLAsTiCC team et al. 2018; Kessler et al. 2019). The relative rates and redshift distribution are identical to those of the data produced for the PLAsTiCC challenge. This is the same data set used to evaluate SCONE's categorical classification performance in Q21. No cuts on individual low S/N ratio light-curve points were made, but

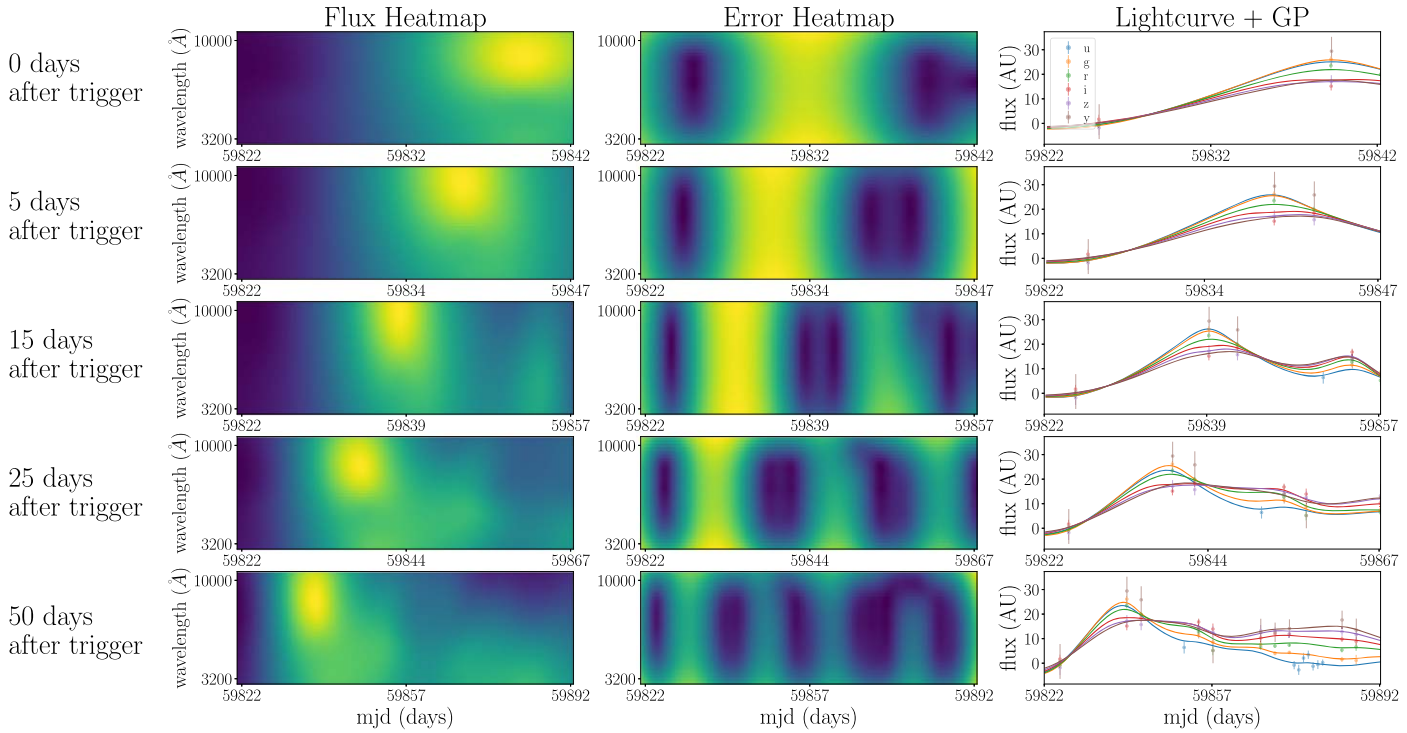


**Figure 1.** An SNII ($z = 0.39$) shown in all five heatmap data sets along with the light curves and Gaussian process fits used to create each heatmap. The flux and flux error measurements from the raw photometry are shown as points with error bars, while the Gaussian process fits to each photometry band are shown as curves. The Gaussian process errors, which are used to create the heatmaps in the middle column, are not shown in the light-curve plots. The $x$-axis limit of the plots in each row are different, as the light curve is truncated according to the label on the left for each row in the figure.

light curves with fewer than two $5\sigma$ detections were removed, as $t_{\rm trigger}$ would be ill-defined in those cases. We note that, in observed data, transient light-curve samples will contain SNe contaminated by other galactic astrophysical sources, but methods such as Sánchez-Sáez et al. (2021) are reliably able to distinguish extragalactic and galactic events. Thus, we can assume the feasibility of creating a pure sample of SN light curves such as the one used in this work.

### 2.2. Trigger Definition

We define a *detection* as any observation exceeding the $5\sigma$ signal-to-noise ratio (S/N) threshold. We define the *trigger* as the next detection that occurs at least one night after the first. In this work, the data set with the least photometric information includes observations up to (and including) the date of trigger. Thus, all SNe in our data sets have at least two epochs of observation. As the date of first detection is also a common choice of trigger date in other transient surveys, the implications of this discrepancy are explored further in Section 3.3. We present results on a data set where the distinction between these two definitions is small, i.e., $t_{\rm trigger} \leqslant t_{\rm first\ detection} + 5$.

### 2.3. Data Sets and Heatmap Creation

#### 2.3.1. $t_{trigger} + N$ Data Sets

To evaluate SCONE's classification performance on light curves at different stages of the supernova lifetime, five sets of heatmaps were created from the simulations described in Section 2.1. All sets of heatmaps take data starting 20 nights prior to the date of trigger ($t_{\rm trigger}$) and end at $N = 0$, 5, 15, 25, and 50 days after the date of trigger, respectively. Hereafter, these are collectively referred to as "$t_{\rm trigger} + N$ data sets."

Prior to training, the light-curve data is processed into heatmaps. We use the approach described by Boone (2019) to apply two-dimensional Gaussian process regression to the raw light-curve data to model the event in the wavelength ($\lambda$) and time ($t$) dimensions. We use the Matérn kernel ($\nu = \frac{3}{2}$) with a fixed 6000 Å characteristic length scale in $\lambda$ and fit for the length scale in $t$. Once the Gaussian process regression model has been trained, we obtain its predictions on a $\lambda$, $t$ grid and call this our "flux heatmap."

It is important to note that the Gaussian processes are fit on light curves truncated at $N$ days after trigger in each data set and not given access to light-curve information past the cutoff date. Thus, though the $\lambda$ axis is not affected by the different choices of $N$, the $t$ range of the input light-curve data varies for each $t_{\rm trigger} + N$ data set. For the data sets in this work, the $\lambda$, $t$ grids were chosen to preserve the shape of the resulting heatmap despite the fact that the number of nights of light-curve data varies between the $t_{\rm trigger} + N$ data sets. $\lambda$ is chosen to be $3000 < \lambda < 10,100$ Å with a 221.875 Å interval for all data sets, while the $t$ interval depends on the number of nights of data: $t_{\rm trigger} - 20 \leqslant t \leqslant t_{\rm trigger} + N$ with a $\frac{N+20}{180}$ day interval, where $N = 0$, 5, 15, 25, 50. This ensures that all heatmaps have size $32 \times 180$.

In addition to the flux heatmap, we also take into account the uncertainties on these predictions at each $\lambda_i$, $t_j$, producing an "error heatmap." We stack these two heatmaps depthwise for each SN light curve and divide by the maximum flux value to constrain all entries to [0,1]. This $32 \times 180 \times 2$ tensor is our input to the convolutional neural network.

An example of the heatmaps and associated light curves of a single SN in all five data sets is shown in Figure 1. Results on the $t_{\rm trigger} + N$ data sets are described in Section 3.2.

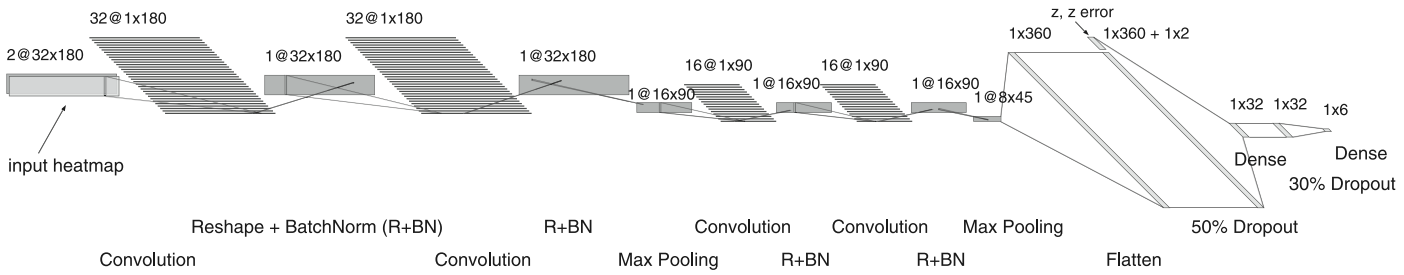


**Figure 2.** SCONE architecture with redshift information for categorical early light-curve classification.

**Table 1**
Training, Validation, and Test Data Set Sizes for the $t_{\rm trigger}+N$ Data Sets.

| Data Set | Number of Each Type | Total Size |
|---|---|---|
| Training | 6148 | 36888 |
| Validation | 769 | 4614 |
| Test | 768 | 4608 |
| Full | 7685 | 46110 |

#### 2.3.2. Bright Supernovae

Our model was also evaluated on the subset of particularly bright supernovae from the $t_{\rm trigger}+0$ and $t_{\rm trigger}+5$ data sets to emulate a real-world use case of SCONE for spectroscopic targeting, as bright supernovae are better candidates for spectroscopic follow-up. "Bright SNe" included in these data sets were chosen to be SNe with last included detection $r<20$ mag. With this threshold, there were 907 SNe in the $t_{\rm trigger}+0$ bright data set and 5088 SNe in the $t_{\rm trigger}+5$ bright data set. As described in more detail in Section 2.4, SCONE was trained with a standard $t_{\rm trigger}+N$ training set combined with 40% of the $t_{\rm trigger}+N$ bright data set, and tested on the $t_{\rm trigger}+N$ bright data set. Results on these data sets are described in Section 3.5.

#### 2.3.3. Mixed Data Set

In order to evaluate SCONE's ability to classify SNe with any number of nights of photometry, a sixth data set (the "mixed" data set) was created from the same PLAsTiCC simulations. Data is taken starting 20 nights prior to the date of trigger (as with the $t_{\rm trigger}+N$ data sets) but truncated at a random night between 0 and 50 days after trigger. Due to the choice of the $t$ interval described in Section 2.3.1, heatmaps with any number of nights of photometry data are all the same size and can thus be mixed in a single data set in this manner. We train SCONE on this mixed data set and evaluate its performance on each of the $t_{\rm trigger}+N$ data sets in Section 3.6.

### 2.4. Data Set Train/Test Split

Due to the importance of class balancing in machine-learning data sets, the same quantity of SNe from each SN type was selected to create the $t_{\rm trigger}+N$ and mixed data sets used to train, validate, and test SCONE. For this purpose, 7685 SNe of each of the six types were randomly chosen, as this was the quantity of the least abundant type. Thus, the size of each full data set was 46,110. An 80/10/10 training/validation/test split was used for all results in this work. The sizes of the training, validation, and test subsets of each data set can be found in Table 1.

For evaluation on the bright data sets, SCONE was trained on a hybrid training set of 40% of the $t_{\rm trigger}+N$ bright data set combined with a $t_{\rm trigger}+N$ training set, prepared as described in Section 2.3.1. Thus, the training set was not quite class-balanced, as the bright data set is not class-balanced but the $t_{\rm trigger}+N$ training set is. The trained model was then evaluated on the full bright data set to produce the results shown in Figure 10. Due to the imbalanced nature of the bright data sets, the confusion matrices in this figure take the place of an accuracy metric, which could be misleading. We chose to include 40% of the bright data set in the training process, to ensure that the model has seen enough of these particularly bright objects to make reasonable predictions.

### 2.5. Model

In this work, we report early light-curve classification results using the vanilla SCONE model developed in Q21 as well as a variant of SCONE that incorporates redshift information. The architecture of SCONE with redshift is shown in Figure 2. Both redshift and redshift error are concatenated with the output of the first dropout layer and used as inputs to the fully connected classifier. The model uses spectroscopic redshift information when available and photometric redshift estimates if not.

Prior to training and testing, the input flux and error heatmaps are divided by the maximum flux value of each heatmap for normalization. This means that absolute brightness information is not used for classification. All results in this work, with and without redshift, used the sparse categorical crossentropy loss function, the Adam optimizer (Kingma & Ba 2014), and trained for 400 epochs with a batch size of 32. SCONE without redshift used a constant 1e-3 learning rate, whereas SCONE with redshift used a constant 5e-4 learning rate.

### 2.6. Computational Performance

The time required for the heatmap creation process was measured using a sample of 100 heatmaps on a single 32-core NERSC Cori Haswell compute node (with Intel Xeon Processor E5-2698 v3). The time required to create one heatmap was $0.03\pm0.01$ seconds. When producing larger-scale data sets, this process is also easily parallelizable over multiple cores or nodes, to further decrease heatmap creation time.

SCONE without redshift has 22,606 trainable parameters and SCONE with redshift has 22,670 trainable parameters, while other photometric classification models require at least hundreds of thousands. The performance gains of this simple but effective model compounded with a small training set make SCONE lightweight and fast to train. The first training epoch on a NVIDIA V100 Volta GPU takes approximately 17 s (4 ms

**Table 2**
Training, Validation, and Test Accuracies *without* Redshift Information for Each Early Light-curve Data Set

| Accuracy without Redshift | Days after Trigger | | | | |
|---|---|---|---|---|---|
| | 0 day | 5 day | 15 days | 25 days | 50 days |
| Training | 58.36 ± 0.14% | 68.92 ± 0.21% | 73.99 ± 0.14% | 76.89 ± 0.29% | 80.93 ± 0.14% |
| Validation | 59.57 ± 0.51% | 70.74 ± 0.59% | 73.31 ± 3.01% | 79 ± 0.84% | 82.5 ± 2.35% |
| Test | 59.66 ± 0.43% | 70.05 ± 0.63% | 73.66 ± 2.36% | 79 ± 0.86% | 82.2 ± 1.8% |

**Note.** These averages and standard deviations were computed from five independent runs of SCONE.

**Table 3**
Training, Validation, and Test Accuracies *with* Redshift Information for Each Early Light-curve Data Set

| Accuracy with Redshift | Days after Trigger | | | | |
|---|---|---|---|---|---|
| | 0 day | 5 day | 15 days | 25 days | 50 days |
| Training | 72.73 ± 0.27% | 79.61 ± 0.3% | 83.07 ± 0.2% | 84.68 ± 0.2% | 87.17 ± 0.26% |
| Validation | 74.78 ± 0.18% | 80.52 ± 1.42% | 83.98 ± 1.15% | 86.75 ± 0.5% | 89.2 ± 0.85% |
| Test | 74.27 ± 0.51% | 80.2 ± 0.93% | 84.14 ± 1.37% | 86.71 ± 1% | 89.04 ± 0.39% |

**Note.** These averages and standard deviations were computed from five independent runs of SCONE.

per batch with a batch size of 32), and subsequent training epochs take approximately 5 s each with TensorFlows data set caching. The first training epoch on a Haswell node takes approximately 12 (625 ms per batch), and subsequent epochs take approximately 6 minutes each. Test time per batch of 32 examples is 3 ms on GPU and 10 ms on a Haswell CPU.

## 3. Results and Discussion

### 3.1. Evaluation Metrics

The *accuracy* of a set of predictions describes the frequency with which the predictions match the true labels. In this case, we define our prediction for each SN example as the class with highest-probability output by the model, and compare this to the true label to obtain an accuracy.

The *confusion matrix* is a convenient visualization of the correct and missed predictions by class, providing a bit more insight into the model's performance. The confusion matrices shown in Figure 4 are normalized such that the ($i$,$j$) entry describes the fraction of the true class, $i$, classified as class $j$. The confusion matrices in Figure 10 are colored by the normalized values, just like Figure 4, but overlaid with absolute (non-normalized) values. For both figures, the ($i$,$i$) entries, or those on the diagonal, describe correct classifications.

The *receiver operating characteristic (ROC) curve* makes use of the output probabilities for each class rather than simply taking the highest-probability class, as the previous two metrics have done. We consider an example to be classified as class $i$ if the output probability for class $i$, or $p_i$, exceeds some threshold $p$ ($p_i > p$). The ROC curve sweeps values of $p$ between 0 and 1 and plots the true-positive rate (TPR) at each value of $p$ against the false-positive rate (FPR).

TPR is the percentage of correctly classified objects in a particular class, or true positives (TP), as a fraction of all examples in that class, true positives and false negatives (TP+FN). Other names for TPR include *recall* and *efficiency*. The values along the diagonal of the normalized confusion matrices in Figure 4 are efficiency values:

$$\text{Efficiency} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

FPR is the percentage of objects incorrectly classified as a particular class, or false positives (FP), as a fraction of all examples not in that class, false positives and true negatives (FP+TN):

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}.$$

The *area under the ROC curve*, or *AUC*, is used to evaluate the classifier from its ROC curve. A perfect classifier would have an AUC of 1, while a random classifier would score (on average) a 0.5.

The *precision* or *purity* of a set of predictions is the percentage of correctly classified objects in a particular predicted class:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

### 3.2. $t_{trigger} + N$ Data Sets

The accuracies our model achieved without redshift on each $t_{\text{trigger}} + N$ data set are described in Table 2, and the accuracies with redshift are described in Table 3. These tables show that redshift unequivocally improves classification performance, especially at early times when there is little photometric data to learn from. The inclusion of redshift information not only increases the average accuracies for each data set but also improves the model's generalizability, as the standard deviations for the validation and test accuracies are lower overall in Table 3.

The largest improvement in accuracy between $t_{\text{trigger}} + N$ data sets occurred between 0 and 5 days after trigger for all data sets. Because the explosion likely reached peak brightness during this period, the light curves truncated at five days after

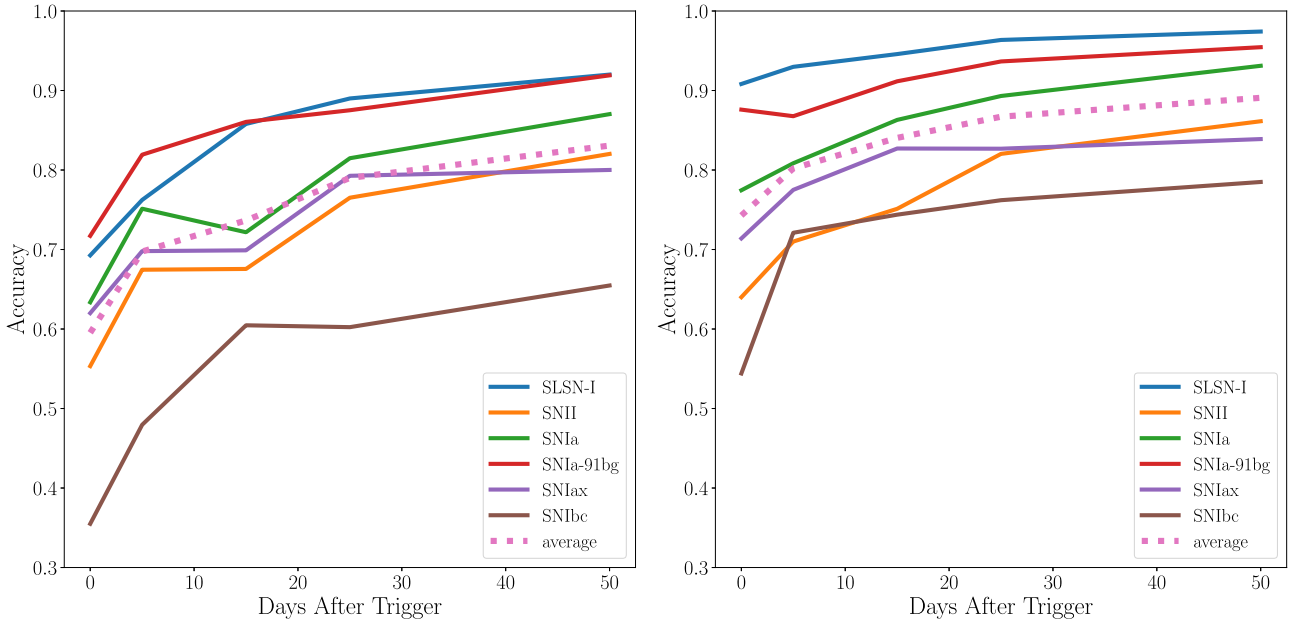


**Figure 3.** Accuracy/efficiency over time for each supernova type without redshift (left) and with redshift (right) for the $t_{\rm trigger} + N$ test data sets. The values used in this plot correspond with the diagonals on each normalized confusion matrix in Figure 4.

trigger include much more information necessary for differentiating between the SN types.

Figure 3 shows the accuracy evolution over time for each supernova type in the test sets. From the test sets with redshift plot on the right, it is clear that the jump in overall accuracy between 0 and 5 days after trigger can be attributed to the sharp accuracy boost experienced by SNIbc at 5 days after trigger. Overall, SNIbc benefited the most from the inclusion of redshift, though classification performance on all types saw improvement. Note that, as described in Section 2.3, all heatmaps are normalized to values between 0 and 1 so absolute flux values are not used to differentiate between types. Thus, the model cannot rely on relative luminosity information.

The confusion matrices for $t_{\rm trigger} + \{0, 5, 50\}$ test sets with and without redshift information are shown in Figure 4. The top two panels are early epoch classification results (0 and 5 days after trigger) and the bottom panel shows late epoch results. The confusion matrices from intermediate epochs (15 and 25 days after trigger) were omitted for brevity.

At the date of trigger (top panel of Figure 4), the incorporation of redshift information primarily prevents confusion between SLSN-I and SNIbc. True SLSN-I events misclassified as SNIbc decreased from 11% on average to 2% with redshift. True SNIbc misclassified as SLSN-I decreased from 16% on average to 2% with redshift. Overall, SLSN-I were classified with 91% accuracy with redshift compared to 69% without redshift, and SNIbc were classified with 54% accuracy with redshift compared to 36% without redshift. All types saw marked improvement in classification performance without redshift from 0 to 5 days after trigger, while classification with redshift saw drastic improvement in SNIbc accuracy but only minor improvement for other types. Finally, the effect of added redshift becomes less noticeable by late epochs, where classification accuracy (along the diagonal) is only mildly improved in the bottom panel of Figure 4.

The confusion matrices in Figure 4 are normalized by true type, meaning that the values in each row sum to unity. Thus, the values along the diagonal are *efficiency* scores. Normalizing by predicted type, such that the values in each column sum to unity, would result in *purity* scores along the diagonal. However, since all data sets used in Figure 4 are class-balanced, the purity scores can be reconstructed from these confusion matrices by dividing each main diagonal value by the sum of the values in its column.

The data sets used for the confusion matrices in Figure 4 were also used to create ROC curves for each SN type. ROC curves for test sets without redshift are shown on the left side of Figure 5, and ROC curves for test sets with redshift are shown on the right. The addition of redshift information seems to most notably improve the model's ability to classify SLSN-I—all three panels on the right show SLSN-I as the highest AUC curve, whereas all three panels on the left show SNIa-91bg with a higher AUC curve than SLSN-I. This is consistent with our earlier observations from the confusion matrices and accuracy plots.

The information in the ROC curves for all $t_{\rm trigger} + N$ data sets is summarized in Figure 6, showing AUC over time plots with and without redshift. The performance looks quite impressive, starting at an average AUC of above 0.9 with redshift at the date of trigger and increasing to 0.975 by 50 days after trigger. Without redshift, average AUC is still respectable, starting at 0.88 and increasing to 0.97.

### 3.3. Approximating a First-detection Trigger Definition

Another common trigger definition used in transient surveys places the trigger at the date of the first detection ($t_{\rm first\ detection}$) rather than the second, which is the definition followed in this work. In order to more directly compare SCONE's results with those of other classifiers following the first detection trigger definition, the distribution of $t_{\rm trigger} - t_{\rm first\ detection}$ was examined as well as SCONE's performance on the subset of the $t_{\rm trigger} + 0$ data set with date of second detection ($t_{\rm trigger}$) at most 5 days after the date of first detection (i.e., $t_{\rm trigger} \leqslant t_{\rm first\ detection} + 5$).

Figure 7 shows that $> 65\%$ of $t_{\rm trigger}$ dates are no more than 5 days after the date of first detection. To further understand the

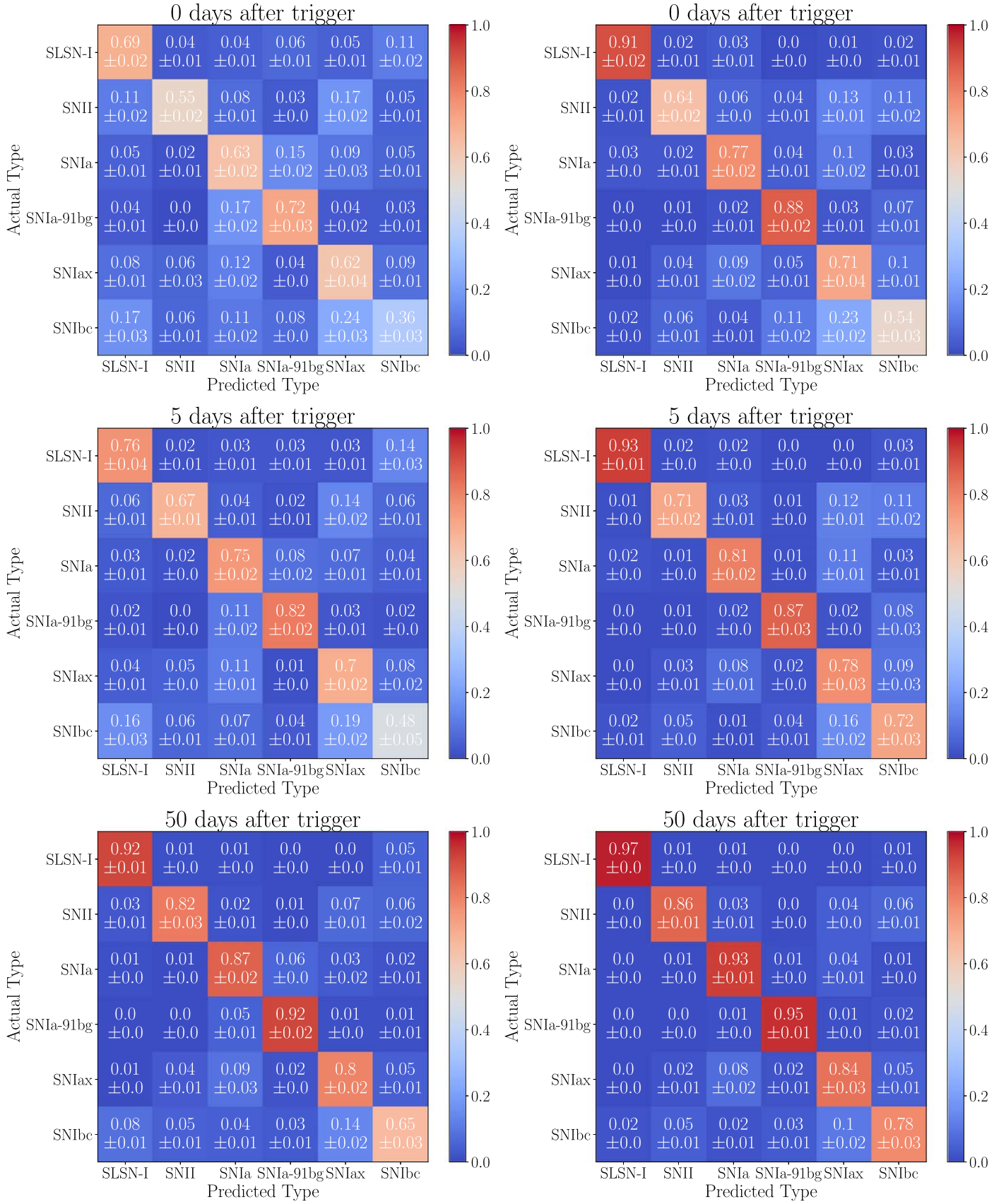


**Figure 4.** Normalized confusion matrices produced by SCONE without (left) and with (right) redshift for the $t_{\text{trigger}} + \{0, 5, 50\}$ test sets (heatmaps created from light curves truncated at 0, 5, and 50 days after the date of trigger). These matrices were made with test set classification performance from five independent runs of SCONE.

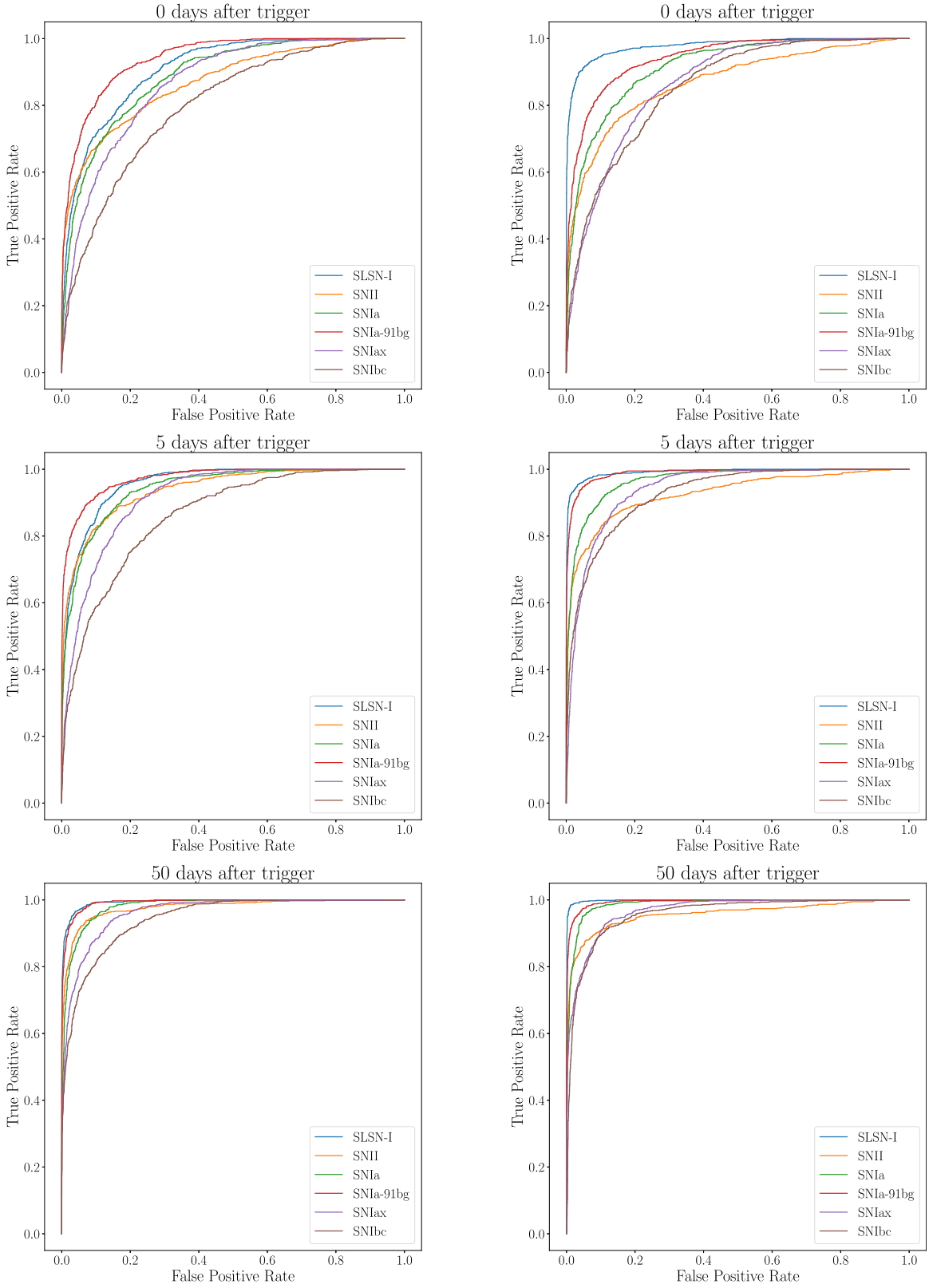


**Figure 5.** Receiver operating characteristic (ROC) curves produced by SCONE without (left) and with (right) redshift for the $t_{\text{trigger}} + \{0, 5, 50\}$ test sets (heatmaps created from light curves truncated at 0, 5, and 50 days after the date of trigger).

direct impact of this choice of trigger definition, SCONE was tested on the subset of the $t_{\text{trigger}} + 0$ data set with date of second detection ($t_{\text{trigger}}$) at most 5 days after the date of first detection. This cut ensures that the light curves used for classification are not given substantially more information than those created with the first detection trigger definition. The normalized confusion matrices for the $t_{\text{trigger}} \leqslant t_{\text{first detection}} + 5$ data set are shown with and without redshift in Figure 8.

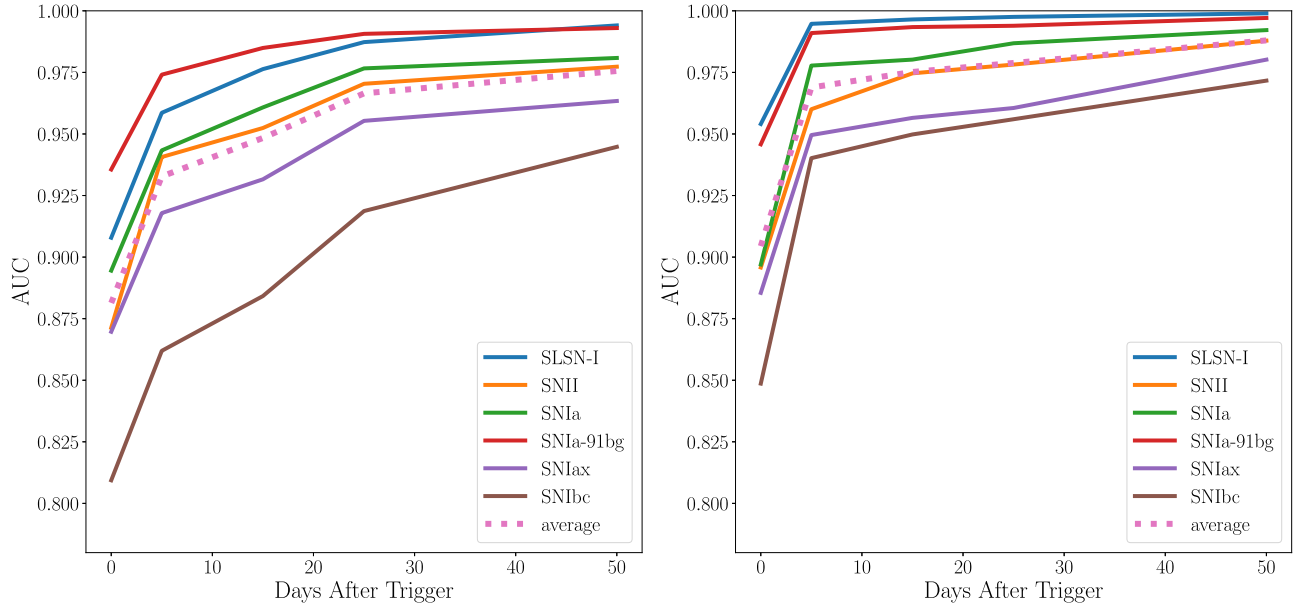


**Figure 6.** Area under the ROC curve (AUC) without (left) and with (right) redshift over time for each supernova type.

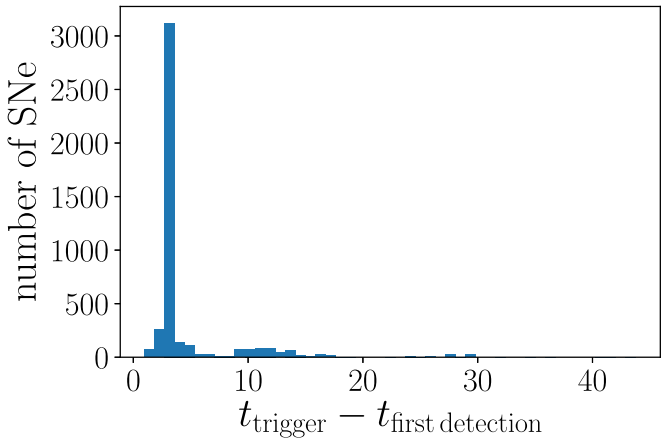


**Figure 7.** Distribution of $t_{\text{trigger}} - t_{\text{first detection}}$ in a SCONE test data set of 4608 SNe.

With redshift, SCONE's performance primarily suffers on SLSN-I and SNIa classification. SLSN-I appears to more strongly resemble SNIax and SNIbc at early times, as the SNIax confusion rose to 8% from 1% and the SNIbc confusion rose to 11% from 2%. SNIa were commonly misclassified as SNIa-91bg at early times, which is not reflected in the $t_{\text{trigger}} + 0$ confusion matrices in Figure 4. Surprisingly, true SNIa-91bg were not misclassified as SNIa despite the prevalence of SNIa misclassified as SNIa-91bg. Without redshift, however, SCONE's performance on the $t_{\text{trigger}} \leqslant t_{\text{first detection}} + 5$ subset very closely resembles the $t_{\text{trigger}} + 0$ results shown in Figure 4.

### 3.4. Baseline Model

A multilayer perceptron model (MLP; Hornik et al. 1989) was developed as a baseline for direct comparison to SCONE. MLP architectures are a simple type of feedforward neural network with at least three layers (input, hidden, output) in which each node in a particular layer is connected to every node in the subsequent layer. They have been successfully used in many general as well as image classification tasks (Liu et al. 2021; Tolstikhin et al. 2021).

The $32 \times 180 \times 2$ input heatmap is split into 180 non-overlapping "patches" of size $32 \times 1$. The patches were chosen to be full height in the wavelength dimension, to remain consistent with the full height convolutional kernels used in SCONE. A $180 \times 64$-dimensional hidden layer is then computed via $h_{1,ij} = \text{relu}(x_i^j W_{1,ji} + b_{1,j})$, where $\text{relu}(x) = \max(0, x)$ is the rectified linear unit, $x^j$ is the $j$th input heatmap patch, $W_1$ is the weight matrix learned by the network, and $b_1$ is the learned bias vector. The dimensionality of the hidden layer is then squashed to a single 64-dimensional vector with global average pooling: $h_{2,i} = \text{average}(h_{1,ij})$. Finally, the output class is computed via $y_k = \sigma(h_{2,i} W_{2,ji} + b_{2,j})_k$, where $\sigma(\vec{x})_k = \frac{e^{x_k}}{\sum_j e^{x_j}}$ is the softmax function, $W_2$ is the learned weight matrix, and $b_2$ is the learned bias vector.

Without redshift, our model achieved a test accuracy of 56%. With redshift, the test accuracy improved to 67.19%. The performance of the MLP on the $t_{\text{trigger}} + 0$ data set with and without redshift is summarized in the confusion matrices in Figure 9. Compared to the performance of SCONE on the $t_{\text{trigger}} + 0$ data set in the top panel of Figure 4, the MLP is less accurate at classifying most SN types, most noticeably with redshift. The degraded but still respectable performance of the MLP on classification both with and without redshift shows that these supernova types can indeed be differentiated in some hyperdimensional space by a neural network, and that SCONE in particular possesses the required discriminatory power for this task.

### 3.5. Bright Supernovae

Bright supernovae, defined as supernovae with last included $r$-band observation $r < 20$ mag, were identified from both the $t_{\text{trigger}} + 0$ and $t_{\text{trigger}} + 5$ data sets. Since fewer (and likely dimmer) observations were included for each supernova in the $t_{\text{trigger}} + 0$ data set, there are much fewer examples of bright supernovae than in the $t_{\text{trigger}} + 5$ data set. The bright

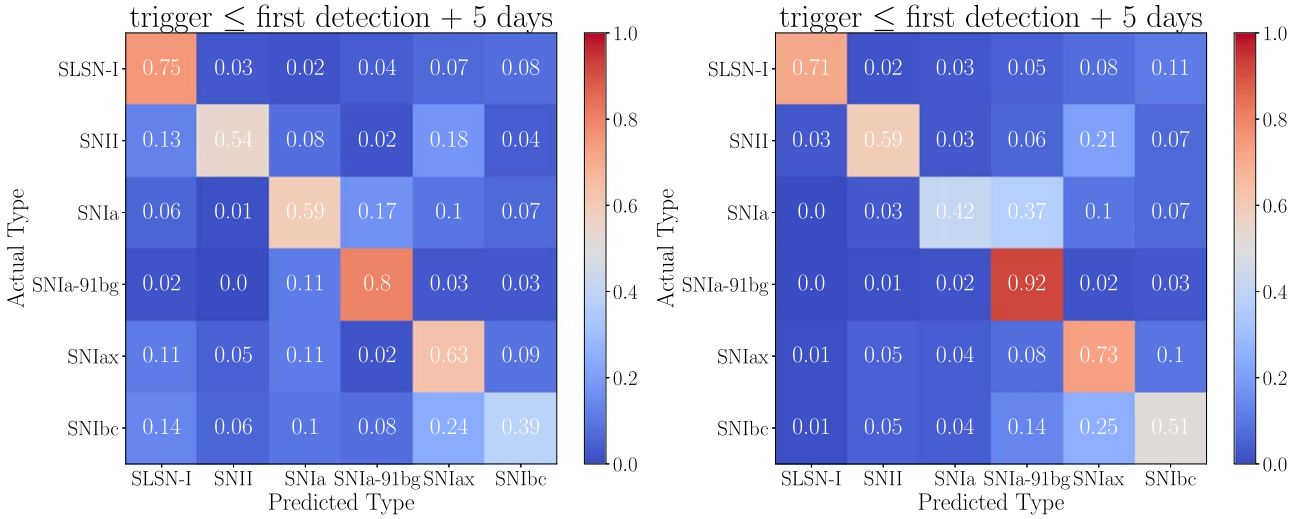


**Figure 8.** Normalized confusion matrices produced by SCONE without (left) and with (right) redshift for the $t_{\text{trigger}} \leqslant t_{\text{first detection}} + 5$ subset of the $t_{\text{trigger}} + 0$ test set. This cut ensures that the light curves used for performance evaluation are not given substantially more information than those created with the first detection trigger definition.

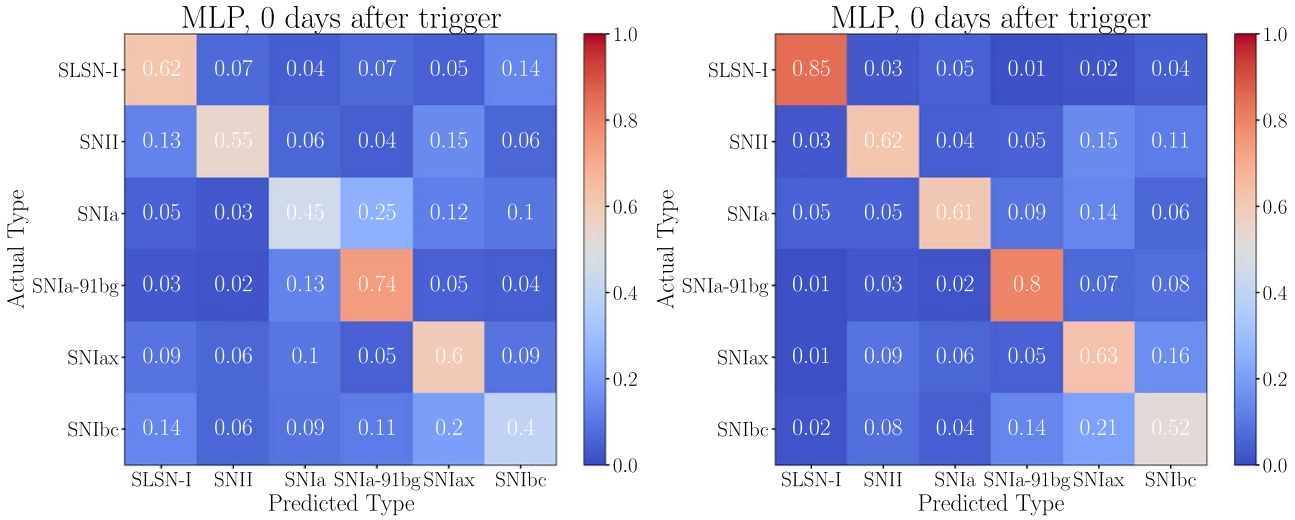


**Figure 9.** Normalized confusion matrices produced by the baseline MLP model without (left) and with (right) redshift for the $t_{\text{trigger}} + 0$ test set (heatmaps created from light curves truncated at the date of trigger).

supernovae subsets of these data sets are referred to as the "bright $t_{\text{trigger}} + N$ data sets."

To evaluate the performance of SCONE on identifying bright supernovae at early epochs, the model was trained on a regular class-balanced $t_{\text{trigger}} + N$ training set, prepared as described in Section 2.4, combined with 40% of the bright $t_{\text{trigger}} + N$ data set. The results of testing the trained SCONE model on the bright $t_{\text{trigger}} + N$ data sets are shown in Figure 10. These confusion matrices, like the ones in Figure 4, are colored by efficiency score. However, since the data set is not class-balanced, the overlaid values are absolute (non-normalized) to preserve information on the relative abundance of each type. Thus, an efficiency (purity) score for each type can be calculated by dividing each main diagonal value by the sum of the values in its row (column). The overall accuracies as well as the total number of SNe in each data set are summarized in Table 4.

The benefits of redshift information are much more pronounced for certain types than others. As also noted in analyses of Figures 4 and 6, the quantity of SNIbc misclassified as SLSN-I was significantly reduced in results from SCONE with redshift information. At the date of trigger, 44.4% of SNIbc were misclassified as SLSN-I without redshift. This contamination rate was reduced to only 3.7% with redshift. However, classification of bright SNIa seems relatively unaffected by the presence of redshift information. Five days after trigger, SNIa were classified with an efficiency/accuracy of 98.6% and a purity score of 98.1% without redshift, and 97.4% efficiency/accuracy and 99.1% purity with redshift.

### 3.6. Mixed Data Set

Training on the $t_{\text{trigger}} + N$ data sets represents one way of deploying SCONE for real-world transient alert applications, while training on a mixed data set is a much less computationally expensive alternative. On one hand, testing a $t_{\text{trigger}} + N$-trained model on a $t_{\text{trigger}} + N$ test set yields the best classification accuracies. However, this approach requires the creation of separate data sets for each choice of $N$, which could be an expensive initial time investment, depending on the number of

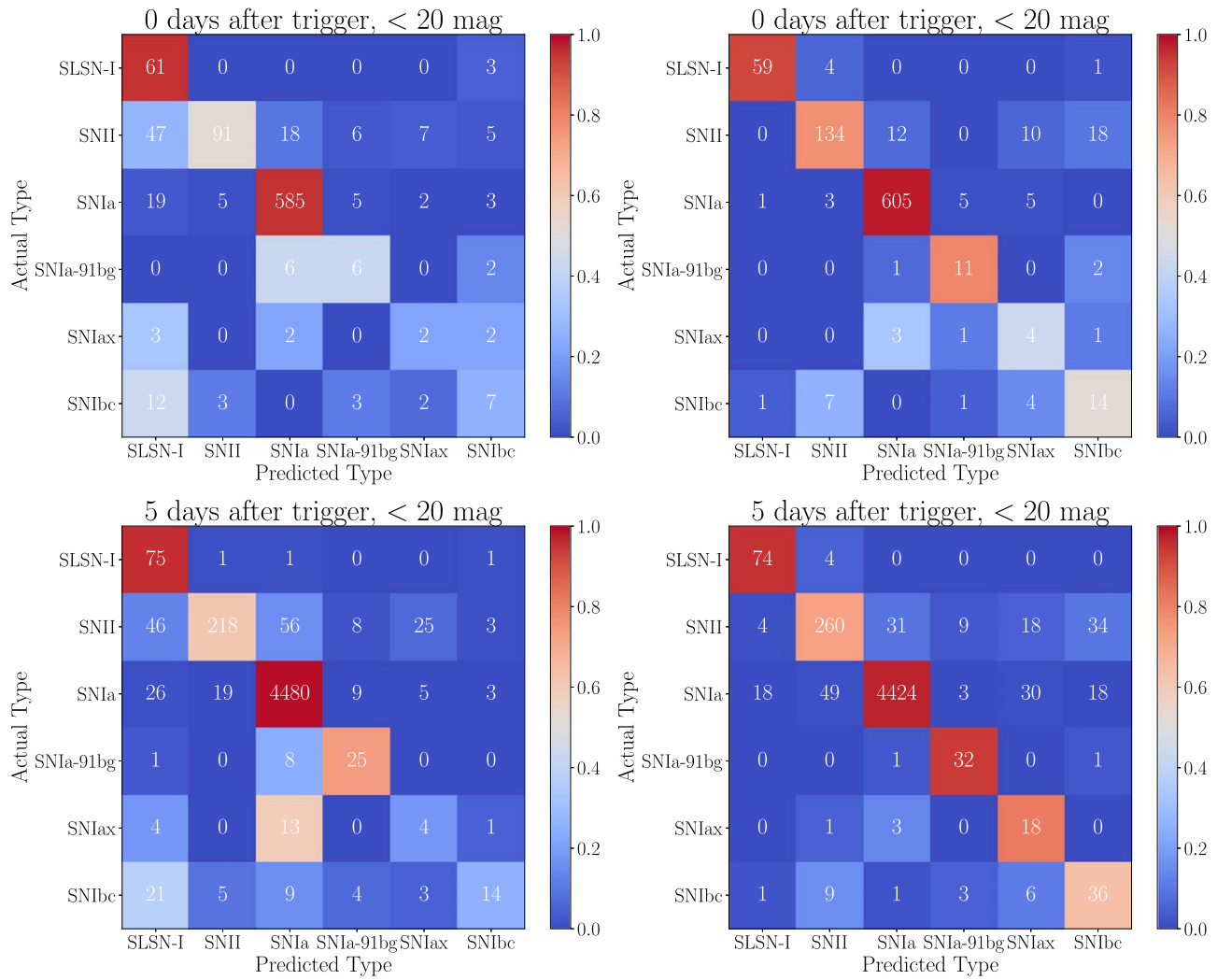


**Figure 10.** Early epoch confusion matrices with (right) and without (left) redshift for the bright supernovae (<20 magnitude) in each $t_{\text{trigger}} + N$ data set. SCONE was trained with a class-balanced $t_{\text{trigger}} + N$ training set combined with 40% of <20 magnitude supernovae. These confusion matrices were created by testing the trained SCONE model on the full <20 magnitude supernovae data set. The confusion matrices are colored according to normalized accuracies, as in Figure 4, and are overlaid with absolute (non-normalized) values since the data set is imbalanced.

**Table 4**
Test Accuracies with and without Redshift Information for the Bright Data Sets

| | Total | Accuracy without $z$ | Accuracy with $z$ |
|---|---|---|---|
| bright $t_{\text{trigger}} + 0$ | 907 | 82.91% | 91.18% |
| bright $t_{\text{trigger}} + 5$ | 5088 | 94.65% | 95.2% |

data sets and size of each data set (see Section 2.6 for computational requirements for heatmap creation). In this work, only five data sets ($N = 0$, 5, 15, 25, 50) were created, but perhaps $N = 0, 1, \ldots, 50$ will be needed to accurately classify real-world transient alerts with any number of nights of photometry. Training on a mixed data set, where each heatmap is created with a random number of nights of photometry after trigger, is a viable alternative for resource- or time-constrained applications.

To directly compare the performance of SCONE trained on the mixed data set and the $t_{\text{trigger}} + N$ data sets, the model trained on a mixed data set was tested on each individual $t_{\text{trigger}} + N$ data set. The accuracies over time split by SN type are summarized in Figure 11. Compared to the results of SCONE trained and tested on each individual $t_{\text{trigger}} + N$ data set (Figure 3), the accuracies are lower but still respectable. The performance at the date of trigger is the most dissimilar, with average accuracy 74% with $z$ for a model trained on $t_{\text{trigger}} + 0$ and 64% with mixed. The performance of the mixed-trained model performs similarly to the $t_{\text{trigger}} + N$-trained model by 5 days after trigger, however, with both averaging just under 80% with $z$. The AUCs over time split by SN type are shown in Figure 12. These AUC plots are comparable to the $t_{\text{trigger}} + N$ AUCs in Figure 6, indicating that the performances of both models are comparable when averaged over all values of the prediction threshold $p$. However, the predicted class for categorical classification is not typically calculated with respect to a threshold; rather, it is defined as the class with the highest prediction confidence for each example. Thus, the AUCs are analogous to analyzing the performance on each type as its own binary classification problem, resulting in slight discrepancies from the accuracies.

### 3.7. Comparison with Existing Literature

At the time of this writing, the only work in existing literature with a similarly strong focus on early photometric classification of supernovae is RAPID (Muthukrishna et al. 2019,

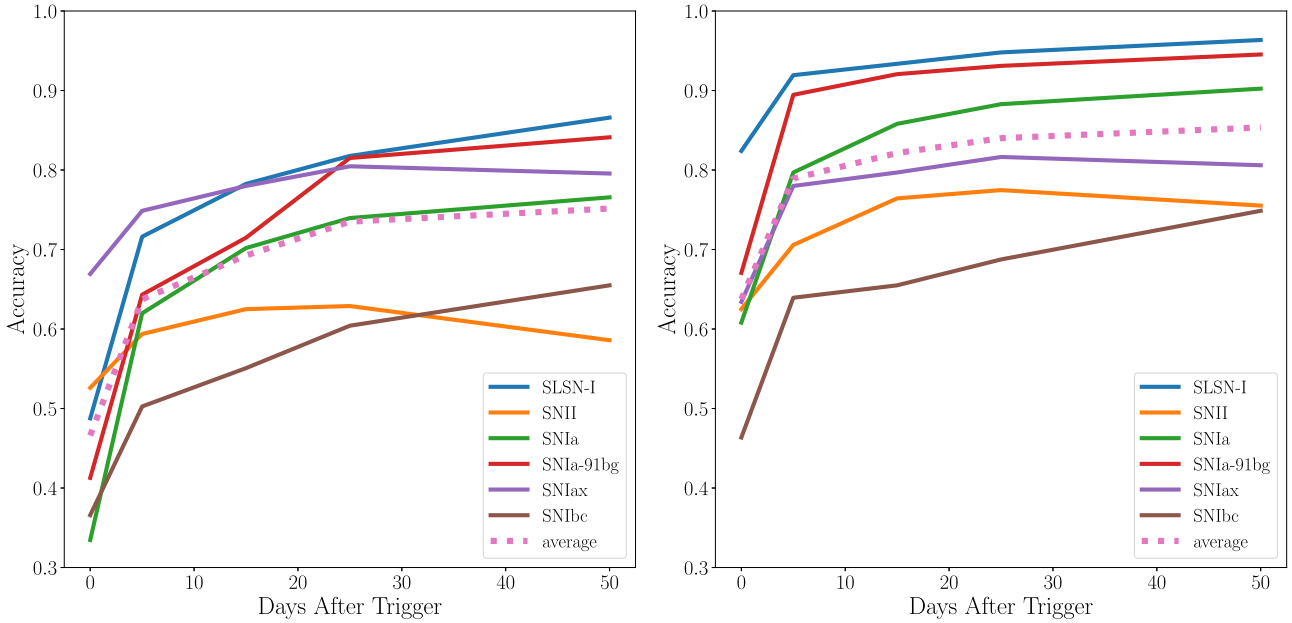


**Figure 11.** Test set accuracy/efficiency without (left) and with (right) redshift over time for SCONE trained on the mixed data set and tested on each individual $t_{\text{trigger}} + N$ data set. The values used in these plots correspond with the diagonals on a normalized confusion matrix.

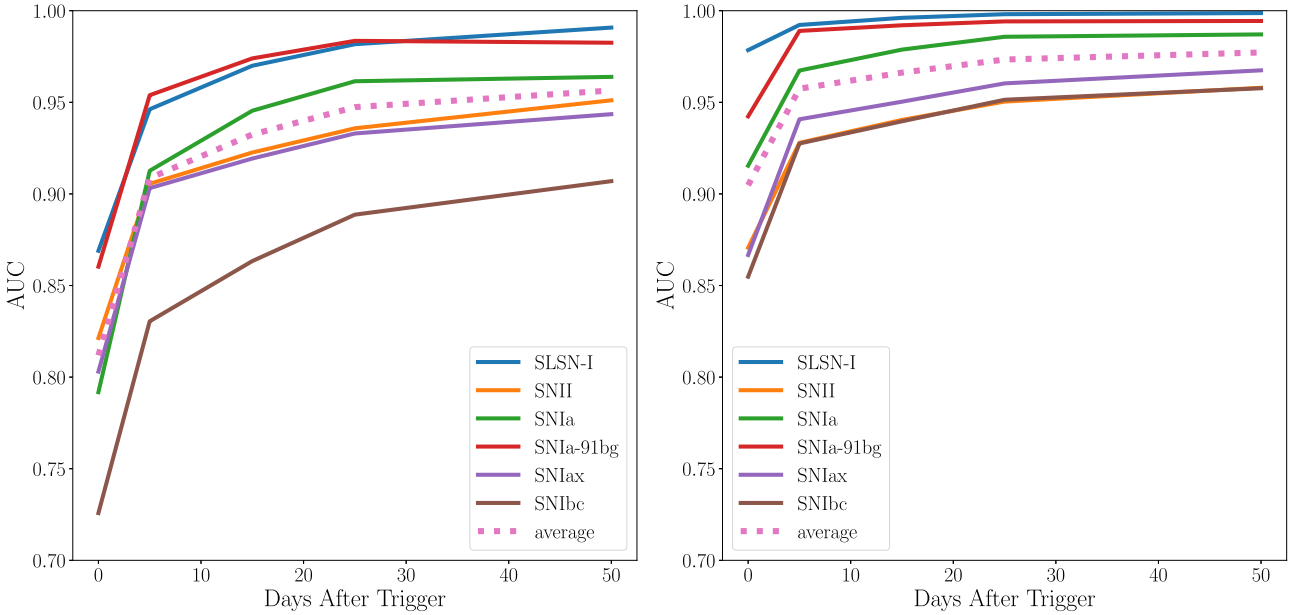


**Figure 12.** Area under the ROC curve (AUC) without (left) and with (right) redshift over time for SCONE trained on the mixed data set and tested on each individual $t_{\text{trigger}} + N$ data set.

hereafter M19), a GRU RNN approach to photometric transient classification that differentiates between 12 transient types, including seven supernova types. RAPID differs in several significant ways from the data, model, and results presented in this work. We highlight some of the differences in these two works below.

### *3.7.1. Comparison of Methods*

The most obvious difference is in the type of neural network architecture used for classification. RAPID uses a unidirectional RNN architecture, which is designed to learn from time-series data chronologically. SCONE employs a convolutional neural network architecture, which is most commonly used for image recognition tasks. In this instance, however, SCONE is designed to read in data chronologically. Convolutional layers in a CNN work by computing functions on a "sliding window" of the input image, thereby allowing the model to learn small-scale structures in the image. This window, or the convolutional kernel, is typically a small square chosen to match the characteristic length scale of structures in the images. SCONE's convolutional kernel, however, is chosen to span the full height of the input heatmap, resulting in a window that slides chronologically along the horizontal, or time, axis.

M19 trained and tested on a data set of simulated Zwicky Transient Facility light curves, which have *g*- and *r*-band photometry, compared to the LSST light curves used in this work, with *ugrizY* photometry bands. In addition to the six

supernova types that this work focuses on, M19 includes four rare transient classes (pair instability supernovae, intermediate luminosity transients, calcium-rich gap transients (CART), and tidal disruption events) as well as point-Ia and KN.

Other differences include the addition of a "pre-explosion" class, rest- versus observer-frame time intervals, and the choice of trigger definition. M19 chooses to include an additional class, "pre-explosion," to describe examples at time steps prior to the occurrence of the transient event. M19 also converts time intervals out of the observer frame by dividing by $1+z$, which is not done in this work in order to ensure that mistakes in redshift estimates will not be propagated to affect the light-curve data. Finally, M19 uses the first-detection trigger date definition, while this work defines $t_{\rm trigger}$ to be the date of the second detection.

#### 3.7.2. Comparison of Results

The results of SCONE classification with redshift (right side panels of Figures 4–6) is used to compare with RAPID's results, as RAPID also incorporates redshift information. As described in the previous section, this work differs in many ways from M19, and the following comparison does not account for these differences; a rigorous comparison of the two models against a single data set is left to a future work.

Most notably, SCONE improves upon RAPID's SNIbc and SNII classification accuracy, while RAPID performs very well at classifying early-time SNIa. From Figure 7 of M19, 12% of SNIbc are correctly classified two days after detection, compared to SCONE's 54% accuracy at the date of trigger. In RAPID's results, 30% of true SNIbc are misclassified as CART, which is not included in the data sets in this work. The second- and third-largest contaminants (SNIax at 19%, then SNIa and SNIa-91bg at 8% each), are both part of this analysis. From Figure 4, we find that SNIax and SNIa-91bg are also major contaminants for SCONE at 23% and 11%, respectively, at the date of trigger and 16% and 4%, respectively, five days after trigger. However, there is no significant contamination from SNIa, with contamination rates at 4% on the date of trigger and 1% five days after trigger.

Two days after detection, SNII is classified at 7% accuracy by RAPID, compared to 64% accuracy at the date of trigger by SCONE. The primary contaminant of SNII for RAPID 2 days after detection is SNIa at 21%, which is not reflected in SCONE's results, where the contamination rate is 6% at the date of trigger and 3% five days after trigger. The second-largest contaminant, SLSN-I, is also not an issue in SCONE's SNII classification. Surprisingly, the improvement over time of RAPID's SNII classification accuracy outpaces its SNIbc classification accuracy, as it is able to achieve 49% accuracy on SNII 40 days after detection, compared to 31% accuracy on SNIbc.

While SCONE's SNIa classification accuracy slowly climbs from 77% at the date of trigger to 93% 50 days after trigger, RAPID is able to classify SNIa at 88% accuracy almost immediately after detection. A future direct comparison will aid in concluding whether this discrepancy is due to differences in the data sets, such as M19's exclusion of $z \geqslant 0.5$ objects, or something more fundamental to the model architectures.

## 4. Conclusions

Our ability to observe the universe has improved in leaps and bounds over the past century, allowing us to find new and rare transient phenomena, enrich our understanding of transient physics, and even make cosmological discoveries aided by observational data. Our photometric observing capabilities greatly outpace the rate at which we can gather the associated spectroscopic information, resulting in a vast trove of photometric data sparsely annotated by spectroscopy. In the era of large-scale sky surveys, with millions of transient alerts per night, an accurate and efficient photometric classifier is essential not only to make use of the photometric data for science analysis, but also to determine the most effective spectroscopic follow-up program early on in the life of the transient.

In this work, we presented SCONE's performance classifying simulated LSST early-time supernova light curves for SN types Ia, II, Ibc, Ia-91bg, Iax, and SLSN-I. As an approach based on neural networks, SCONE avoids the time-intensive manual process of feature selection and engineering, and requires only raw photometric data as input. We showed that the incorporation of redshift estimates as well as errors on those estimates significantly improved classification accuracy across the board, and was especially noticeable at very early times. Notably, this is the first application of convolutional neural networks to this problem.

SCONE was tested on three types of data sets: data sets of light curves that were truncated at 0, 5, 15, 25, and 50 days after trigger ($t_{\rm trigger}+N$ data sets); bright ($<$20 magnitude) subsets of the $t_{\rm trigger}+\{0, 5\}$ data sets; and a data set of light curves truncated at a random number of nights between 0 and 50 ("mixed"). Without redshift, SCONE was able to classify $t_{\rm trigger}+0$ light curves with 60% overall accuracy, which increases to 82% at 50 days after trigger. SCONE with redshift information starts at 74% overall accuracy at the date of trigger and improves to 89% 50 days after trigger. Confusion matrices, ROC plots, and accuracy over time as well as AUC over time plots of results with and without redshift were presented to better understand classification performance and identify areas of improvement. For the bright subsets, overall accuracy is $>$90% at the date of trigger with redshift and over 80% without. These results improve to around 95% accuracy both with and without redshift by five days after trigger. The overall accuracy over time of a model trained on a mixed data, tested on the $t_{\rm trigger}+N$ data sets, shows some degradation in accuracy at very early epochs, but may be a worthwhile lightweight alternative to the more resource-intensive process of creating many $t_{\rm trigger}+N$ data sets.

We showed that SCONE's performance with redshift is competitive with existing work on early classification, such as M19, while improving on computational time requirements. SCONE has a lightweight preprocessing step and can achieve impressive performance with a small training set. It requires only hundredths of a second to preprocess each light curve into a heatmap, and seconds for each training epoch on GPU. This makes SCONE a great candidate for incorporation into alert brokers for LSST and future wide-field sky surveys.

In future work, we plan to apply this model to real data to further validate the approach. We also plan to extend SCONE to classify both full-duration and early light curves for more transient and variable classes in the PLAsTiCC simulations.

This work was supported by DOE grant DE-FOA-0002424, NASA Grant NNH15ZDA001N-WFIRST, and NSF grant AST-2108094. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a

U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under contract No. DE-AC02-05CH11231.

## ORCID iDs

Helen Qu https://orcid.org/0000-0003-1899-9791
Masao Sako https://orcid.org/0000-0003-2764-7093

## References

Aguirre, C., Pichara, K., & Becker, I. 2019, MNRAS, 482, 5078
Allam, T. J., & McEwen, J. D. 2021, arXiv:2105.06178
Armstrong, P., Tucker, B. E., Rest, A., et al. 2021, MNRAS, 507, 3125
Boone, K. 2019, AJ, 158, 257
Boone, K. 2021, AJ, 162, 275
Carrasco-Davis, R., Reyes, E., Valenzuela, C., et al. 2021, AJ, 162, 231
Charnock, T., & Moss, A. 2017, ApJL, 837, L28
Filippenko, A. V. 2005, in ASP Conf. Ser. 332, The Fate of the Most Massive Stars, ed. R. Humphreys & K. Stanek (San Francisco, CA: ASP), 34
Freedman, W. L., Madore, B. F., Hatt, D., et al. 2019, ApJ, 882, 34
Frieman, J. A., Bassett, B., Becker, A., et al. 2008, AJ, 135, 338
Guillochon, J., Nicholl, M., Villar, V. A., et al. 2018, ApJS, 236, 6
Guy, J., Sullivan, M., Conley, A., et al. 2010, A&A, 523, A7
Hložek, R., Ponder, K. A., Malz, A. I., et al. 2020, arXiv:2012.12392
Hochreiter, S., & Schmidhuber, J. 1997, Neural Comp., 9, 1735
Hornik, K., Stinchcombe, M., & White, H. 1989, NN, 2, 359
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Jha, S. W. 2017, Handbook of Supernovae (Berlin: Springer), 375
Karpenka, N. V., Feroz, F., & Hobson, M. P. 2013, MNRAS, 429, 1278
Kasen, D., & Bildsten, L. 2010, ApJ, 717, 245
Kessler, R., Bassett, B., Belov, P., et al. 2010, PASP, 122, 1415
Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009, PASP, 121, 1028
Kessler, R., Conley, A., Jha, S., & Kuhlmann, S. 2010a, arXiv:1001.5210
Kessler, R., Guy, J., Marriner, J., et al. 2013, ApJ, 764, 48
Kessler, R., Narayan, G., Avelino, A., et al. 2019, PASP, 131, 094501
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Kodi Ramanah, D., Arendse, N., & Wojtak, R. 2021, arXiv:2107.12399
Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, in Proc. 25th International Conf. on Neural Information Processing Systems, Vol. 1 (Red Hook, NY: Curran Associates Inc.), 1097
LeCun, Y., Boser, B., Denker, J. S., et al. 1989, Neural Comp., 1, 541
LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Proc. IEEE, 86, 2278
Liu, H., Dai, Z., So, D. R., & Le, Q. V. 2021, arXiv:2105.08050
Modjaz, M., Blondin, S., Kirshner, R. P., et al. 2014, AJ, 147, 99
Möller, A., & de Boissière, T. 2020, MNRAS, 491, 4277
Moss, A. 2018, arXiv:1810.06441
Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, PASP, 131, 118002
Naul, B., Bloom, J. S., Pérez, F., & van der Walt, S. 2018, NatAs, 2, 151
Nicholl, M., Guillochon, J., & Berger, E. 2017, ApJ, 850, 55
Pasquet, J., Pasquet, J., Chaumont, M., & Fouchez, D. 2019, A&A, 627, A21
Perets, H. B., Gal-Yam, A., Mazzali, P. A., et al. 2010, Natur, 465, 322
Perlmutter, S., Turner, M. S., & White, M. 1999, PhRvL, 83, 670
Pierel, J. D. R., Rodney, S., Avelino, A., et al. 2018, PASP, 130, 114504
Poznanski, D., Maoz, D., & Gal-Yam, A. 2007, AJ, 134, 1285
Pursiainen, M., Childress, M., Smith, M., et al. 2018, MNRAS, 481, 894
Qu, H. 2021, helenqu/scone: early lightcurve classification release, v1.1.0, Zenodo, doi:10.5281/zenodo.5602043
Qu, H., Sako, M., Möller, A., & Doux, C. 2021, AJ, 162, 67
Richards, J. W., Homrighausen, D., Freeman, P. E., Schafer, C. M., & Poznanski, D. 2012, MNRAS, 419, 1121
Riess, A. G. 1998, AJ, 116, 1009
Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, ApJ, 876, 85
Sako, M., Bassett, B., Becker, A., et al. 2008, AJ, 135, 348
Sako, M., Bassett, B., Connolly, B., et al. 2011, ApJ, 738, 162
Sánchez-Sáez, P., Reyes, I., Valenzuela, C., et al. 2021, AJ, 161, 141
Smith, M., Sullivan, M., Wiseman, P., et al. 2020, MNRAS, 494, 4426
Sollerman, J., Yang, S., Schulze, S., et al. 2021, A&A, 655, A105
Sullivan, M., Howell, D. A., Perrett, K., et al. 2006, AJ, 131, 960
The PLAsTiCC team, Allam, T. Jr., Bahmanyar, A., et al. 2018, arXiv:1810.00001
Tolstikhin, I., Houlsby, N., Kolesnikov, A., et al. 2021, arXiv:2105.01601
Villar, V. A., Berger, E., Metzger, B. D., & Guillochon, J. 2017, ApJ, 849, 70
Villar, V. A., Cranmer, M., Contardo, G., Ho, S., & Yao-Yu Lin, J. 2020a, arXiv:2010.11194
Villar, V. A., Hosseinzadeh, G., Berger, E., et al. 2020b, ApJ, 905, 94
Woosley, S. E., Pinto, P. A., Martin, P. G., & Weaver, T. A. 1987, ApJ, 318, 664
Zeiler, M. D., & Fergus, R. 2014, in European Conf. on Computer Vision (Cham: Springer), 818