Bayesian Neural Networks for classification tasks in the Rubin big data era

Anais Möller [1] **Thibault de Boissière** [2]

Abstract

Upcoming surveys such as the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST) will detect up to 10 million time-varying sources in the sky every night for ten years. This information will be transmitted in a continuous stream to brokers that will select the most promising events for a variety of science cases using machine learning algorithms. We study the benefits and challenges of Bayesian Neural Networks (BNNs) for this type of classification tasks. BNNs are found to be accurate classifiers which also provide additional information: they quantify the classification uncertainty which can be harnessed to analyse this upcoming data avalanche more efficiently.

1. Introduction

We are entering into a new era of big data time-domain astronomy. Upcoming surveys such as Vera C. Rubin Observatory Legacy Survey of Space Time (LSST) will detect up to ten million time-domain events every night for over a decade. LSST will emit an alert stream with time-domain even data within minutes of observation. The Rubin Community brokers will then receive that stream in real-time. Brokers will enrich and filter these alerts to select the most promising candidates for a variety of science cases. The selected LSST brokers are ALERCE (Förster et al., 2021), AMPEL (Nordin et al., 2019), BABAMUL, ANTARES (Narayan et al., 2018), FINK (Möller et al., 2021), LASAIR (Smith et al., 2019) and PITT-GOOGLE. Many of these brokers are currently processing the Zwicky Transient Facility (ZTF) alert stream which is an order of magnitude less than what is expected from Rubin.

Classification algorithms are a core part of brokers. They provide scores that can be used to select candidates for specific astrophysical phenomena such as supernovae, kilonovae, microlensing, AGNs, and many others. In recent years, a wide variety of classification algorithms have been developed for time-domain astronomy which have shown excellent performance in classification tasks (Leoni et al., 2021; Muthukrishna et al., 2019; Villar et al., 2019; Villar & et al., 2020; Godines et al., 2019; Ishida et al., 2019)

Bayesian Neural Networks are a promising classification method that provide classification scores as well as uncertainties that can reflect the model's confidence in the prediction. BNNs go beyond point estimates and yield a distribution of classification probabilities. The final prediction is typically computed as the mean probability of this distribution. The standard deviation of this distribution is typically used as an estimation of the prediction uncertainty. The astronomical community has recently started to use BNNs for classification tasks (Walmsley et al., 2020; Möller & de Boissière, 2019; Möller et al., 2022).

In this work, we use the SUPERNNOVA (SNN) classification algorithm (Möller & de Boissière, 2019) to evaluate the performance and interpretability of BNNs for the new Rubin era. Since the Rubin observing strategy is yet to be defined, we use the Dark Energy Survey (DES) SN fields as a proxy for Rubin's Deep Drilling Fields. Our benchmark task is the classification of type Ia vs non Ia supernovae light-curves. Supernovae are bright stellar explosions that fade away within weeks.

We use a simulated dataset for training and evaluation. These simulations contain realistic DES light-curves from Type Ia models, peculiar Ia and core-collapse supernovae. An example of the simulated light-curves can be seen in Figure 1. More details on these simulations can be found in (Möller et al., 2022). Additionally, we use real light-curves of type Ia supernovae candidates observed by the Zwicky Transient Facility (ZTF) for evaluation in Section 4.2. These light-curves where obtained using the Fink broker API[1].

1.1. Bayesian Neural Networks (BNNs)

Scientific application of machine learning methods often require estimating the model's predictive uncertainty. A

[1]Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Mail Number H29, PO Box 218, 31122 Hawthorn, VIC, Australia. [2]Canva inc., Sydney, Australia. Correspondence to: Anais Möller <amoller@swin.edu.au>.

Proceedings of the 39^{th} International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.

[1]https://fink-portal.org/api

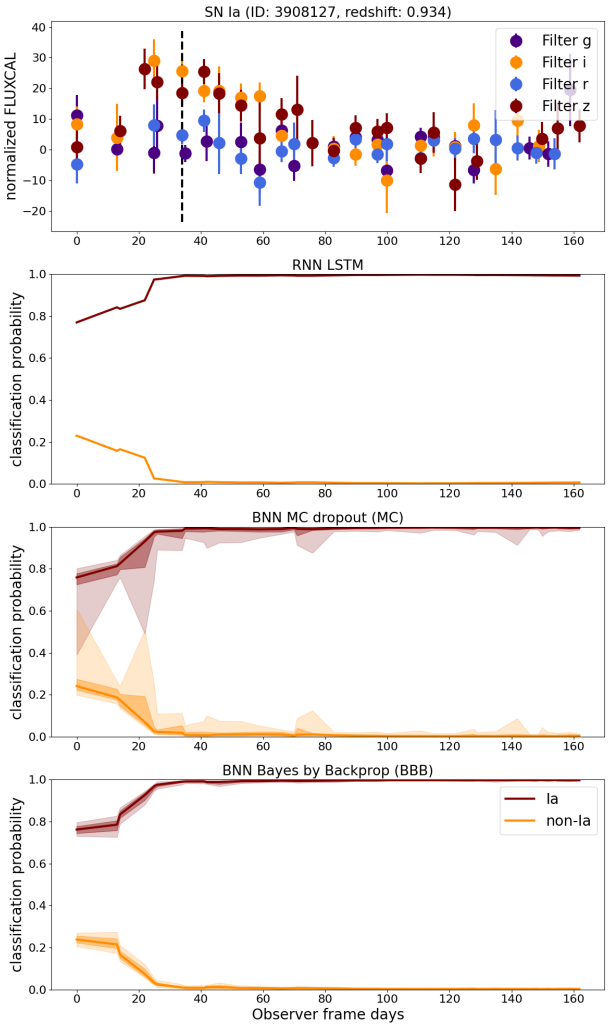

Figure 1. Simulated DES light-curve and SNN classifications using three different neural networks: RNN, BNN MC, and BNN BBB. Top row shows the SN light-curve (normalised flux in band-passes g, r, i, z; time in Observer Frame days) and its time of maximum brightness in a dashed line. Bottom rows shows the classification scores for each method (SN Ia: maroon, non SN Ia: orange). The BNN methods provide classification uncertainties (shadowed regions show 68 per cent and 95 per cent contours).

popular way to do so is to cast NN training under a bayesian light where the goal is to learn a probability distribution over possible neural networks. This problem is typically untractable analytically. In practice, approximate methods, such as variational inference, are used. A review of BNNs and their use in astronomy can be found in (Charnock et al., 2020).

We use two BNN implementations to approximate the posterior distribution of weights: MC dropout (MC; Gal & Ghahramani, 2015) and Bayes by Backprop (BBB; Fortunato et al., 2017). MC provides a Bayesian interpretation of recurrent dropout when the dropout mask is the same at all time steps. BBB learns an approximate posterior distribution of weights using variational inference. Both methods have been previously implemented and tested on simulations in SUPERNNOVA (Möller & de Boissière, 2019).

For both methods, to obtain the classification probability distribution, we sample the predictions from our BNN 50 times. In the following we compute the classification probability, $\widehat{p}_i$ for a given light-curve, $\mathbf{x}_i$ as:

$$\widehat{p}_i = mean\Big\{p_j(\mathbf{x}_i)\Big\}_{i=1:n_s} \tag{1}$$

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 $\mathbf{x}_i$.

To evaluate classification uncertainties we compute the model uncertainty for a given light-curve $\mathbf{x}_i$ as:

$$\widehat{\sigma}_i = std\Big\{p_j(\mathbf{x}_i)\Big\}_{i=1:n_s} \tag{2}$$

where std is the sample standard deviation.

2. Calibration for BNNs

Classification probabilities that reflect the real likelihood of events being correctly assigned to a target are said to be calibrated. Calibration is extremely important in the Rubin context to carry out precision cosmology analyses.

We use reliability diagrams (DeGroot & Fienberg, 1983) to evaluate our model's calibration in Figure 2. We evaluate this calibration with complete light-curves spanning $\approx 100-150$ days. We find all models to be close to perfectly calibrated with some excess on the fraction of positives at low probability in particular for BNN BBB.

3. Early classification accuracy

To identify promising time-domain events swiftly, it is necessary to use algorithms that allow accurate classification with only a couple of observations. This is increasingly necessary as surveys such as Rubin will detect millions of transients per night which need to be disentangled together with the scarce follow-up resources which need to be optimized.

We evaluate the performance of a traditional RNN and the two BNNs in SNN for early classification (around maximum brightness) in Figure 3. We consider a light-curve classified as type Ia if its classification probability is $\widehat{p}_{Ia} > 0.5$.

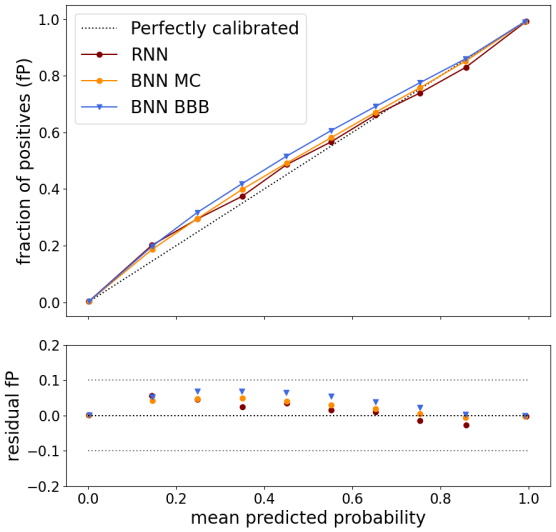

Figure 2. Calibration: reliability diagrams for an RNN and two BNNs: MC dropout (MC) and Bayes by Backprop (BBB).

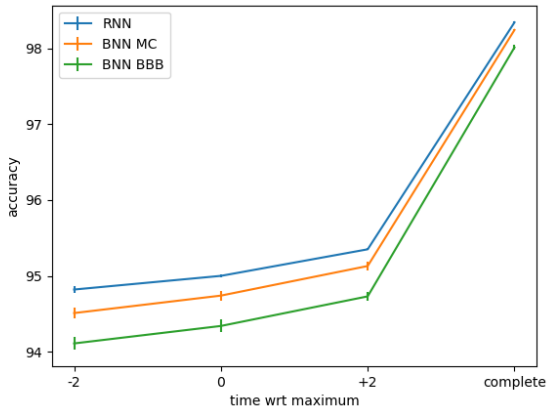

Figure 3. Accuracy of type Ia vs. non Ia supernova classification with respect to time of maximum brightness in the rest frame of a supernova. We evaluate an RNN and two BNNs: MC dropout (MC) and Bayes by Backprop (BBB) for the same task. Experiments are repeated 5 times with different random seeds. The reported accuracy is the mean accuracy. The reported error bar is the standard deviation.

Early classification is found to be highly-accurate. As more observations become available we also find that the accuracy increases. We note that BNNs are shown to have slightly lower accuracies than a LSTM RNN. This may be improved with careful tuning of the BNN prior parameters.

In the following, we continue exploring BNNs in this classification task with complete light-curves.

4. Interpretability: Out of distribution (OOD) or anomalies

In this Section we present two tests designed to evaluate the interpretability of BNN predictions for OOD or anomalies.

4.1. Entropy

Entropy has been used as a proxy for the model's confidence in its predictions and thus an interesting metric to evaluate BNNs on (Fortunato et al., 2017). Confident predictions should yield low entropy. For a dataset $\mathcal{D} : [\mathbf{x}_1, ...\mathbf{x}_N]$ with N light-curves and a classification model m, the entropy of $\mathcal{D}$ under m is defined as:

$$H_m[\mathcal{D}] = \sum_{i=1}^{N} p(\mathbf{y}_i|\mathbf{x}_i) log\left(\frac{1}{p(\mathbf{y}_i|\mathbf{x}_i)}\right). \quad (3)$$

where $p(\mathbf{y}_i|\mathbf{x}_i)$ is the classification probability given the light-curve i.

Table 1. Delta entropy between out-of-distribution events and supernova classification.

MODEL	RANDOM	REVERSE	SHUFFLE
RNN	-0.02	0.01	0.03
MC	-0.02	0.05	0.1
BBB	-0.02	0.08	0.06

For two given sets of predictions, we can define the *entropy gap* ΔH by:

$$\Delta H = \quad \overline{H_{m1}}[\mathcal{D}] - \overline{H_{m2}}[\mathcal{D}] \quad (4)$$

where we evaluate the entropy gap over the same dataset for two given models (m_1, m_2). It is expected that predictions on OOD datasets would have a large entropy if the algorithm is robust. Thus the ΔH between these OOD and a SN-like dataset should be positive.

We generate three different types of OOD events: time reversed light-curves, randomly shuffled light-curves and random fluxes. We evaluate the entropy of these predictions in Table 1. For reverse and shuffled light-curves we find, as expected, a high entropy gap when compared to supernova light-curve classification. This behaviour is not seen in random light-curves which may be attributed to their possible resemblance to noisy low signal-to-noise supernovae.

Table 2. Model classification accuracy, duration of classification for 1000 light-curves and number of samples used for the results.

MODEL	ACCURACY	TIME(S)	NUMBER SAMPLING
RNN	98.29	10	NA
BBB	98.09	315	50
BBB	97.89	129	20
BBB	97.89	35	5
MC	98.39	342	50
MC	98.39	142	20
MC	98.39	39	5

4.2. DES-like vs. ZTF light-curves

We now explore the predictions obtained for a DES-like dataset, which is similar to the training set, and SNe Ia candidates from ZTF selected by the FINK broker. These candidates have been selected as probable SNe Ia by classifiers such as SUPERNNOVA trained on ZTF-like data.

MC models (resp. BBB) trained on DES-like simulations of type Ia supernovae obtain on average a median classification probability of 0.9 (resp. 0.9) and standard deviation of the classification probability distribution of 0.002 (resp. 0.001 for BBB). Contrast this with an average classification probability of 0.6 (resp. 0.6) and standard deviation of 0.19 (resp. 0.27) on the early SNe Ia candidates from the Fink broker. Clearly, the models are less confident in their prediction for this new dataset showing an expected interpretable behaviour.

5. Scalability

As Rubin's data volume will be unprecedented, we require fast classification algorithms. SNN RNN has been benchmarked in the FINK broker, classifying 2500 light-curves per second/core.

Here, we use a simpler benchmark to assess the classification accuracy of BNNs with respect of the number of samples of their probability distribution. We evaluate the classification of 1000 light-curve using one node and the embedded SNN database.

In Table 2 we show the classification accuracy evolution as a function of number of samples of the probability distribution. We find the decrease in accuracy is small (< 0.2 for BBB while < 0.01 for MC) compared to the reduction of classification time in the algorithms (up to one order of magnitude). Thus, for brokers it could be envisaged to reduce the number of samples to provide fast classification with these algorithms. We note that an evaluation on the impact of this sampling in the estimated uncertainties is left for future work.

6. Summary

The use of Bayesian Neural Networks (BNNs) for classification has began within the astronomical community. In this work, we explore the use of BNNs in classification tasks. We performed calibration and scalability benchmarks and explored the interpretability of BNNs outputs.

BNNs are found to be highly accurate in classification tasks with calibrated probabilities. We have also found quantitative evidence that the prediction confidence of our BNNs decreases for out-of-distribution events.

With the advent of large surveys discovering thousands of transients every night, it will be imperative to prioritize follow-up using partial light-curve classification. SUPERNNOVA BNNs achieve high accuracy on this challenging task. Future work will expand on the study of robustness of BNNs, in particular supplementing probability thresholds with prediction uncertainties to improve classification.

Software and Data

Software used in this work is open source and available in GitHub. Trained models are available on request.

Acknowledgements

We thank the anonymous reviewers for feedback to improve this work.

This work was developed within the Fink community and made use of the Fink community broker resources. Fink is supported by LSST-France and CNRS/IN2P3.

References

Charnock, T., Perreault-Levasseur, L., and Lanusse, F. Bayesian Neural Networks. *arXiv e-prints*, art. arXiv:2006.01490, June 2020.

DeGroot, M. H. and Fienberg, S. E. The comparison and evaluation of forecasters. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 32(1/2):12–22, 1983. ISSN 00390526, 14679884. URL http://www.jstor.org/stable/2987588.

Förster, F., Cabrera-Vives, G., Castillo-Navarrete, E., Estévez, P. A., Sánchez-Sáez, P., Arredondo, J., Bauer, F. E., Carrasco-Davis, R., Catelan, M., Elorrieta, F., Eyheramendy, S., Huijse, P., Pignata, G., Reyes, E., Reyes, I., Rodríguez-Mancini, D., Ruz-Mieres, D., Valenzuela, C., Álvarez-Maldonado, I., Astorga, N., Borissova, J., Clocchiatti, A., De Cicco, D., Donoso-Oliva, C., Hernández-García, L., Graham, M. J., Jordán, A., Kurtev, R., Mahabal, A., Maureira, J. C., Muñoz-Arancibia, A., Molina-Ferreiro, R., Moya, A., Palma, W., Pérez-

Carrasco, M., Protopapas, P., Romero, M., Sabatini-Gacitua, L., Sánchez, A., San Martín, J., Sepúlveda-Cobo, C., Vera, E., and Vergara, J. R. The Automatic Learning for the Rapid Classification of Events (ALeRCE) Alert Broker. , 161(5):242, May 2021. doi: 10.3847/1538-3881/abe9bc.

Fortunato, M., Blundell, C., and Vinyals, O. Bayesian Recurrent Neural Networks. *ArXiv e-prints*, art. arXiv:1704.02798, April 2017.

Gal, Y. and Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *ArXiv e-prints*, art. arXiv:1506.02142, June 2015.

Godines, D., Bachelet, E., Narayan, G., and Street, R. A. A machine learning classifier for microlensing in wide-field surveys. *Astronomy and Computing*, 28:100298, July 2019. doi: 10.1016/j.ascom.2019.100298.

Ishida, E. E. O., Beck, R., González-Gaitán, S., de Souza, R. S., Krone-Martins, A., Barrett, J. W., Kennamer, N., Vilalta, R., Burgess, J. M., Quint, B., Vitorelli, A. Z., Mahabal, A., and Gangler, E. Optimizing spectroscopic follow-up strategies for supernova photometric classification with active learning. *Monthly Notices of the Royal Astronomical Society*, 483:2–18, February 2019. doi: 10.1093/mnras/sty3015.

Leoni, M., Ishida, E. E. O., Peloton, J., and Möller, A. Fink: early supernovae Ia classification using active learning. *arXiv e-prints*, art. arXiv:2111.11438, November 2021.

Möller, A. and de Boissière, T. SuperNNova: an open-source framework for Bayesian, neural network-based supernova classification. *Monthly Notices of the Royal Astronomical Society*, 491(3):4277–4293, 12 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz3312. URL https://doi.org/10.1093/mnras/stz3312.

Möller, A., Peloton, J., Ishida, E. E. O., Arnault, C., Bachelet, E., Blaineau, T., Boutigny, D., Chauhan, A., Gangler, E., Hernandez, F., Hrivnac, J., Leoni, M., Leroy, N., Moniez, M., Pateyron, S., Ramparison, A., Turpin, D., Ansari, R., Allam, Tarek, J., Bajat, A., Biswas, B., Boucaud, A., Bregeon, J., Campagne, J.-E., Cohen-Tanugi, J., Coleiro, A., Dornic, D., Fouchez, D., Godet, O., Gris, P., Karpov, S., Nebot Gomez-Moran, A., Neveu, J., Plaszczynski, S., Savchenko, V., and Webb, N. FINK, a new generation of broker for the LSST community. , 501(3):3272–3288, March 2021. doi: 10.1093/mnras/staa3602.

Möller, A., Smith, M., Sako, M., Sullivan, M., Vincenzi, M., Wiseman, P., Armstrong, P., Asorey, J., Brout, D., Carollo, D., Davis, T. M., Frohmaier, C., Galbany, L., Glazebrook, K., Kelsey, L., Kessler, R., Lewis, G. F., Lidman, C., Malik, U., Nichol, R. C., Scolnic, D., Tucker, B. E., Abbott, T. M. C., Aguena, M., Allam, S., Annis, J., Bertin, E., Bocquet, S., Brooks, D., Burke, D. L., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., Castander, F. J., Conselice, C., Costanzi, M., Crocce, M., da Costa, L. N., De Vicente, J., Desai, S., Diehl, H. T., Doel, P., Everett, S., Ferrero, I., Finley, D. A., Flaugher, B., Friedel, D., Frieman, J., García-Bellido, J., Gerdes, D. W., Gruen, D., Gruendl, R. A., Gschwend, J., Gutierrez, G., Herner, K., Hinton, S. R., Hollowood, D. L., Honscheid, K., James, D. J., Kuehn, K., Kuropatkin, N., Lahav, O., March, M., Marshall, J. L., Menanteau, F., Miquel, R., Morgan, R., Palmese, A., Paz-Chinchón, F., Pieres, A., Malagón, A. A. P., Romer, A. K., Roodman, A., Sanchez, E., Scarpine, V., Schubnell, M., Serrano, S., Sevilla-Noarbe, I., Suchyta, E., Tarle, G., Thomas, D., To, C., and Varga, T. N. The Dark Energy Survey 5-year photometrically identified Type Ia Supernovae. , June 2022. doi: 10.1093/mnras/stac1691.

Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., and Hložek, R. RAPID: Early Classification of Explosive Transients using Deep Learning. *arXiv e-prints*, art. arXiv:1904.00014, 03 2019.

Narayan, G., Zaidi, T., Soraisam, M. D., Wang, Z., Lochner, M., Matheson, T., Saha, A., Yang, S., and ANTARES Collaboration. Machine-learning-based Brokers for Real-time Classification of the LSST Alert Stream. , 236(1):9, May 2018. doi: 10.3847/1538-4365/aab781.

Nordin, J., Brinnel, V., van Santen, J., Bulla, M., Feindt, U., Franckowiak, A., Fremling, C., Gal-Yam, A., Giomi, M., Kowalski, M., Mahabal, A., Miranda, N., Rauch, L., Reusch, S., Rigault, M., Schulze, S., Sollerman, J., Stein, R., Yaron, O., van Velzen, S., and Ward, C. Transient processing and analysis using AMPEL: alert management, photometry, and evaluation of light curves. , 631:A147, November 2019. doi: 10.1051/0004-6361/201935634.

Smith, K. W., Williams, R. D., Young, D. R., Ibsen, A., Smartt, S. J., Lawrence, A., Morris, D., Voutsinas, S., and Nicholl, M. Lasair: The transient alert broker for LSST:UK. *Research Notes of the AAS*, 3(1):26, jan 2019. doi: 10.3847/2515-5172/ab020f. URL https://doi.org/10.3847%2F2515-5172%2Fab020f.

Villar, V. A. and et al. SuperRAENN: A Semisupervised Supernova Photometric Classification Pipeline Trained on Pan-STARRS1 Medium-Deep Survey Supernovae. , 905 (2):94, December 2020. doi: 10.3847/1538-4357/abc6fd.

Villar, V. A., Berger, E., Miller, G., Chornock, R., Rest, A., Jones, D. O., Drout, M. R., Foley, R. J., Kirshner, R., Lunnan, R., Magnier, E., Milisavljevic, D., Sanders,

N., and Scolnic, D. Supernova Photometric Classification Pipelines Trained on Spectroscopically Classified Supernovae from the Pan-STARRS1 Medium-deep Survey. , 884(1):83, October 2019. doi: 10.3847/1538-4357/ab418c.

Walmsley, M., Smith, L., Lintott, C., Gal, Y., Bamford, S., Dickinson, H., Fortson, L., Kruk, S., Masters, K., Scarlata, C., Simmons, B., Smethurst, R., and Wright, D. Galaxy Zoo: probabilistic morphology through Bayesian CNNs and active learning. , 491(2):1554–1574, Jan 2020. doi: 10.1093/mnras/stz2816.