Transfer Learning for Transient Classification: From Simulations to Real Data and ZTF to LSST

Rithwik Gupta [1 2] **Daniel Muthukrishna** [1] **Nabeel Rehemtulla** [3 4 5] **Ved Shah** [3 4 5]

Abstract

The shortcomings of existing astronomical transient classification tools are extremely pressing as the the Legacy Survey of Space and Time (LSST) will demand performant classifiers to exploit its capabilities after commencing imminently. (i) Classifiers trained on simulations struggle with real data; (ii) models developed for one survey do not generalize to others; and (iii) new surveys require prohibitively large amounts of labeled training data. We demonstrate that transfer learning can overcome these challenges by repurposing models trained either on simulations or data from other surveys. We show that transfer learning from a model trained on simulated ZTF light curves reduces the amount of real labeled ZTF examples needed by 95% relative to a model trained from scratch and yielding similar performance. Similarly, when adapting ZTF models for LSST simulations, transfer learning achieves 94% of the baseline performance while requiring only 30% of the training data. These findings have significant implications for the early operations of LSST, suggesting that reliable automated classification will be possible soon after the survey begins, rather than waiting months or years to accumulate sufficient training data.

1. Introduction

With the development of advanced survey telescopes, we are entering a new era for astronomical study. The Vera C. Rubin Observatory's LSST (Ivezić et al., 2019) is expected to generate tens of millions of transient alerts per night, an order of magnitude more than current state-of-the-art surveys. This unprecedented volume of data has motivated the development of numerous machine learning-based alert and light-curve classification methods (e.g. Charnock & Moss, 2017; Muthukrishna et al., 2019; Möller & de Boissière, 2020; Boone, 2019; Gomez et al., 2020; Villar et al., 2019; 2020; Carrasco-Davis et al., 2021; Boone, 2021; Qu et al., 2021; Gagliano et al., 2021; Pimentel et al., 2023; Gomez et al., 2023; Sheng et al., 2023; de Soto et al., 2024; Rehemtulla et al., 2024; Shah et al., 2025).

Past studies have been limited by three key constraints. First, classification pipelines are often trained and validated on simulated data, but struggle to generalize to real observations due to subtle differences between simulated and observed light curves (e.g., Muthukrishna et al., 2019). Second, models developed for specific surveys typically require complete retraining when applied to data from different telescopes, despite the underlying physics of the transients remaining unchanged. Third, achieving robust classification performance requires large amounts of labeled training data— a resource-intensive process requiring considerable time and expensive spectroscopic follow-up observations.

To address these limitations, we explore transfer learning, a method that has been successfully applied across many fields in machine learning to adapt models trained for one task to similar applications (Zhuang et al., 2021). The technique works by reusing features and relationships that a model has already learned, even when the specific data distributions differ. This is particularly effective in two scenarios that directly address the challenges in astronomical classification: (i) when labeled data in the target domain is limited or (ii) when training new models from scratch is computationally expensive. By reusing learned features and fine-tuning only specific components of the model, transfer learning can dramatically reduce both the amount of labeled data and computational resources needed to achieve high performance on new tasks.

The potential of transfer learning is particularly relevant as we prepare for LSST as it provides a solution to the gap between simulated and real data. Deep learning models require large amounts of training data, which has led to the development of sophisticated light curve simulations (e.g.,

---

[1]Massachusetts Institute of Technology, Cambridge, MA 02139, USA [2]Irvington High School, Fremont, CA 94538, USA [3]Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA [4]Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Evanston, IL 60201, USA [5]NSF-Simons AI Institute for the Sky (SkAI), Chicago, IL 60611, USA. Correspondence to: Daniel Muthukrishna <danmuth@mit.edu>, Rithwik Gupta <rithwikca2020@gmail.com>.

Kessler et al., 2019; Narayan & ELAsTiCC Team, 2023) and numerous classifiers trained on these simulations (e.g., Boone, 2019; Hložek et al., 2023). However, simulations do not perfectly capture real observations, making training on labeled survey data necessary to actually use models in production (e.g., Rehemtulla et al., 2024). While previous work has treated simulated and real data independently, with models often struggling to generalize between the two, transfer learning provides a framework to leverage features learned from simulations while adapting to real observations. Similarly, while most classifiers are built for specific surveys, transfer learning enables models trained on existing facilities to be adapted for new ones, significantly reducing the data requirements for achieving high performance. However, the initial model architecture determines the feasibility of transfer learning across different surveys. In this work, we incorporate the observation wavelength directly in the model input to facilitate transfer learning across different surveys with varying passbands.

We demonstrate two key benefits of transfer learning for astronomical transient classification: (i) transfer learning reduces the amount of data needed to generalize from simulated data to real observations, and (ii) transfer learning reduces the amount of labeled data needed to generalize from one survey to another.

2. Data

In this work, we use three distinct datasets to evaluate transfer learning for transient classification: simulated light curves that match the observing properties of the Zwicky Transient Facility (ZTF; Bellm et al., 2019), real observations from the ZTF Bright Transient Survey (BTS; Fremling et al., 2020; Perley et al., 2020; Rehemtulla et al., 2024), and simulated light curves matching the expected LSST observing properties (Narayan & ELAsTiCC Team, 2023). Each dataset provides different opportunities to test transfer learning between simulated and real data, as well as between different surveys. The class distribution for each dataset is summarized in Table 1. Further details about each dataset used can be found in Appendix B.

3. Methods

3.1. Classifiers

We build a Recurrent Neural Network (RNN) classifier with Gated Recurrent Units (GRU; Cho et al., 2014) following the architecture used in Muthukrishna et al. (2019) and updated in Gupta et al. (2024; 2025). As illustrated in Figure 2, we include a parallel fully-connected (dense) neural network to process any available metadata, which in this work consists only of Milky Way extinction. While we do not use other host information in our work, our classification framework is designed to incorporate both time-dependent and time-independent features. We present results using this architecture, but we expect transfer learning results to be broadly consistent across other classification architectures that use flexible input formats capable of handling data from different surveys.

To enable transfer learning in this work, we use a survey-agnostic input representation. The input vector, $\boldsymbol{X}_s$, for a transient s, is a $4 \times N_t$ matrix, where N_t is the maximum number of timesteps for any light curve in our datasets. $N_t = 656$ in this work, however most light curves have far fewer observations. Each row i of the input matrix $\boldsymbol{X}_s$ contains the following information:

$$\boldsymbol{X}_{si} = [\lambda_p, t_{si}, f_{si}, \epsilon_{si}], \tag{1}$$

where t_{si} is the time of the ith observation scaled between 0 and 1, λ_p is the median wavelength of the passband, f_{si} is the flux scaled by dividing by 500, and ϵ_{si} is the corresponding flux error with the same scaling. We select a constant scaling factor of 500 as it is close to the mean flux in our dataset and enables real-time classification by avoiding peak-based scaling, which would depend on observations that may not yet be available in real-time scenarios. Our model does not require redshift information, making it useful in the early observations of LSST when redshifts may not be known for many sources.

Using this input method has two key advantages. First, it circumvents the need for interpolation between observations, which can be problematic for sparsely sampled light curves. Second, it passes all passband information through the same channel, which allows us to easily repurpose a trained model for a different survey with a different number of passbands.

To counteract class imbalances in our dataset, we use a weighted categorical cross-entropy loss to train our model. The weight of a training object w_c is inversely proportional to the fraction of transients from the class c in the training set. We train our models using the Adam optimizer (Kingma & Ba, 2014) over 100 epochs using EarlyStopping implemented in keras (Chollet, 2015).

3.2. Transfer Learning

To assess the effectiveness of transfer learning, we define a source dataset and a target dataset. We first train a classifier on the source dataset, and then adapt it for the target dataset with the specific focus of analysing how much labelled data we need from the target dataset to achieve good performance. For all analyses in this work, ZTF Sims serves as the source dataset.

We are particularly interested in transfer learning from simulated to real data (ZTF to BTS) and, in preparation for the Vera Rubin Observatory, from ZTF to LSST data. Thus, we

Class	SNIa	SNIa-91bg	SNIax	SNIb	SNIc	SNIc-BL	SNII	SNIIn	SNIIb	TDE	SLSN-I	AGN
ZTF Sims	11587	13000	13000	5267	1583	1423	13000	13000	12323	11354	12880	10561
LSST Sims	51745	12272	12012	37261	24513	10331	71806	34016	23408	28285	28322	32681
BTS Data	4628 (SNIa)			339 (SNIb/c)			1212 (SNII)			22	46	–

Table 1. Class distribution for the datasets used in this work, namely the ZTF Sims (Muthukrishna et al., 2021), LSST Sims (ELaSTiCC; Narayan & ELAsTiCC Team, 2023), and BTS Data (Rehemtulla et al., 2024).

run two experiments, using either BTS Data or LSST Sims as the target dataset. In both cases, we intentionally limit the amount of labelled data available in the target dataset.

Figure 2 summarises our model architecture and experimental setup for transfer learning. We initially train a model to classify a transient into one of the n classes in the ZTF Sims. After pretraining on the ZTF Sims, we replace the classification layer with a Multi-Layer Perceptron (MLP) followed by a new classification layer. The new classification layer varies in size depending on the number of classes, m, in the target dataset. The complexity of the MLP is determined by how different the target dataset is from the source dataset. We describe the details of each of these experiments in the following subsections.

We note that the primary benefit of transfer learning comes from pretraining rather than from specific choices about which layers to freeze, as unfreezing the entire network yields similar performance. The choice of freezing strategy primarily acts to reduce computational costs during fine-tuning.

3.2.1. FINE-TUNING FOR BTS

When fine-tuning our model on the BTS data, we replace the classification layer with a new one for the $m = 5$ BTS classes and freeze all other layers of the pretrained model. Since BTS and ZTF Sims share the same observing properties, we only retrain the new output layer and freeze the rest of the model weights. This prevents the network from modifying the features it has already learned about transients from its pretraining on the ZTF Sims.

3.2.2. FINE-TUNING FOR LSST

Transferring from ZTF to LSST presents an additional challenge: adapting from two passbands (gr) to six ($ugrizy$). While our input format (see eq. 1) enables the model to learn a coarse relationship between flux and wavelength, the model requires retraining to capture transient behaviour across the new passbands.

For LSST, we unfreeze both the initial GRU layers (blue in Figure 2) and the final classification layers (yellow in Figure 2). We found that only retraining the final layers, as we did for BTS, resulted in poor performance. The initial GRU layers, which process the time-series information, need to

be retrained to effectively handle the more complex and sparser LSST light curves.

While unfreezing the entire network performs similarly to our selective freezing approach, we choose to freeze the dense layers to reduce computational costs while maintaining equivalent performance.

4. Results

To evaluate the effectiveness of transfer learning, we compare classifiers trained with transfer learning against those trained from scratch on the target dataset. We focus on two key experiments: transferring from simulated to real data (ZTF Sims to BTS) and transferring between surveys (ZTF to LSST). In both cases, we analyse how much labelled data from the target dataset is needed to achieve good classification performance. As we vary the amount of labelled data available while training, we use 80% of the limited data to train the model, and 20% to validate its performance and determine optimal training time. The remaining data (outside of the data provided during training) is used in the test set to evaluate the model.

4.1. Simulated to Real Data

To evaluate transfer learning from simulations to real data, we compare two approaches: fine-tuning a model pretrained on ZTF Sims and training a model directly on BTS data. For both approaches, we train with varying amounts of BTS data, sampling from the full dataset to maintain realistic class distributions.

We evaluate the performance using the Area Under the macro-averaged Receiver Operating Characteristics Curve (AUROC). The Receiver Operating Characteristic (ROC) curve plots the False Positive Rate against the True Positive Rate over a range of classification probability thresholds.

Figure 1 [left] shows that transfer learning from simulated data significantly reduces the amount of real labelled data needed to build an effective classifier. A model pretrained on ZTF Sims achieves the baseline performance with just 5% of the data. This reduction in required training data is particularly valuable given the time and expense of obtaining spectroscopic labels. We restrict our evaluation at 4,000 labelled training samples as going further would leave too
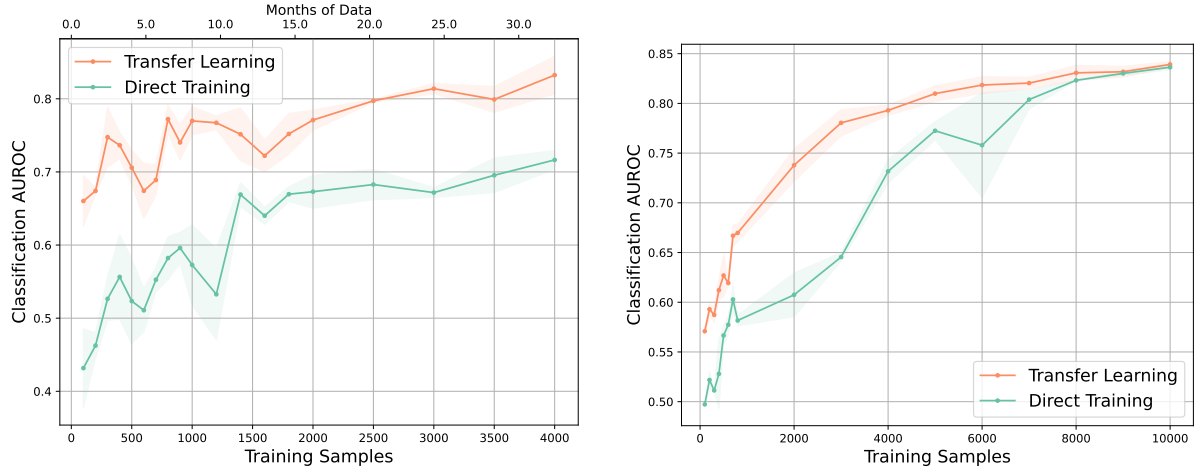
Figure 1. Transfer learning performance from ZTF sims to BTS [left] and ZTF Sims to LSST Sims [right]. The x-axis shows the number of randomly sampled training light curves provided to the model and the y-axis shows the classification performance as the macro-averaged Area under the Receiver Operating Characteristic (ROC) curve for each class (AUROC). The error bars show the standard deviation across 10 independent training runs.

few TDEs in the test set for robust evaluation. Even at this point, transfer learning provides a significant performance improvement when compared to direct training.

The performance improvement in low-data scenarios is primarily driven by minority classes, particularly TDEs and SLSNe-I, which transfer learning can effectively classify despite having very few labelled examples. These rare classes also account for the large variance observed at small training set sizes in Figure 1, where classifiers become highly sensitive to each additional training sample. The class-specific results in Figure 3 confirm that most performance variability stems from the TDE and SLSN-I classes. Importantly, transfer learning still provides substantial benefits when applied to only the common transient classes, as demonstrated in Appendix D.

4.2. ZTF to LSST

To evaluate transfer learning between surveys, we compare models pretrained on ZTF Sims with those trained directly on LSST simulations.

As depicted in Figure 1 [right], transfer learning significantly reduces the data requirements for developing LSST classifiers. A model pretrained on ZTF achieves 94% of the baseline performance while requiring only 30% of the training data. Despite the additional complexity of adapting to LSST's six passbands (compared to ZTF's two), pretraining helps the model learn features and capture some survey-agnostic patterns that are intrinsic to the astrophysical phenomena, thus reducing the amount of new training data required. This suggests that existing classification models from current surveys such as ZTF can be efficiently

adapted for LSST, enabling the rapid deployment of reliable classifiers early in the survey.

5. Conclusion

Observations of the transient universe are entering a new era with the advent of wide-field surveys such as LSST. The unprecedented data volumes from LSST will require robust automated transient classification systems that can be deployed rapidly after the survey begins. Traditionally, developing reliable classifiers for a new survey requires accumulating substantial amounts of labeled data over months or years of operations. In this work, we demonstrate that transfer learning provides an effective solution to this challenge.

Our results show a significant advantage of transfer learning for astronomical transient classification: fine-tuning existing models requires 70% to 95% less labeled data than training from scratch. A model pretrained on ZTF simulations achieves equivalent performance on real data while requiring only 25% of the labeled examples. Similarly, when adapting ZTF models for LSST, transfer learning maintains 94% of the baseline performance with just 30% of the training data. These results suggest that classifiers trained on existing surveys or simulations can be efficiently adapted to new data sources.

The success of transfer learning between both simulated and real data, as well as between different surveys, demonstrates that these models learn generalizable features of astronomical transients. This has important implications for future surveys such as LSST, suggesting that reliable automated classification will be possible soon after observations begin.

4

Several promising directions remain for future work. First, combining data from multiple existing surveys could provide more robust base models that capture a wider range of transient behavior. Further, applying these techniques to a broader range of astronomical phenomena could extend the benefits of transfer learning beyond transient classification.

As we prepare for the era of LSST, our results demonstrate that transfer learning will be crucial for developing the next generation of astronomical classification systems, enabling rapid scientific discovery through efficient adaptation of existing models. The code used in this work is publicly available[1].

Acknowledgements

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., Earl, N., Starkman, N., Bradley, L., Shupe, D. L., Patil, A. A., Corrales, L., Brasseur, C. E., N''othe, M., Do-

nath, A., Tollerud, E., Morris, B. M., Ginsburg, A., Vaher, E., Weaver, B. A., Tocknell, J., Jamieson, W., van Kerkwijk, M. H., Robitaille, T. P., Merry, B., Bachetti, M., G''unther, H. M., Aldcroft, T. L., Alvarado-Montes, J. A., Archibald, A. M., B'odi, A., Bapat, S., Barentsen, G., Baz'an, J., Biswas, M., Boquien, M., Burke, D. J., Cara, D., Cara, M., Conroy, K. E., Conseil, S., Craig, M. W., Cross, R. M., Cruz, K. L., D'Eugenio, F., Dencheva, N., Devillepoix, H. A. R., Dietrich, J. P., Eigenbrot, A. D., Erben, T., Ferreira, L., Foreman-Mackey, D., Fox, R., Freij, N., Garg, S., Geda, R., Glattly, L., Gondhalekar, Y., Gordon, K. D., Grant, D., Greenfield, P., Groener, A. M., Guest, S., Gurovich, S., Handberg, R., Hart, A., Hatfield-Dodds, Z., Homeier, D., Hosseinzadeh, G., Jenness, T., Jones, C. K., Joseph, P., Kalmbach, J. B., Karamehmetoglu, E., Kaluszy'nski, M., Kelley, M. S. P., Kern, N., Kerzendorf, W. E., Koch, E. W., Kulumani, S., Lee, A., Ly, C., Ma, Z., MacBride, C., Maljaars, J. M., Muna, D., Murphy, N. A., Norman, H., O'Steen, R., Oman, K. A., Pacifici, C., Pascual, S., Pascual-Granado, J., Patil, R. R., Perren, G. I., Pickering, T. E., Rastogi, T., Roulston, B. R., Ryan, D. F., Rykoff, E. S., Sabater, J., Sakurikar, P., Salgado, J., Sanghi, A., Saunders, N., Savchenko, V., Schwardt, L., Seifert-Eckert, M., Shih, A. Y., Jain, A. S., Shukla, G., Sick, J., Simpson, C., Singanamalla, S., Singer, L. P., Singhal, J., Sinha, M., SipHocz, B. M., Spitler, L. R., Stansby, D., Streicher, O., Sumak, J., Swinbank, J. D., Taranu, D. S., Tewary, N., Tremblay, G. R., Val-Borro, M. d., Van Kooten, S. J., Vasovi'c, Z., Verma, S., de Miranda Cardoso, J. V., Williams, P. K. G., Wilson, T. J., Winkel, B., Wood-Vasey, W. M., Xue, R., Yoachim, P., Zhang, C., Zonca, A., and Astropy Project Contributors. The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package. *ApJ*, 935(2): 167, August 2022. doi: 10.3847/1538-4357/ac7c74.

Bellm, E. C., Kulkarni, S. R., Graham, M. J., Dekany, R., Smith, R. M., Riddle, R., Masci, F. J., Helou, G., Prince, T. A., Adams, S. M., Barbarino, C., Barlow, T., Bauer, J., Beck, R., Belicki, J., Biswas, R., Blagorodnova, N., Bodewits, D., Bolin, B., Brinnel, V., Brooke, T., Bue, B., Bulla, M., Burruss, R., Cenko, S. B., Chang, C.-K., Connolly, A., Coughlin, M., Cromer, J., Cunningham, V., De, K., Delacroix, A., Desai, V., Duev, D. A., Eadie, G., Farnham, T. L., Feeney, M., Feindt, U., Flynn, D., Franckowiak, A., Frederick, S., Fremling, C., Gal-Yam, A., Gezari, S., Giomi, M., Goldstein, D. A., Golkhou, V. Z., Goobar, A., Groom, S., Hacopians, E., Hale, D., Henning, J., Ho, A. Y. Q., Hover, D., Howell, J., Hung, T., Huppenkothen, D., Imel, D., Ip, W.-H., Ivezić, Ž., Jackson, E., Jones, L., Juric, M., Kasliwal, M. M., Kaspi, S., Kaye, S., Kelley, M. S. P., Kowalski, M., Kramer, E., Kupfer, T., Landry, W., Laher, R. R., Lee, C.-D.,

---

[1] https://github.com/rithwik-g/transient-transfer-learning

Lin, H. W., Lin, Z.-Y., Lunnan, R., Giomi, M., Mahabal, A., Mao, P., Miller, A. A., Monkewitz, S., Murphy, P., Ngeow, C.-C., Nordin, J., Nugent, P., Ofek, E., Patterson, M. T., Penprase, B., Porter, M., Rauch, L., Rebbapragada, U., Reiley, D., Rigault, M., Rodriguez, H., van Roestel, J., Rusholme, B., van Santen, J., Schulze, S., Shupe, D. L., Singer, L. P., Soumagnac, M. T., Stein, R., Surace, J., Sollerman, J., Szkody, P., Taddia, F., Terek, S., Van Sistine, A., van Velzen, S., Vestrand, W. T., Walters, R., Ward, C., Ye, Q.-Z., Yu, P.-C., Yan, L., and Zolkower, J. The Zwicky Transient Facility: System Overview, Performance, and First Results. *PASP*, 131(995):018002, January 2019. doi: 10.1088/1538-3873/aaecbe.

Boerner, T. J., Deems, S., Furlani, T. R., Knuth, S. L., and Towns, J. Access: Advancing innovation: Nsf's advanced cyberinfrastructure coordination ecosystem: Services & support. In *Practice and Experience in Advanced Research Computing 2023: Computing for the Common Good*, PEARC '23, pp. 173–176, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450399852. doi: 10.1145/3569951.3597559. URL https://doi.org/10.1145/3569951.3597559.

Boone, K. Avocado: Photometric Classification of Astronomical Transients with Gaussian Process Augmentation. *AJ*, 158(6):257, December 2019. doi: 10.3847/1538-3881/ab5182.

Boone, K. Parsnip: Generative models of transient light curves with physics-enabled deep learning. *AJ*, 162(6):275, December 2021. ISSN 1538-3881. doi: 10.3847/1538-3881/ac2a2d. URL http://dx.doi.org/10.3847/1538-3881/ac2a2d.

Cabrera-Vives, G., Moreno-Cartagena, D., Astorga, N., Reyes-Jainaga, I., Förster, F., Huijse, P., Arredondo, J., Muñoz Arancibia, A. M., Bayo, A., Catelan, M., Estévez, P. A., Sánchez-Sáez, P., Álvarez, A., Castellanos, P., Gallardo, P., Moya, A., and Rodriguez-Mancini, D. ATAT: Astronomical Transformer for time series and Tabular data. *A&A*, 689:A289, September 2024. doi: 10.1051/0004-6361/202449475.

Carrasco-Davis, R., Reyes, E., Valenzuela, C., Förster, F., Estévez, P. A., Pignata, G., Bauer, F. E., Reyes, I., Sánchez-Sáez, P., Cabrera-Vives, G., Eyheramendy, S., Catelan, M., Arredondo, J., Castillo-Navarrete, E., Rodríguez-Mancini, D., Ruz-Mieres, D., Moya, A., Sabatini-Gacitúa, L., Sepúlveda-Cobo, C., Mahabal, A. A., Silva-Farfán, J., Camacho-Iñiguez, E., and Galbany, L. Alert classification for the alerce broker system: The real-time stamp classifier. *AJ*, 162(6):231, November 2021. ISSN 1538-3881. doi: 10.3847/

1538-3881/ac0ef1. URL http://dx.doi.org/10.3847/1538-3881/ac0ef1.

Charnock, T. and Moss, A. Deep Recurrent Neural Networks for Supernovae Classification. *ApJl*, 837:L28, March 2017. doi: 10.3847/2041-8213/aa603d.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734. Association for Computational Linguistics, 2014. doi: 10.3115/v1/D14-1179.

Chollet, F. keras. https://github.com/fchollet/keras, 2015.

de Soto, K. M., Villar, A., Berger, E., Gomez, S., Hosseinzadeh, G., Branton, D., Campos, S., DeLucchi, M., Kubica, J., Lynn, O., Malanchev, K., and Malz, A. I. Superphot+: Realtime fitting and classification of supernova light curves, 2024. URL https://arxiv.org/abs/2403.07975.

Fremling, C., Miller, A. A., Sharma, Y., Dugas, A., Perley, D. A., Taggart, K., Sollerman, J., Goobar, A., Graham, M. L., Neill, J. D., Nordin, J., Rigault, M., Walters, R., Andreoni, I., Bagdasaryan, A., Belicki, J., Cannella, C., Bellm, E. C., Cenko, S. B., De, K., Dekany, R., Frederick, S., Golkhou, V. Z., Graham, M. J., Helou, G., Ho, A. Y. Q., Kasliwal, M. M., Kupfer, T., Laher, R. R., Mahabal, A., Masci, F. J., Riddle, R., Rusholme, B., Schulze, S., Shupe, D. L., Smith, R. M., van Velzen, S., Yan, L., Yao, Y., Zhuang, Z., and Kulkarni, S. R. The Zwicky Transient Facility Bright Transient Survey. I. Spectroscopic Classification and the Redshift Completeness of Local Galaxy Catalogs. *ApJ*, 895(1):32, May 2020. doi: 10.3847/1538-4357/ab8943.

Gagliano, A., Narayan, G., Engel, A., Carrasco Kind, M., and LSST Dark Energy Science Collaboration. GHOST: Using Only Host Galaxy Information to Accurately Associate and Distinguish Supernovae. *ApJ*, 908(2):170, February 2021. doi: 10.3847/1538-4357/abd02b.

Gomez, S., Berger, E., Blanchard, P. K., Hosseinzadeh, G., Nicholl, M., Villar, V. A., and Yin, Y. Fleet: A redshift-agnostic machine learning pipeline to rapidly identify hydrogen-poor superluminous supernovae. *ApJ*, 904(1):74, nov 2020. doi: 10.3847/1538-4357/abbf49. URL https://dx.doi.org/10.3847/1538-4357/abbf49.

Gomez, S., Berger, E., Blanchard, P. K., Hosseinzadeh, G., Nicholl, M., Hiramatsu, D., Villar, V. A., and Yin,

Y. The first two years of fleet: An active search for superluminous supernovae. *ApJ*, 949(2):114, jun 2023. doi: 10.3847/1538-4357/acc536. URL https://dx.doi.org/10.3847/1538-4357/acc536.

Gupta, R., Muthukrishna, D., and Lochner, M. A Classifier-Based Approach to Multi-Class Anomaly Detection Applied to Astronomical Time-Series. *arXiv e-prints*, art. arXiv:2408.08888, August 2024. doi: 10.48550/arXiv.2408.08888.

Gupta, R., Muthukrishna, D., and Lochner, M. A classifier-based approach to multiclass anomaly detection for astronomical transients. *RAS Techniques and Instruments*, 4: rzae054, January 2025. doi: 10.1093/rasti/rzae054.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

Hložek, R., Malz, A. I., Ponder, K. A., Dai, M., Narayan, G., Ishida, E. E. O., Allam, Jr., T., Bahmanyar, A., Bi, X., Biswas, R., Boone, K., Chen, S., Du, N., Erdem, A., Galbany, L., Garreta, A., Jha, S. W., Jones, D. O., Kessler, R., Lin, M., Liu, J., Lochner, M., Mahabal, A. A., Mandel, K. S., Margolis, P., Martínez-Galarza, J. R., McEwen, J. D., Muthukrishna, D., Nakatsuka, Y., Noumi, T., Oya, T., Peiris, H. V., Peters, C. M., Puget, J. F., Setzer, C. N., Siddhartha, Stefanov, S., Xie, T., Yan, L., Yeh, K. H., and Zuo, W. Results of the Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC). *ApJs*, 267(2):25, August 2023. doi: 10.3847/1538-4365/accd6a.

Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science and Engineering*, 9(03):90–95, may 2007. ISSN 1558-366X. doi: 10.1109/MCSE.2007.55.

Ivezić, Ž., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S. F., Andrew, J., and et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *ApJ*, 873:111, March 2019. doi: 10.3847/1538-4357/ab042c.

Kessler, R., Narayan, G., Avelino, A., Bachelet, E., Biswas, R., Brown, P. J., Chernoff, D. F., Connolly, A. J., Dai, M., Daniel, S., Stefano, R. D., Drout, M. R., Galbany, L., Gonzá lez-Gaitán, S., Graham, M. L., Hložek, R., Ishida, E. E. O., Guillochon, J., Jha, S. W., Jones, D. O., Mandel, K. S., Muthukrishna, D., O'Grady, A., Peters,

C. M., Pierel, J. R., Ponder, K. A., Prša, A., Rodney, S., and and, V. A. V. Models and simulations for the photometric LSST astronomical time series classification challenge (PLAsTiCC). *PASP*, 131(1003):094501, jul 2019. doi: 10.1088/1538-3873/ab26f1. URL https://doi.org/10.1088%2F1538-3873%2Fab26f1.

Kingma, D. P. and Ba, J. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, art. arXiv:1412.6980, December 2014. doi: 10.48550/arXiv.1412.6980.

Masci, F. J., Laher, R. R., Rusholme, B., Shupe, D. L., Groom, S., Surace, J., Jackson, E., Monkewitz, S., Beck, R., Flynn, D., Terek, S., Landry, W., Hacopians, E., Desai, V., Howell, J., Brooke, T., Imel, D., Wachter, S., Ye, Q.-Z., Lin, H.-W., Cenko, S. B., Cunningham, V., Rebbapragada, U., Bue, B., Miller, A. A., Mahabal, A., Bellm, E. C., Patterson, M. T., Jurić, M., Golkhou, V. Z., Ofek, E. O., Walters, R., Graham, M., Kasliwal, M. M., Dekany, R. G., Kupfer, T., Burdge, K., Cannella, C. B., Barlow, T., Van Sistine, A., Giomi, M., Fremling, C., Blagorodnova, N., Levitan, D., Riddle, R., Smith, R. M., Helou, G., Prince, T. A., and Kulkarni, S. R. The Zwicky Transient Facility: Data Processing, Products, and Archive. *PASP*, 131(995): 018003, January 2019. doi: 10.1088/1538-3873/aae8ac.

McKinney, W. Data structures for statistical computing in python. In van der Walt, S. and Millman, J. (eds.), *Proceedings of the 9th Python in Science Conference*, pp. 51 – 56, 2010.

Möller, A. and de Boissière, T. SuperNNova: an open-source framework for Bayesian, neural network-based supernova classification. *MNRAS*, 491(3):4277–4293, January 2020. doi: 10.1093/mnras/stz3312.

Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., and Hložek, R. RAPID: Early classification of explosive transients using deep learning. *PASP*, 131 (1005):118002, sep 2019. doi: 10.1088/1538-3873/ab1609. URL https://doi.org/10.1088%2F1538-3873%2Fab1609.

Muthukrishna, D., Mandel, K. S., Lochner, M., Webb, S., and Narayan, G. Real-time Detection of Anomalies in Multivariate Time Series of Astronomical Data. *NeurIPS 2021*, art. arXiv:2112.08415, December 2021.

Narayan, G. and ELAsTiCC Team. The Extended LSST Astronomical Time-series Classification Challenge (ELAsTiCC). In *American Astronomical Society Meeting Abstracts*, volume 241 of *American Astronomical Society Meeting Abstracts*, pp. 117.01, January 2023.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P.,

Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Perley, D. A., Fremling, C., Sollerman, J., Miller, A. A., Dahiwale, A. S., Sharma, Y., Bellm, E. C., Biswas, R., Brink, T. G., Bruch, R. J., De, K., Dekany, R., Drake, A. J., Duev, D. A., Filippenko, A. V., Gal-Yam, A., Goobar, A., Graham, M. J., Graham, M. L., Ho, A. Y. Q., Irani, I., Kasliwal, M. M., Kim, Y.-L., Kulkarni, S. R., Mahabal, A., Masci, F. J., Modak, S., Neill, J. D., Nordin, J., Riddle, R. L., Soumagnac, M. T., Strotjohann, N. L., Schulze, S., Taggart, K., Tzanidakis, A., Walters, R. S., and Yan, L. The Zwicky Transient Facility Bright Transient Survey. II. A Public Statistical Sample for Exploring Supernova Demographics. *ApJ*, 904(1):35, November 2020. doi: 10.3847/1538-4357/abbd98.

Pimentel, Ó., Estévez, P. A., and Förster, F. Deep Attention-based Supernovae Classification of Multiband Light Curves. *AJ*, 165(1):18, January 2023. doi: 10.3847/1538-3881/ac9ab4.

Qu, H., Sako, M., Möller, A., and Doux, C. Scone: Supernova classification with a convolutional neural network. *AJ*, 162(2):67, July 2021. ISSN 1538-3881. doi: 10.3847/1538-3881/ac0824. URL http://dx.doi.org/10.3847/1538-3881/ac0824.

Rehemtulla, N., Miller, A. A., Laz, T. J. D., Coughlin, M. W., Fremling, C., Perley, D. A., Qin, Y.-J., Sollerman, J., Mahabal, A. A., Laher, R. R., Riddle, R., Rusholme, B., and Kulkarni, S. R. The zwicky transient facility bright transient survey. iii. btsbot: Automated identification and follow-up of bright transients with deep learning. *ApJ*, 972(1):7, aug 2024. doi: 10.3847/1538-4357/ad5666. URL https://dx.doi.org/10.3847/1538-4357/ad5666.

Shah, V. G., Gagliano, A., Malanchev, K., Narayan, G., and Collaboration, T. L. D. E. S. Oracle: A real-time, hierarchical, deep-learning photometric classifier for the lsst, 2025. URL https://arxiv.org/abs/2501.01496.

Sheng, X., Nicholl, M., Smith, K. W., Young, D. R., Williams, R. D., Stevance, H. F., Smartt, S. J., Srivastav, S., and Moore, T. Neural engine for discovering luminous events (needle): identifying rare transient candidates in real time from host galaxy images, 2023. URL https://arxiv.org/abs/2312.04968.

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. *ApJ*, 884(1):83, October 2019. doi: 10.3847/1538-4357/ab418c.

Villar, V. A., Hosseinzadeh, G., Berger, E., Ntampaka, M., Jones, D. O., Challis, P., Chornock, R., Drout, M. R., Foley, R. J., Kirshner, R. P., Lunnan, R., Margutti, R., Milisavljevic, D., Sanders, N., Pan, Y.-C., Rest, A., Scolnic, D. M., Magnier, E., Metcalfe, N., Wainscoat, R., and Waters, C. SuperRAENN: A Semisupervised Supernova Photometric Classification Pipeline Trained on Pan-STARRS1 Medium-Deep Survey Supernovae. *ApJ*, 905 (2):94, December 2020. doi: 10.3847/1538-4357/abc6fd.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2021. doi: 10.1109/JPROC.2020.3004555.

A. Visual Methodology

Figure 2 summarizes our methodology.

B. Dataset Details

B.1. Simulated ZTF Data (ZTF Sims)

The simulated ZTF dataset is described in §2 of both Gupta et al. (2025) and Muthukrishna et al. (2021), and is based on the models developed for the Photometric LSST Astronomical Time-Series Classification (PLAsTiCC; Kessler et al., 2019). Each transient has flux and flux error measurements in the g and r passbands with a median cadence of 3 days in each band[2]. The dataset includes light curves from 12 transient classes: SNIa, SNIa-91bg, SNIax, SNIb, SNIc, SNIc-BL, SNII, SNIIn, SNIIb, TDE, SLSN-I, and AGN.

B.2. ELAsTiCC (LSST Sims)

The LSST simulations are drawn from the Extended LSST Astronomical Time Series Classification Challenge (ELAsTiCC; Narayan & ELAsTiCC Team, 2023), which builds upon the original PLAsTiCC simulations with improved models and more realistic survey properties. Light curves are simulated in all six LSST passbands ($ugrizy$) with the expected LSST cadence and depth. ELAsTiCC also estimates the frequency of each class to create a realistic distribution of observation rates with the notable upsampling of minority classes to make model training more efficient[3]. The classes we use from ELaSTiCC directly overlap with the classes from ZTF.

B.3. Bright Transient Survey Data (BTS Data)

For real observations, we use data from the ZTF BTS. The aim of BTS is to systematically classify every bright ($m < 18.5$ mag) extragalactic transient discovered in the public ZTF alert stream (Masci et al., 2019), providing a spectroscopically complete sample. The dataset adopted is queried and processed following the procedures in (Rehemtulla et al., 2024) and considers data up to August 2024. Light curves are available in the ZTF g and r filters. We use data from three broad transient classes, SNIa, SNII, and SNIb/c, and two rarer transient classes, TDE and SLSN-I. We use these broader classifications due to the small number of transients classified in most subtypes.
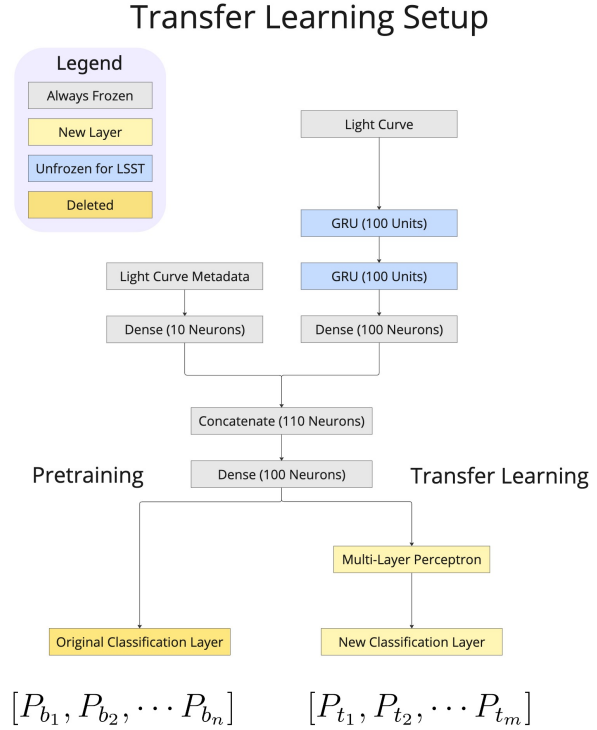

Figure 2. Schematic illustrating our neural network architecture and transfer learning methodology. The model is first pretrained on ZTF simulations to output probabilities $[P_{b_1}, P_{b_2}, ..., P_{b_n}]$ for each of the n classes in the source dataset. For transfer learning, we replace the output classification layer with a new one that outputs probabilities $[P_{t_1}, P_{t_2}, ..., P_{t_m}]$ for the m classes in the target dataset. When transferring to BTS data, we freeze the pretrained model and only train the new classification layer. When transferring to LSST, we also unfreeze the initial GRU layers to learn the additional passbands while keeping the dense layers frozen. The complexity of the additional Multi-Layer Perceptron (MLP) layers depends on the similarity between the source and target datasets - zero neurons for BTS (which is similar to ZTF sims) and two layers for LSST (which has different passbands).
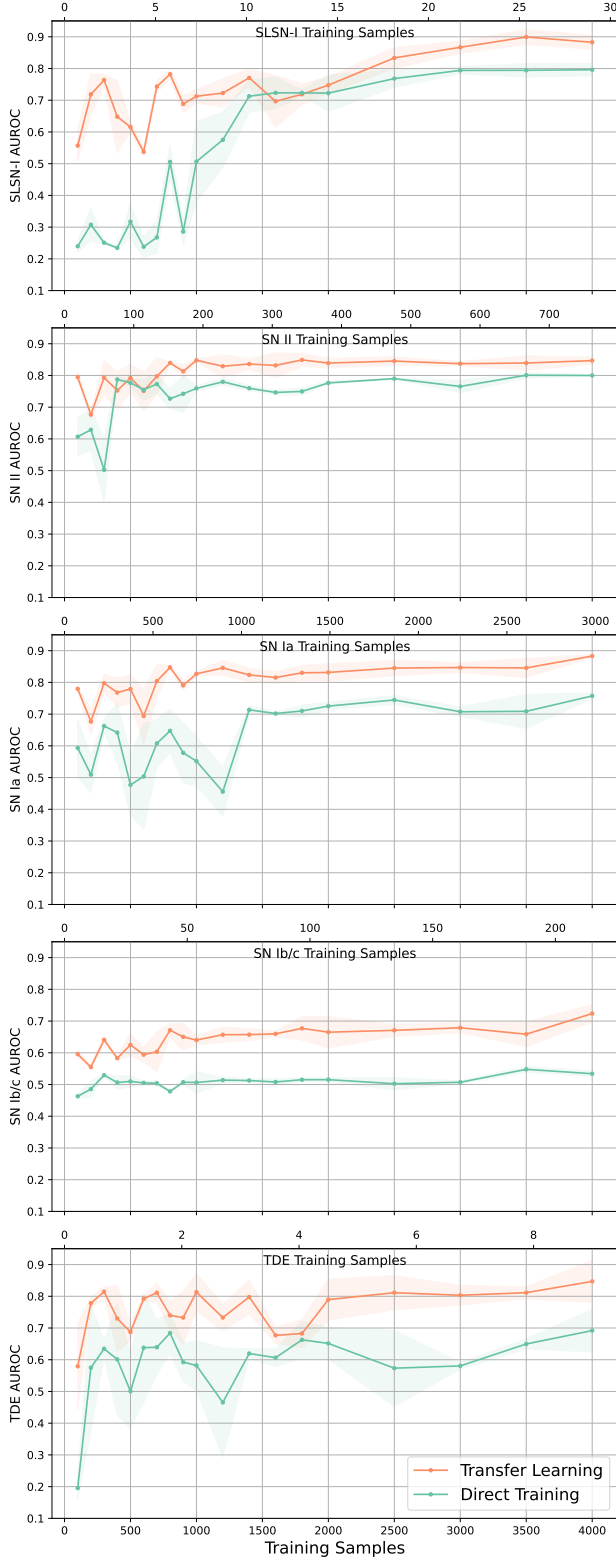
Figure 3. Class-specific classification AUROCs for transfer learning from simulated ZTF data to real BTS data.
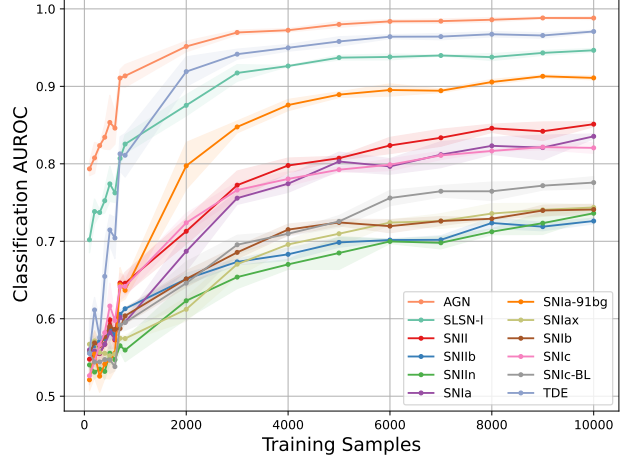

Figure 4. Class-specific classification AUROCs for transfer learning from ZTF sims to LSST sims.

C. Class-based Performance

C.1. BTS

Figure 3 shows the class-specific performance of transfer learning compared to direct training on BTS data. Transfer learning provides significant improvement across all transient classes, with the most pronounced benefits observed for rare classes such as SLSNe-I and TDEs, which have very few training samples.

C.2. LSST

Figure 4 shows the class-specific performance for transfer learning from ZTF simulations to LSST simulations. The relative performance across different transient classes reflects trends observed in other photometric classifiers such as RAPID (Muthukrishna et al., 2019), ATAT (Cabrera-Vives et al., 2024), and ORACLE (Shah et al., 2025). Classes with distinctive photometric signatures, such as AGNs with their characteristic stochastic variability over long timescales, achieve higher classification accuracy than supernova subtypes with more subtle differences. For example, distinguishing between SNIb and SNIc based solely on photometric rise times remains challenging, as these subtle features do not always manifest clearly in broad-band observations.
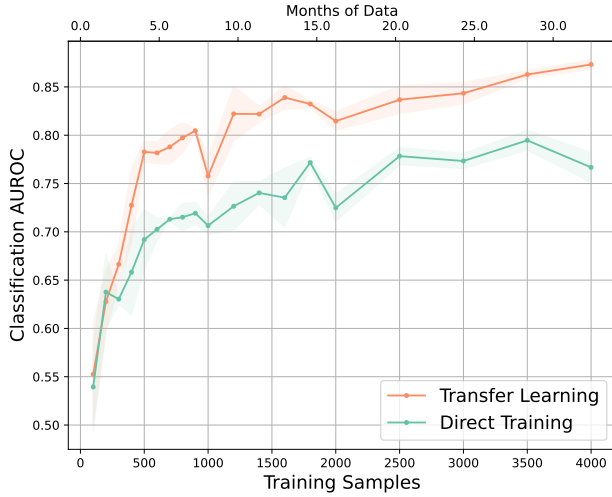
Figure 5. A recreation of Figure 1 [left] with TDEs and SLSN-Is omitted.

D. Performance without Minority Classes

The transfer learning performance improvement shown in Figure 1 [left] is partly attributed to minority classes (SLSN-I and TDE), which have very few labelled observations but are relatively easy to classify (Muthukrishna et al., 2019; Gupta et al., 2025). To test whether transfer learning benefits extend beyond these rare classes, Figure 5 shows the same analysis using only the three most common transient types (SNIa, SNII, and SNIb/c). Transfer learning still provides substantial improvement even when focusing solely on these well-represented classes.

E. Training Time and Computational Efficiency

Another advantage of transfer learning beyond reducing data requirements is reduced training time, which is especially important in the era of data-driven astronomy. We find that models initialised with pretrained weights converge 25% faster than those trained from scratch for two reasons. First, the model begins with an understanding of general transient behaviour, requiring fewer iterations to learn class-specific features. Second, with many layers frozen, there are fewer parameters to optimise during training.

The reduction in training time, while not critical on modern GPUs (training from scratch takes approximately 15 minutes on a Tesla V100), becomes significant when processing large datasets or when computational resources are limited.

For example, when testing multiple model architectures or performing cross-validation with many training runs, the 25% reduction in training time per model can substantially reduce total computation time.

---

[2]The public MSIP ZTF survey has since changed to a 2-day median cadence

[3]None of the upsampled classes are used in this work