Simulation-Based Pretraining and Domain Adaptation for Astronomical Time Series Tasks with Minimal Labeled Data

Rithwik Gupta¹² Daniel Muthukrishna¹ Jeroen Audenaert¹

Abstract

Astronomical time-series analysis faces a critical limitation: the scarcity of labeled observational data. We present a pre-training approach that leverages simulations, significantly reducing the need for labeled examples from real observations. Our models, trained on simulated data from multiple astronomical surveys (ZTF and LSST), learn generalizable representations that transfer effectively to downstream tasks. Using classifier-based architectures enhanced with contrastive and adversarial objectives, we create domain-agnostic models that demonstrate substantial performance improvements over baseline methods in classification, redshift estimation, and anomaly detection when fine-tuned with minimal real data. Remarkably, our models exhibit effective zero-shot transfer capabilities, achieving comparable performance on future telescope (LSST) simulations when trained solely on existing telescope (ZTF) data. Furthermore, they generalize to very different astronomical phenomena (namely variable stars from NASA's Kepler telescope) despite being trained on transient events, demonstrating cross-domain capabilities. Our approach provides a practical solution for building general models when labeled data is scarce, but domain knowledge can be encoded in simulations.

1. Introduction

Time-series analysis in astronomy often requires substantial labeled data for supervised learning approaches. Models have been developed to classify variable stars and transient events (e.g. Muthukrishna et al., 2019; Hložek et al., 2023; Audenaert, 2025), detect anomalies (Villar et al., 2021; Muthukrishna et al., 2022; Perez-Carrasco et al., 2023), and estimate physical parameters such as redshift (e.g Qu & Sako, 2023; Zhang et al., 2024). However, these models are typically trained for specific telescopes or individual tasks, with limited ability to transfer learned representations across different instruments. Furthermore, they require extensive labeled data from new telescopes to achieve adequate performance, limiting their effectiveness during the initial survey stages.

Foundation models (FMs) have transformed natural language processing and computer vision by learning generalizable representations from large quantities of data (Bommasani et al., 2022). However, astronomical data presents unique challenges: (1) limited publicly available labeled data; (2) instrument-specific characteristics that hinder crosssurvey generalization (Audenaert et al., 2025); and (3) complex physical phenomena that require domain expertise to model effectively. Recent work has begun to develop foundation models for astrophysical data (e.g. Parker et al., 2024; Donoso-Oliva et al., 2023; Smith et al., 2024; Zhang et al., 2024; Audenaert et al., 2025). While these approaches demonstrate cross-modal learning, no existing approach achieves zero-shot transfer between different telescopes for time-domain phenomena. Moreover, existing FMs typically use self-supervised methods that do not take advantage of the astronomical class structure that is fundamental to astronomical understanding.

Astronomy has a significant advantage over many domains: decades of physical understanding encoded in simulations. Astronomers have developed detailed models of astrophysical phenomena that generate synthetic light curves (e.g. PLAsTiCC Modelers, 2019). While these simulations contain systematic differences from real observations, they capture the underlying physical processes and can provide effective pretraining for transfer to real data (Gupta et al., 2025).

Motivated by the success of supervised classifier-based methods for anomaly detection (Gupta et al., 2024), we propose a novel approach to learning generalizable representations for astronomical time series that:

- 1. Leverages the latent space of classifiers pretrained on simulations from multiple astronomical surveys
- 2. Develops domain-agnostic representations through adversarial and contrastive learning objectives

¹Massachusetts Institute of Technology, Cambridge, MA 02139, USA ²Irvington High School, Fremont, CA 94538, USA. Correspondence to: Daniel Muthukrishna <danmuth@mit.edu>, Rithwik Gupta <rithwikca2020@gmail.com>.

ML4Astro 2025, Vancouver, CA. Copyright 2025 by the author(s).



Figure 1. Overview of our simulation-based pre-training methodology showing three model variants (Classifier, Adversarial, Contrastive) and evaluation on both in-domain and cross-domain tasks. We first pre-train various classifiers and domain-agnostic models using simulated data. We then evaluate these models on various downstream tasks, including zero-shot estimation for new telescopes.

- 3. Enables effective downstream task performance with minimal labeled real data
- 4. Demonstrates zero-shot transfer capabilities between different survey telescopes

Unlike existing approaches that focus on single-survey applications, our method specifically targets cross-survey transfer learning using physics-informed simulations.

This approach is particularly valuable for upcoming surveys like the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST), which will produce millions of time-series alerts nightly (Ivezić et al., 2019). Having models ready to analyze LSST data from day one—without requiring extensive new labeled datasets—would dramatically accelerate scientific discovery.¹

2. Datasets and Benchmarks

2.1. Pretraining Data

We pretrain our models using 151,468 simulated astronomical transients from two survey telescopes: 87,080 from the Zwicky Transient Facility (ZTF; Bellm et al., 2018) and 64,388 from LSST. We evaluate three distinct pretraining configurations: (1) ZTF simulations only, (2) combined ZTF and LSST simulations, and (3) combined ZTF and LSST simulations with domain-agnostic objectives (adversarial or contrastive training).

The simulated transients encompass eight astronomical classes: Type Ia, Ib/c, II, Iax, and Ia-91bg supernovae; Tidal Disruption Events; Super-Luminous Supernovae Type I; and

Active Galactic Nuclei (see Table 3 for class distributions).

ZTF simulations use established astrophysical models (Kessler et al., 2019; Muthukrishna et al., 2019) reflecting the currently operational survey's characteristics including two photometric bands (gr) with limiting magnitude $r \sim 21$ mag. LSST simulations (Narayan & ELAsTiCC Team, 2023) model the Rubin Observatory beginning operations in late 2025: six photometric bands (ugrizy) with significantly deeper sensitivity ($r \sim 25$ mag).

Each transient is represented as a multichannel variablelength time series (light curve²) with observations formatted as $[\lambda_p, t_i, f_i]$, where λ_p indicates the median passband wavelength, t_i is the time since the first observation (in days), and f_i is the normalized flux (brightness measurement). This standardized representation enables cross-survey compatibility by encoding temporal and spectral information in a telescope-agnostic format.

2.2. Evaluation Datasets and Downstream Tasks

We evaluate pretrained models by fine-tuning on downstream tasks using real observational data and additional simulated datasets to assess in-domain performance, crossdomain generalization, and zero-shot transfer capabilities.

ZTF Real Data: 3,747 spectroscopically confirmed transients from the ZTF Bright Transient Survey (Rehemtulla et al., 2024) spanning three task types:

• *Classification*: Macro-averaged Area Under the ROC Curve (AUROC) for distinguishing Type Ia (771 objects), Type Ib/c (2,828 objects), and Type II (148

¹The code used in this work is publicly available: https://github.com/Rithwik-G/astrofm2.0 and https://github.com/Rithwik-G/Kepler-FM

²See Figure 1 for a plot of example light curves

objects) supernovae

- *Redshift Estimation*: Mean Squared Error (MSE) for predicting spectroscopic redshifts using photometric light curve data from the same 3,747 objects
- Anomaly Detection (AD): Binary classification AU-ROC treating 38 rare transient types as anomalies. The models are not provided any anomalous data during training.

Simulated Redshift Estimation: 10,000 ZTF and 9,990 LSST simulated objects with redshift labels for regression tasks (MSE evaluation). These datasets are distinct from the pretraining classification data and the pretrained models have no prior exposure to redshift information and must learn this novel regression task during fine-tuning. For zeroshot evaluation, we use a subset of 2,596 LSST objects restricted to ZTF's redshift range (z < 0.4), where models fine-tuned exclusively on ZTF redshift data predict LSST redshifts without any LSST-specific training redshifts.

Kepler Real Data (Cross-Domain): 9,501 variable stars from NASA's Kepler telescope (Audenaert et al., 2021) spanning eight stellar variability classes (aperiodic, constant, contact binaries, Delta Scuti, Beta Cephei, eclipsing binaries, Gamma Doradus, SPB variables) to test generalization from transient to variable star *classification* (macroaveraged AUROC). Kepler is a space telescope designed for high-cadence (30-minute) photometric monitoring in a single broad optical band, contrasting sharply with the multi-band, 3-day cadence observations of ground-based transient surveys like ZTF and LSST.

2.3. Evaluation Scenarios

To simulate realistic deployment challenges when applying pretrained models to new surveys with varying data availability, we evaluate under three scenarios:

- **Limited Scenario:** Fine-tunes pretrained models using only 512 randomly selected labeled objects per task. We repeat experiments with five different random samples to account for selection variance.
- **Full Scenario:** Fine-tunes pretrained models using all available labeled data for each task, representing performance after extensive labeling efforts.
- Zero-Shot Scenario: Applied exclusively to LSST redshift estimation, where models fine-tuned only on ZTF redshift data are directly applied to LSST simulations without any LSST-specific fine-tuning data. This tests immediate model deployment capability on new surveys before domain-specific labeling becomes available.

The zero-shot redshift evaluation deliberately restricts LSST objects to ZTF's redshift range because our goal is leverag-

ing existing telescope knowledge for immediate application to new surveys. While LSST will observe significantly higher redshifts than ZTF (see Figure 2), zero-shot models cannot reasonably extrapolate beyond their training distribution.

3. Methods

We develop three simulation-based pretraining approaches for learning generalizable representations of astronomical time series: a classifier trained on ZTF and LSST simulations, and a classifier enhanced with domain-agnostic objectives using either an adversarial objective or a contrastive objective for explicit survey alignment.

3.1. Base Classifier

Neural network classifiers have demonstrated the ability to learn meaningful representations of astronomical phenomena that transfer effectively to tasks beyond their original classification objective, including anomaly detection and morphological analysis (Walmsley et al., 2022; Etsebeth et al., 2023; Gupta et al., 2025). Building on these successes, our base model trains a classifier on simulated transients from both ZTF and LSST to distinguish between the eight astronomical transient classes.

Architecture: We employ a Gated Recurrent Unit (GRU; Cho et al. 2014) network that processes variable-length time series inputs. Each observation $[\lambda_p, t_i, f_i]$ is processed sequentially through the 100-unit GRU layers, with the final hidden state passed through two fully-connected mutililayer perceptron (MLP) layers to produce class predictions. The model is trained using a standard categorical cross-entropy loss H on the final softmax predictions. We extract the penultimate MLP layer as our learned latent representation $z_i \in \mathbb{R}^{128}$ for downstream tasks. We refer to the classifier with its final classification layer removed as the encoder C_L throughout this work, such that $z_i = C_L(X_i)$ for each input light curve X_i .

While this classifier learns structured representations that distinguish astronomical classes within each survey (see Fig. 4 of Gupta et al., 2024), it does not produce a telescopeagnostic latent space. Objects from the same class but different surveys remain separated in the latent space (see Figures 4 and 5 for UMAP visualizations of the latent space).

3.2. Adversarial Training

Different telescopes have distinct observational characteristics (photometric bands, sensitivity, cadence, noise) that can cause models to learn survey-specific artifacts rather than the underlying physics. To create domain-agnostic representations that work across different astronomical surveys, we implement an adversarial training framework that encourages the model to learn telescope-invariant features useful for astronomical classification.

Architecture: We extend the base classifier with an additional discriminator network D that attempts to predict whether a representation came from ZTF or LSST simulations. The discriminator is a two-layer MLP that takes the latent representation z_i as input to predict the survey telescope.

Training: The classifier $C(\mathbf{X}_i)$ and the discriminator $D(\mathbf{z}_i)$ are trained simultaneously with competing objectives,

Classifier loss:
$$\mathcal{L}_C = H(C(\mathbf{X}_i), c_i)$$
 (1)

Discriminator loss:
$$\mathcal{L}_D = H(D(\boldsymbol{z}_i), O_i)$$
 (2)

where H denotes the cross-entropy loss, c_i is the class label, and $O_i \in \{\text{ZTF}, \text{LSST}\}$ indicates the source survey. The model is trained to both classify the transient correctly and to confuse the discriminator, thereby learning a latent space that is useful for classification yet agnostic to the survey domain. The adversarial training process is detailed further in Appendix B.2.

3.3. Supervised Contrastive Training

While adversarial training encourages domain-agnostic representations, it does not explicitly enforce that objects of the same class cluster together across surveys. Hence, as a comparative method, we use supervised contrastive learning (Khosla et al., 2020), which explicitly pulls together representations of the same class from different surveys while pushing apart different classes. We use a variant of the contrastive loss proposed in Chen et al. (2020) for model pretraining where we treat all samples belonging to the same class as positive pairs. This modification provides a more explicit signal for class-based alignment and naturally encourages unification across domains, as long as examples from the same class are drawn from both surveys. The total loss is the sum of the classifier loss \mathcal{L}_C and the supervised contrastive loss \mathcal{L}_{SCL} . The training details and algorithm is described in Appendix B.3.

3.4. Baselines

We compare our pretrained models against models trained directly on downstream tasks without pretraining, which reflects the standard approach in prior work. We also train classifiers on individual surveys (notably ZTF-only) as additional baselines to isolate the benefits of multi-survey pretraining.

3.5. Downstream Tasks

For downstream tasks, we freeze the pretrained encoder and attach task-specific multi-layer perceptrons (MLPs). This preserves the learned cross-survey representations while adapting to specific objectives (classification, regression, anomaly detection). The fine tuning for each task is set up as follows (see Appendix B.4 for further details):

- **Classification/Redshift estimation:** Two-layer MLP with appropriate output activation (softmax for classification, linear for regression).
- **Anomaly Detection:** Following Gupta et al. (2025), we train an isolation forest on frozen latent representations, treating rare transient classes as anomalies
- **Cross-Domain Tasks (Kepler):** For novel domains, we unfreeze the encoder layers to allow limited adaptation while preserving the core learned representations
- **Zero-Shot Transfer:** For zero-shot evaluation on LSST redshift estimation, we use a two-layer MLP to finetune the pretrained models on ZTF redshift data only, and then evaluate on LSST simulations without any LSST training data. We test two approaches: direct MLP prediction and kNN (k=100) using distanceweighted averaging of the closest ZTF embeddings.

4. Results

We evaluate our simulation-based pretraining models on multiple downstream tasks using real observational data from ZTF (3,747 transients) and Kepler (9,501 variable stars), as well as simulated redshift estimation data from ZTF (10,000 objects) and LSST (9,990 objects). We compare both using the "Full" dataset for fine-tuning and using only a "Limited" set of 512 objects. Our experiments demonstrate that pretraining on simulations provides substantial improvements over training from scratch (no pretraining), with particularly pronounced benefits for cross-survey generalization and zero-shot transfer.

4.1. Simulation-Based Pretraining Effectiveness

Substantial Improvements from Simulation Pretraining: Table 1 shows that the pretrained models consistently outperform no-pretraining baselines across all downstream tasks. Most remarkably, our pretrained models fine-tuned with only 512 labeled examples (Limited scenario) outperform baseline models trained on the full dataset (3747 labeled transients) for both classification and anomaly detection tasks on real data.

Cross-Domain Generalization: Our models demonstrate impressive generalization capabilities, successfully transferring from explosive transient phenomena to stellar variability. On Kepler variable star classification, the best pretrained classifier achieves 0.968 ± 0.018 AUROC compared to 0.901 ± 0.003 without pretraining. This cross-domain

Simulation-Based Pretraining and Domain Adaptation for Astronomical Time Series

			ZTF Re	al Data	Kepler	Simulations			
Model	Classification (AUROC)		$\begin{array}{c c} \textbf{Redshift} \\ (\textbf{MSE} \times 10^3) \end{array}$		Anomaly Det. (AUROC)		Classification (AUROC)	Redshift (MSE)	
	Lim.	Full	Lim.	Full	Lim.	Full	Full	ZTF	LSST
Baseline: No Pretraining	0.637 ±0.005	0.853 ±0.013	6.02 ±0.05	3.85 ±0.31	0.498 ±0.013	0.527 ±0.027	0.901 ±0.003	0.079 ±0.022	0.289 ±0.018
Classifier (ZTF only)	0.875 ±0.020	0.904 ±0.012	4.91 ±0.10	3.87 ±0.36	0.605 ±0.025	0.596 ±0.008		0.028 ±0.003	0.252 ±0.004
Classifier (ZTF + LSST)	0.879 ±0.011	0.910 ±0.013	4.79 ±0.06	3.82 ±0.39	0.622 ±0.018	0.616 ±0.036	0.968 ±0.006	0.026 ±0.002	0.177 ±0.008
Classifier + Contrastive	0.886 ±0.026	0.914 ±0.005	4.87 ±0.03	3.73 ±0.17	0.584 ±0.018	0.576 ±0.028	0.946 ±0.021	0.028 ±0.002	0.191 ±0.013
Classifier + Adversarial	0.844 ±0.016	0.853 ±0.012	5.20 ±0.04	4.19 ±0.42	0.559 ±0.024	0.546 ±0.077	0.925 ±0.015	0.030 ±0.003	0.197 ±0.014

Table 1. Performance comparison across downstream tasks for each pretrained model. We report mean \pm standard deviation from five different random fine-tuning samples. Bold indicates best performance per task. The Classifier (ZTF + LSST), Contrastive, and Adversarial models represent our main contributions. For Kepler classification, "—" indicates the model was not evaluated on this cross-domain task.

	Model	Redshi ZTF	ifting Data LSST	LSST MSE
сt	No Pretraining	No	Yes	0.0750 ± 0.0252
ire	No Pretraining	Yes	Yes	0.0614 ± 0.0033
ã	Classifier	Yes	Yes	0.0579 ± 0.0061
	No Pretraining	Yes	No	0.1869 ± 0.0127
pol	Classifier	Yes	No	0.1035 ± 0.0081
ero-s	Classifier + Contrastive	Yes	No	$\textbf{0.0727} \pm \textbf{0.0056}$
	Classifier + Adversarial	Yes	No	$\textbf{0.0744} \pm \textbf{0.0063}$
N	Classifier + Contrastive kNN	Yes	No	0.0854 ± 0.0040

Table 2. Zero-shot redshift estimation performance on LSST simulations. Direct training methods use LSST labels during training, while zero-shot methods are trained only on ZTF data and evaluated on LSST without any LSST training redshifts. Bold indicates best zero-shot performance. Domain-agnostic models (Contrastive, Adversarial) achieve performance comparable to direct training methods.

transfer is particularly remarkable given the distinct observational characteristics of space-based Kepler versus groundbased ZTF/LSST surveys and the fundamental differences between explosive transients and variable stars. The substantial performance improvement on very different data demonstrates an unexpected emergent capability of our pretraining method.

4.2. Zero-Shot Transfer Performance

Table 2 reveals that the additional contrastive and adversarial objectives provide their most significant advantages in zero-shot scenarios. Most remarkably, our contrastive zero-shot approach achieves 0.0727 ± 0.0056 MSE, actually outperforming a model trained directly on LSST redshift data (0.0750 \pm 0.0252 MSE) and approaching the performance of models trained on both ZTF and LSST redshifts (0.0614 \pm 0.0033 MSE).

This performance demonstrates a clear trade-off between fine-tuning and zero-shot capabilities. While the standard classifier outperforms contrastive and adversarial variants on supervised fine-tuning tasks (Table 1), it shows substantially degraded zero-shot performance (0.1035 \pm 0.0081 MSE versus 0.0727 \pm 0.0056 MSE for contrastive). These results indicate that explicit cross-survey alignment through domain-agnostic objectives is essential for effective zero-shot transfer but may introduce constraints that limit performance when target-domain fine-tuning data is available.

5. Conclusion

We present a simulation-based pretraining approach that addresses the scarcity of labeled astronomical data by leveraging synthetic light curves from multiple surveys. Our models demonstrate substantial improvements over training from scratch, with pretrained models using only 512 real labeled examples significantly outperforming baselines trained on full datasets with $7 \times$ more data for classification and anomaly detection tasks.

Key findings include: (1) effective cross-domain transfer from explosive transients (ZTF and LSST ground-based surveys) to stellar variability (Kepler space telescope) despite fundamental astrophysical and instrumental differences, revealing unexpected emergent capabilities, and (2) domain-agnostic training objectives that enable zero-shot performance matching models trained directly on target survey data, with our contrastive model achieving comparable LSST redshift performance without any LSST training data.

These findings have immediate relevance for LSST beginning operations in late 2025. Our models can be deployed from day one without extensive domain-specific labeling, dramatically accelerating early science returns. The demonstrated improvement in data efficiency addresses the perennial challenge of limited expert annotations in astronomy. Our approach provides a practical framework for scientific domains where simulations encode substantial domain expertise but labeled observational data remains scarce.

Acknowledgements

This work used Bridges-2 at Pittsburgh Supercomputing Center through allocation PHY240105 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program (Boerner et al., 2023), which is supported by U.S. National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

This work made use of the python programming language and the following packages: numpy (Harris et al., 2020), matplotlib (Hunter, 2007), scikit-learn (Pedregosa et al., 2011), pandas (McKinney, 2010), astropy (Astropy Collaboration et al., 2022), and pytorch (Paszke et al., 2019).

References

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., Donath, 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. *The Astrophysical Journal*, 935(2):167, August 2022. doi: 10.3847/1538-4357/ac7c74.

- Audenaert, J. From stellar light to astrophysical insight: automating variable star research with machine learning. *arXiv e-prints*, art. arXiv:2507.03093, July 2025. doi: 10.48550/arXiv.2507.03093.
- Audenaert, J., Kuszlewicz, J. S., Handberg, R., Tkachenko, A., Armstrong, D. J., Hon, M., Kgoadi, R., Lund, M. N., Bell, K. J., Bugnet, L., Bowman, D. M., Johnston, C., García, R. A., Stello, D., Molnár, L., Plachy, E., Buzasi, D., and Aerts, C. Tess data for asteroseismology (t'da) stellar variability classification pipeline: Setup and application to the kepler q9 data. *The Astronomical Journal*, 162(5):209, October 2021. ISSN 1538-3881. doi: 10.3847/1538-3881/ac166a. URL http: //dx.doi.org/10.3847/1538-3881/ac166a.
- Audenaert, J., Muthukrishna, D., Gregory, P., Hogg, D., and Villar, V. A. Causal Foundation Models: Disentangling Physics from Instrument Properties. *ICML 2025 Work*shop on Foundation Models for Structured Data, 2025.
- 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, Ż., 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., 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., Sistine, A. V., 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. *Publications of the Astronomical Society of the Pacific*, 131(995):018002, dec 2018. doi: 10.1088/1538-3873/aaecbe. URL https://doi. org/10.1088%2F1538-3873%2Faaecbe.

- 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.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., Donahue, C., Doumbouya, M., Durmus, E., Ermon, S., Etchemendy, J., Ethayarajh, K., Fei-Fei, L., Finn, C., Gale, T., Gillespie, L., Goel, K., Goodman, N., Grossman, S., Guha, N., Hashimoto, T., Henderson, P., Hewitt, J., Ho, D. E., Hong, J., Hsu, K., Huang, J., Icard, T., Jain, S., Jurafsky, D., Kalluri, P., Karamcheti, S., Keeling, G., Khani, F., Khattab, O., Koh, P. W., Krass, M., Krishna, R., Kuditipudi, R., Kumar, A., Ladhak, F., Lee, M., Lee, T., Leskovec, J., Levent, I., Li, X. L., Li, X., Ma, T., Malik, A., Manning, C. D., Mirchandani, S., Mitchell, E., Munyikwa, Z., Nair, S., Narayan, A., Narayanan, D., Newman, B., Nie, A., Niebles, J. C., Nilforoshan, H., Nyarko, J., Ogut, G., Orr, L., Papadimitriou, I., Park, J. S., Piech, C., Portelance, E., Potts, C., Raghunathan, A., Reich, R., Ren, H., Rong, F., Roohani, Y., Ruiz, C., Ryan, J., Ré, C., Sadigh, D., Sagawa, S., Santhanam, K., Shih, A., Srinivasan, K., Tamkin, A., Taori, R., Thomas, A. W., Tramèr, F., Wang, R. E., Wang, W., Wu, B., Wu, J., Wu, Y., Xie, S. M., Yasunaga, M., You, J., Zaharia, M., Zhang, M., Zhang, T., Zhang, X., Zhang, Y., Zheng, L., Zhou, K., and Liang, P. On the opportunities and risks of foundation models, 2022. URL https://arxiv.org/abs/2108.07258.
- Boone, K. Avocado: Photometric classification of astronomical transients with gaussian process augmentation. *The Astronomical Journal*, 158(6):257, dec 2019. doi: 10.3847/1538-3881/ab5182. URL https://doi.org/10.3847%2F1538-3881%2Fab5182.
- Borucki, W. J., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Dupree, A. K., Gautier, T. N., Geary, J. C., Gilliland, R., Gould, A., Howell, S. B., Jenkins, J. M., Kondo, Y., Latham,

D. W., Marcy, G. W., Meibom, S., Kjeldsen, H., Lissauer, J. J., Monet, D. G., Morrison, D., Sasselov, D., Tarter, J., Boss, A., Brownlee, D., Owen, T., Buzasi, D., Charbonneau, D., Doyle, L., Fortney, J., Ford, E. B., Holman, M. J., Seager, S., Steffen, J. H., Welsh, W. F., Rowe, J., Anderson, H., Buchhave, L., Ciardi, D., Walkowicz, L., Sherry, W., Horch, E., Isaacson, H., Everett, M. E., Fischer, D., Torres, G., Johnson, J. A., Endl, M., MacQueen, P., Bryson, S. T., Dotson, J., Haas, M., Kolodziejczak, J., Van Cleve, J., Chandrasekaran, H., Twicken, J. D., Quintana, E. V., Clarke, B. D., Allen, C., Li, J., Wu, H., Tenenbaum, P., Verner, E., Bruhweiler, F., Barnes, J., and Prsa, A. Kepler Planet-Detection Mission: Introduction and First Results. *Science*, 327(5968):977, February 2010. doi: 10.1126/science.1185402.

- Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A simple framework for contrastive learning of visual representations, 2020. URL https://arxiv.org/abs/2002.05709.
- 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.
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014.
- Donoso-Oliva, C., Becker, I., Protopapas, P., Cabrera-Vives, G., Vishnu, M., and Vardhan, H. Astromer: A transformer-based embedding for the representation of light curves. *Astronomy and Astrophysics*, 670: A54, February 2023. ISSN 1432-0746. doi: 10.1051/0004-6361/202243928. URL http://dx.doi.org/10.1051/0004-6361/202243928.
- Etsebeth, V., Lochner, M., Walmsley, M., and Grespan, M. Astronomaly at scale: Searching for anomalies amongst 4 million galaxies, 2023.
- Gupta, R., Muthukrishna, D., and Lochner, M. A classifierbased approach to multi-class anomaly detection for astronomical transients, 2024. URL https://arxiv. org/abs/2403.14742.
- Gupta, R., Muthukrishna, D., and Lochner, M. A classifierbased approach to multiclass anomaly detection for astronomical transients. *RAS Techniques and Instruments*, 4: rzae054, January 2025. doi: 10.1093/rasti/rzae054.

- Gupta, R., Muthukrishna, D., Rehemtulla, N., and Shah, V. Transfer learning for transient classification: From simulations to real data and ztf to lsst. *Monthly Notices of the Royal Astronomical Society: Letters*, pp. slaf074, 07 2025. ISSN 1745-3925. doi: 10. 1093/mnrasl/slaf074. URL https://doi.org/10. 1093/mnrasl/slaf074.
- 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). *The Astrophysical Journal Supplement Series*, 267(2):25, August 2023. doi: 10.3847/1538-4365/accd6a.
- Huang, H., Muthukrishna, D., Nair, P., Zhang, Z., Fausnaugh, M., Majumder, T., Foley, R. J., and Ricker, G. R. Predicting the age of astronomical transients from real-time multivariate time series. *arXiv preprint arXiv:2311.17143*, 2023.
- 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. *The Astrophysical Journal*, 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). *Publications of the Astronomical Society of the Pacific*, 131(1003):094501, jul 2019. doi: 10.1088/1538-3873/ab26f1. URL https://doi.org/10.1088%2F1538-3873%2Fab26f1.

- Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., and Krishnan, D. Supervised contrastive learning. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), Advances in Neural Information Processing Systems, volume 33, pp. 18661–18673. Curran Associates, Inc., 2020. URL https://proceedings.neurips. cc/paper_files/paper/2020/file/ d89a66c7c80a29b1bdbab0f2a1a94af8-Paper. pdf.
- Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization, 2017.
- Koch, D. G., Borucki, W. J., Basri, G., Batalha, N. M., Brown, T. M., Caldwell, D., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Gautier, Thomas N., I., Geary, J. C., Gilliland, R. L., Gould, A., Jenkins, J., Kondo, Y., Latham, D. W., Lissauer, J. J., Marcy, G., Monet, D., Sasselov, D., Boss, A., Brownlee, D., Caldwell, J., Dupree, A. K., Howell, S. B., Kjeldsen, H., Meibom, S., Morrison, D., Owen, T., Reitsema, H., Tarter, J., Bryson, S. T., Dotson, J. L., Gazis, P., Haas, M. R., Kolodziejczak, J., Rowe, J. F., Van Cleve, J. E., Allen, C., Chandrasekaran, H., Clarke, B. D., Li, J., Quintana, E. V., Tenenbaum, P., Twicken, J. D., and Wu, H. Kepler Mission Design, Realized Photometric Performance, and Early Science. The Astrophysical Journal Letters, 713(2):L79-L86, April 2010. doi: 10.1088/2041-8205/713/2/L79.
- Liu, F. T., Ting, K. M., and Zhou, Z.-H. Isolation forest. In 2008 eighth ieee international conference on data mining, pp. 413–422. IEEE, 2008.
- McInnes, L., Healy, J., and Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction, 2020.
- 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.
- Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., and Hložek, R. RAPID: Early classification of explosive transients using deep learning. *Publications of the Astronomical Society of the Pacific*, 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 large-scale transient surveys. *Monthly Notices of the Royal Astronomical Society*, 517(1):393–419, November 2022. doi: 10.1093/mnras/stac2582.
- Narayan, G. and ELAsTiCC Team. The Extended LSST Astronomical Time-series Classification Challenge (ELAs-TiCC). In American Astronomical Society Meeting Abstracts, volume 241 of American Astronomical Society Meeting Abstracts, pp. 117.01, January 2023.
- Parker, L., Lanusse, F., Golkar, S., Sarra, L., Cranmer, M., Bietti, A., Eickenberg, M., Krawezik, G., McCabe, M., Morel, R., Ohana, R., Pettee, M., Régaldo-Saint Blancard, B., Cho, K., and Ho, S. Astroclip: a cross-modal foundation model for galaxies. *Monthly Notices of the Royal Astronomical Society*, 531(4):4990–5011, June 2024. ISSN 1365-2966. doi: 10.1093/mnras/stae1450. URL http: //dx.doi.org/10.1093/mnras/stae1450.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. Pytorch: An imperative style, high-performance deep learning library, 2019. URL https://arxiv.org/abs/1912.01703.
- 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.
- Perez-Carrasco, M., Cabrera-Vives, G., Hernandez-García, L., Förster, F., Sanchez-Saez, P., Arancibia, A. M. M., Arredondo, J., Astorga, N., Bauer, F. E., Bayo, A., Catelan, M., Dastidar, R., Estévez, P. A., Lira, P., and Pignata, G. Alert classification for the alerce broker system: The anomaly detector. *The Astronomical Journal*, 166(4):151, sep 2023. doi: 10.3847/1538-3881/ace0c1. URL https: //dx.doi.org/10.3847/1538-3881/ace0c1.
- PLAsTiCC Modelers. Libraries & Recommended Citations for using PLAsTiCC Models, March 2019. URL https: //doi.org/10.5281/zenodo.2612896.
- Qu, H. and Sako, M. Photo-zsnthesis: Converting type ia supernova lightcurves to redshift estimates via deep learning, 2023. URL https://arxiv.org/abs/ 2305.11869.

- 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. *The Astrophysical Journal*, 972(1):7, aug 2024. doi: 10.3847/ 1538-4357/ad5666. URL https://dx.doi.org/ 10.3847/1538-4357/ad5666.
- Smith, M. J., Roberts, R. J., Angeloudi, E., and Huertas-Company, M. Astropt: Scaling large observation models for astronomy, 2024. URL https://arxiv.org/ abs/2405.14930.
- Villar, V. A., Cranmer, M., Berger, E., Contardo, G., Ho, S., Hosseinzadeh, G., and Lin, J. Y.-Y. A deep-learning approach for live anomaly detection of extragalactic transients. *The Astrophysical Journal Supplement Series*, 255 (2):24, 2021.
- Walmsley, M., Scaife, A. M. M., Lintott, C., Lochner, M., Etsebeth, V., Géron, T., Dickinson, H., Fortson, L., Kruk, S., Masters, K. L., Mantha, K. B., and Simmons, B. D. Practical galaxy morphology tools from deep supervised representation learning. *Monthly Notices of the Royal Astronomical Society*, 513(2):1581–1599, 02 2022. ISSN 0035-8711. doi: 10.1093/mnras/stac525. URL https: //doi.org/10.1093/mnras/stac525.
- Zhang, G., Helfer, T., Gagliano, A. T., Mishra-Sharma, S., and Villar, V. A. Maven: A multimodal foundation model for supernova science, 2024. URL https://arxiv. org/abs/2408.16829.

A. Dataset Details

A.1. ZTF and LSST simulations

Table 3 provides the detailed class distribution for our simulated pretraining datasets. The redshift distribution is illustrated in Figure 2.

A.1.1. ZERO-SHOT REDSHIFT RANGE RESTRICTION

We implement a redshift cut (z < 0.4) for zero-shot LSST evaluation to match the ZTF training distribution and to prevent distribution shift. This reduces the LSST evaluation set from 9,990 to 2,596 objects but ensures that performance differences reflect instrumental transfer rather than extrapolation failure.

LSST will observe transients up to $z \sim 1.2$ (Figure 2), but zero-shot models trained only on ZTF data (z < 0.4) cannot reliably predict redshifts beyond their training range. Evaluating on the full LSST redshift range would conflate two distinct problems: instrumental differences between surveys and extrapolation to unseen redshift regimes. For higher-redshift LSST objects, we perform survey-specific fine-tuning (Table 1) for each pretraining approach.

A.2. ZTF Real Data

Table 4 details the class distribution for our real ZTF evaluation dataset from the Bright Transient Survey (BTS). All objects have spectroscopic confirmation with the redshift distribution illustrated in Figure 3. The redshift distribution is much lower as the BTS preferences classifying brighter, and hence lower redshift transients. We do not use any real light curves for pretraining and reserve them solely for evaluation

For anomaly detection experiments, we designate the following rare transient classes as anomalies: Tidal Disruption Events (TDE), Calcium-rich transients, Intermediate Luminosity Red Transients (ILRT), Luminous Blue Variables (LBV), Luminous Red Novae (LRN), Super-Luminous Supernovae Types I and II (SLSN-I, SLSN-II), peculiar Type Ia subtypes (SN Ia-91T, SN Ia-91bg), Type Ibn supernovae (SN Ibn), Broad-Line Type Ic supernovae (SN Ic-BL), and Type Icn supernovae (SN Icn). These classes represent < 1% of our real transient dataset and were selected as anomalies because of their low observation rates.

A.3. Kepler Variable Stars

We evaluate cross-domain generalization using 9,501 variable star light curves from the Kepler mission (Koch et al., 2010; Borucki et al., 2010), processed by Audenaert et al. (2021). The dataset spans eight stellar variability classes from aperiodic variables to regular pulsators (Table 5).

To ensure compatibility with our transient survey pretraining data, we downsample the high-cadence Kepler observations from 1024 measurements (30-minute sampling) to \sim 205 measurements by averaging every five consecutive observations. This preprocessing matches the typical observation count in ZTF and LSST light curves while preserving the essential variability signatures needed for classification.

This cross-domain evaluation tests whether models pretrained on explosive transients can generalize to fundamentally different astrophysical phenomena—stellar pulsations and eclipsing systems—despite having no exposure to periodic variability during training.

B. Model and Training Details

B.1. Classifier Training

We enable cross-survey compatibility by using a surveyagnostic input $[\lambda_p, t_i, f_i]$ format described in Section 2. This input method (Huang et al., 2023; Gupta et al., 2024; 2025) specifically allows for the usage of the same model across surveys, something not facilitated by many previous input methods.

Training uses categorical cross-entropy loss with early stopping when validation loss plateaus for 5 epochs. We extract representations from the penultimate layer (128dimensional) as our learned embeddings $z_i = C_L(X_i)$ for downstream tasks.

B.2. Adversarial Training Algorithm

Our adversarial pretraining is summarized in Algorithm 1. Here, X_i denotes the input light curve, c_i is its class label, and $O_i \in \{\text{ZTF}, \text{LSST}\}$ indicates the observatory. The latent representation is extracted as $z_i = C_L(X_i)$, where C_L is the penultimate layer of the classifier. The categorical cross-entropy loss is denoted by H(p,q), where p is a predicted distribution and q is a target one-hot vector.

B.3. Contrastive Training Algorithm

For our supervised contrastive loss, we use the contrastive objective proposed in Chen et al. (2020) for model pretraining. It is formally defined as follows:

$$\ell_{i,j} = -\log \frac{\exp\left(\sin\left(\boldsymbol{z}_{i}, \boldsymbol{z}_{j}\right)/\tau\right)}{\sum_{k=1}^{2N} \mathbb{1}_{[k\neq i]} \exp\left(\sin\left(\boldsymbol{z}_{i}, \boldsymbol{z}_{k}\right)/\tau\right)} \quad (3)$$

where: z_i, z_j are the latent representations from the classifier, $\sin(z_i, z_j)$ denotes cosine similarity: $\sin(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}, \tau > 0$ is a temperature parameter that scales the similarity scores, and the loss is computed for all pairs (i, j) where X_i and X_j share the same class label.

Dataset	Astronomical Transient Classes								Total
Dataset	SNIa	SNIa-91bg	SNIax	SNIb/c	SNII	TDE	SLSN-I	AGN	1000
Pretraining Data (Classification)									
ZTF Simulations	9,436	10,663	10,681	6,769	31,193	9,260	10,451	8,627	87,080
LSST Simulations	8,427	6,079	8,298	7,664	9,465	9,686	6,947	7,822	64,388
Downstream Task Data (Redshift Estimation	ı)								
ZTF Simulations	967	1,140	1,124	702	3,213	932	1,053	869	10,000
LSST Simulations	1,236	1,238	1,226	1,286	1,248	1,279	1,253	1,234	9,990
Zero-Shot LSST Subset	265	273	535	258	333	215	446	271	2,596

Table 3. Class distribution of simulated astronomical transients. Classification data is used for pretraining, while redshift data (completely separate objects) tests downstream transfer learning. The zero-shot subset contains LSST objects within ZTF's redshift range (z < 0.4) for realistic evaluation.

Task	SNIa	SNIb/c	SNII	Anomaly	Total
Classification	771 (107)	2,828 (350)	148 (12)	0	3,747
Redshift Estimation	771 (107)	2,828 (350)	148 (12)	0	3,747
Anomaly Detection	771 (107)	2,828 (350)	148 (12)	0 (38)	3,785

Table 4. Real ZTF transient data distribution across downstream tasks. Numbers in parentheses indicate evaluation set sizes. Limited scenarios use 512 randomly selected objects per task from the training portion. All tasks share the same underlying object pool from the ZTF Bright Transient Survey.



Figure 2. Redshift distributions for ZTF and LSST simulated datasets. Most ZTF simulated transients span z = 0.0 - 0.4 while LSST simulations extend to $z \sim 3.0$, reflecting the deeper photometric sensitivity of LSST ($r \sim 25$ mag) compared to ZTF ($r \sim 21$ mag). For zero-shot evaluation, we restrict LSST objects to ZTF's redshift range (z < 0.4) to test transfer learning within the training distribution, as extrapolation beyond this range would is not easily feasible for zero-shot models.

Task	Aperiodic	Constant	Contact	DSCT/BCEP	Eclipse	GDOR/SPB	Instr	RR/CEP	Solar	Total
Kepler Classification	831	1,000	2,260	772	974	630	1,171	63	1,800	9,501

Table 5. Distribution of Kepler variable star data across eight stellar variability classes used for cross-domain evaluation. This dataset tests generalization from explosive transients (pretraining) to variable star phenomena, representing a fundamentally different astronomical domain with space-based, high-cadence observations.



Figure 3. Redshift distribution for our dataset of real ZTF objects

A	lgorithm	1	Ac	lversarial	Training
---	----------	---	----	------------	----------

Require: Dataset $\{(X_i, c_i, O_i)\}_{i=1}^N$: light curves, class labels, and observatory labels

Require: Classifier C, Discriminator D

- 1: Initialize C and D with random weights
- 2: repeat
- 3: // Step 1: Train the discriminator
- 4: Freeze the classifier C
- 5: For each sample, compute latent representation $z_i = C_L(X_i)$
- 6: Compute discriminator loss:

$$\mathcal{L}_D = H(D(\boldsymbol{z}_i), O_i)$$

- 7: Update D to minimize \mathcal{L}_D
- 8: // Step 2: Train the classifier
- 9: Freeze the discriminator D
- 10: Compute classifier loss with adversarial objective:

$$\mathcal{L}_C = H(C(\mathbf{X}_i), c_i) - H(D(C_L(\mathbf{X}_i)), O_i)$$

11: Update C to minimize \mathcal{L}_C

12: **until** convergence

The total supervised contrastive loss is computed by summing $\ell_{i,j}$ over all valid positive pairs in a batch. This encourages latent vectors from the same class to be close together, while implicitly pushing apart representations from other classes.

Our contrastive pretraining is summarized in Algorithm 2. We set $\tau = 0.5$ similar to the default set in previous work (Chen et al., 2020).

Algorithm 2 Supervised Contrastive Training

Require: Training set $\{(X_i, c_i)\}_{i=1}^N$; light curves, class labels

Require: Classifier C, temperature parameter τ

- 1: Initialize C with random weights
- 2: repeat
- 3: Compute classification loss

$$\mathcal{L}_C = H(C(\boldsymbol{X}_i), c_i)$$

- 4: Compute latent representations $z_i = C_L(X_i)$
- 5: Initialize total contrastive loss $\mathcal{L}_{SCL} \leftarrow 0$, counter $M \leftarrow 0$
- 6: for each anchor sample $i \in \{1, \dots, N\}$ do

7: Let
$$P(i) = \{j \neq i : c_j = c_i\}$$

- 8: for each $j \in P(i)$ do
- 9: Compute pairwise contrastive loss:

$$\ell_{i,j} = -\log \frac{\exp\left(\sin\left(\boldsymbol{z}_{i}, \boldsymbol{z}_{j}\right)/\tau\right)}{\sum_{k \neq i} \exp\left(\sin\left(\boldsymbol{z}_{i}, \boldsymbol{z}_{k}\right)/\tau\right)}$$

- 10: Accumulate loss: $\mathcal{L}_{SCL} \leftarrow \mathcal{L}_{SCL} + \ell_{i,j}$
- 11: Increment counter: $M \leftarrow M + 1$

12: end for

- 13: end for
- 14: Compute mean contrastive loss: $\mathcal{L}_{SCL} \leftarrow \mathcal{L}_{SCL}/M$
- 15: Compute total loss: $\mathcal{L} \leftarrow \mathcal{L}_{SCL} + \mathcal{L}_C$
- 16: Update C to minimize \mathcal{L}
- 17: until convergence

B.4. Fine-Tuning

B.4.1. GENERAL TASKS

For in-domain tasks (ZTF/LSST), we freeze the pretrained encoder and fine-tune only the task-specific MLP head. For cross-domain tasks (Kepler), we unfreeze the initial GRU layers to enable domain adaptation while preserving core learned representations. These is motivated by previous transfer learning research (Gupta et al., 2025).

B.4.2. ANOMALY DETECTION

Following (Gupta et al., 2024), we use a classifier-based approach where a fine-tuning classifier is trained on a set of



Figure 4. UMAP visualization of learned representations across all classes. The standard classifier maintains distinct ZTF (brown) and LSST (blue) clusters, while domain-agnostic training methods appear to unify surveys within each astronomical class slightly better.



Figure 5. Class-level alignment for Type II supernovae only. Domain-agnostic methods achieve better survey overlap in the latent representations while preserving astronomical class structure, indicating at possible cross-survey transfer capabilities.

normal data. The penultimate layer of this classifier is then used as a latent space for anomaly detection and an isolation forest (Liu et al., 2008) is trained using this latent space to detect anomalies. This method has state-of-the-art (SoTA) performance for anomaly detection on real data.

B.5. Architecture and Training

We implement a GRU-based encoder (Cho et al., 2014) with 100 hidden units and two fully-connected layers, producing 128-dimensional latent representations. GRUs provide superior efficiency compared to LSTMs while outperforming standard RNNs (Chung et al., 2014), making them well-suited for astronomical time series (Boone, 2019; Gupta et al., 2024; Muthukrishna et al., 2019).

Training uses the Adam optimizer (Kingma & Ba, 2017) with early stopping after 5 epochs without validation improvement. Convergence times on V100 GPUs: 10 minutes (Classifier), 20 minutes (Adversarial), 45 minutes (Contrastive). The final experiments required an estimated 15-25 GPU hours, however ideation and experimentation required considerably more.

C. Qualitative Latent Space Analysis

We visualize the learned representations using UMAP (McInnes et al., 2020) to assess domain alignment quality. Figure 4 shows that standard classifiers maintain distinct ZTF and LSST clusters, indicating survey-specific artifacts in the learned embeddings. In contrast, both adversarial and contrastive training appear to merge these embedding, creating slightly more unified representations where survey identity becomes secondary to astronomical class. Future work should explore this claim quantitatively.

Figure 5 demonstrates this alignment at the class level, showing Type II supernovae from both surveys occupying the same latent region. This class-level unification may validate that our domain-agnostic objectives preserve astronomical semantics while eliminating instrumental biases - precisely the behavior required for effective zero-shot transfer.

Fig. 4 and 5 show a UMAP (McInnes et al., 2020) visualization of the penultimate layer of our neural network classifiers. As seen, contrastive and adversarial models help unify the distributions of ZTF and LSST in the latent space down to the class level.

D. Further Analysis

D.1. Adversarial vs. Contrastive Loss

Incorporating the specialized training techniques proposed in this work does not improve model performance on supervised downstream tasks (Table 1), which is expected since these loss functions are designed for cross-survey alignment rather than task-specific optimization. Adversarial training shows a slight performance decrease compared to standard classifiers. We attribute this to the adversarial netowrk effectively placing a penalty term for domain alignment, whereas the contrastive model receives explicit supervision through class-based positive pairs.

On zero-shot tasks, both adversarial and contrastive models substantially outperform standard classifiers (Table 2), demonstrating that the domain-agnostic embedding enables meaningful cross-survey transfer. We note that training with both contrastive and adversarial losses simultaneously does not improve performance in comparison to a purely contrastive model.

D.2. Performance on Simulations

Fine-tuned models consistently underperform on real data compared to simulations, highlighting the domain gap between synthetic and observational data. For redshift estimation, our best model achieves R^2 score of 0.431 ± 0.053 on real ZTF data versus 0.580 ± 0.011 on simulations. Models can leverage these physics-based simulations as effective starting points but still require labeled real data to perform well. This gap between real data and simulations is why there is a significant performance gap between the Limited and Full evaluation scenarios. In other words, models trained on simulations need to be fine-tuned on real data to work well.

D.3. Anomaly Detection

Anomaly detection is the only task in which we do not see a performance improvement from the Limited to Full settings (Table 1). By definition, anomalies are objects that astronomers find interesting. By using human-defined simulations to pretrain, they are naturally equipped to detect specifically what humans find interesting. The gap between simulations and real data is what anomaly detection pipelines are trying to fill. Further analysis of anomaly detection specifically is out of the scope of this work, however we hope that future researchers analyze the nature of this task and how domain expertise can be incorporated into it.

E. kNN Zero-Shot Estimation

Aside from fine-tuning an MLP, we also evaluate using a k Nearest Neighbors approach for zero-shot estimation (Zhang et al., 2024; Parker et al., 2024). To perform zeroshot redshift estimation for an LSST object, we first find the k = 100 closest ZTF embeddings to the LSST light curve embedding in the latent space. Then, we use the distanceweighted average of the corresponding redshifts to estimate the final redshift of the LSST object. As seen in Table 2, this zero-shot estimation method performs worse than using a directly trained MLP.