In Search of the Unknown Unknowns: A Multi-Metric Distance Ensemble for Out of Distribution Anomaly Detection in Astronomical Surveys

Siddharth Chaini[*]
University of Delaware
chaini@udel.edu

Federica B. Bianco
University of Delaware
Rubin Observatory

Ashish Mahabal
Caltech

Abstract

Distance-based methods involve the computation of distance values between features and are a well-established paradigm in machine learning. In anomaly detection, anomalies are identified by their large distance from normal data points. However, the performance of these methods often hinges on a single, user-selected distance metric (*e.g.*, Euclidean), which may not be optimal for the complex, high-dimensional feature spaces common in astronomy. Here, we introduce a novel anomaly detection method, Distance Multi-Metric Anomaly Detection (DiMMAD), which uses an ensemble of distance metrics to find novelties.

Using multiple distance metrics is effectively equivalent to using different geometries in the feature space. By using a robust ensemble of diverse distance metrics, we overcome the metric-selection problem, creating an anomaly score that is not reliant on any single definition of distance. We demonstrate this multi-metric approach as a tool for simple, interpretable scientific discovery on astronomical time series – (1) with simulated data for the upcoming Vera C. Rubin Observatory Legacy Survey of Space and Time, and (2) real data from the Zwicky Transient Facility.

We find that DiMMAD excels at out-of-distribution anomaly detection – anomalies in the data that might be new classes – and beats other state-of-the-art methods in the goal of maximizing the diversity of new classes discovered. For rare in-distribution anomaly detection, DiMMAD performs similarly to other methods, but may allow for improved interpretability. All our code is open source: DiMMAD is implemented within `DistClassiPy`: https://github.com/sidchaini/distclassipy/, while all code to reproduce the results of this paper is available here: https://github.com/sidchaini/dimmad/.

1 Introduction

The imminent Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST) is poised to revolutionize time-domain astronomy, generating time series alerts capturing changes in brightness for millions of objects every night [1]. While LSST's unique capability – its high resolution, large

---

[*]sidchaini.github.io

depth of view, and high cadence – will enable unprecedented studies of known phenomena, its greatest promise lies in its potential to discover entirely new astrophysical objects, often referred to as the 'unknown unknowns' or 'true novelties' [2]. Historically, these discoveries have often been serendipitous, and have drastically improved our physical understanding of the Universe (*e.g.* pulsars [3], 'Oumuamua [4], *etc.*). However, the sheer scale and volume of LSST data requires us to 'plan for serendipity', where machine learning will be crucial in enabling the discovery of unusual astrophysical phenomena [*e.g.* 5–17].

The task of finding the novel phenomena can be framed as an anomaly detection (AD) problem, and can be seen through two different lens: (1) Out of Distribution (OOD) detection – identifying objects from a very different class than a training set of known classes (*e.g.*, finding supernovae when trained on variable stars), and (2) Rare In-Distribution detection – identifying extreme or rare examples which are more like a different subclass rather than a new class in itself (*e.g.*, finding superluminous supernovae when trained on common supernovae). While both are important to astrophysics research, OOD detection offers the best chance at finding the 'unknown unknowns' [18] to enable the discovery of fundamentally new physics.

The search for these anomalies in time-domain astronomy is typically a two-stage process. The raw, irregularly-sampled time series data (light curves) must be encoded into a fixed-length feature vector. This can be accomplished through feature extraction [*e.g.* 19, 20] or learned via deep learning methods such as representation learning [*e.g.* 8, 21]. Then, an anomaly detection algorithm is applied to this feature space to identify outliers (some common methods are described later in section 3).

A significant challenge, however, is that standard AD algorithms can struggle to prioritize OOD objects [*e.g.* 13]. This can obscure true novelties by saturating the list of top anomaly candidates with statistically rare and interesting, but astrophysically known objects. To address this, we propose a novel, semi-supervised AD method for identifying anomalies from a given feature representation: Distance Multi-Metric Anomaly Detection (DiMMAD).

2 Methods

Our anomaly detection method DiMMAD builds on `DistClassiPy`, a distance-based classifier that uses multiple distance metrics [22]. The authors successfully applied it to the classification of variable stars by extracting variability features from light curves. Their classification is based on the 'distance' of the test object features to the centroids of training object features. With the use of the right distance metric for a classification problem, their classification performance matched the state-of-the-art, while being significantly faster and more interpretable. However, they showed that the performance hinges on choosing the right distance metric for a given classification problem, as the best choice depends on the topology of the feature space. Here, we extend this method to anomaly detection, and use an ensemble of distance metrics to overcome the metric selection problem altogether.

2.1 Distance metrics

A distance metric is a function $d(x, y)$ that quantifies the similarity between two points, obeying three core axioms: identity of indiscernibles, symmetry, and the triangle inequality (refer to Appendix, [22]). While the Euclidean metric is most commonly used, many other functions satisfy these axioms, each imposing a unique geometry on the feature space (shown in Figure 1). While individual distances have been used for AD before ([*e.g.* 23, 24]), DiMMAD uses an ensemble of distance metrics. We chose 16 distance metrics[2] that performed well on light curve classification with `DistClassiPy` [22], but we removed two metrics from the set used in [22] (Additive Symmetric χ^2 & Maryland Bridge) to increase diversity and decrease redundancy, which is crucial for the ensemble to achieve a robust, consensus-based anomaly score capable of handling different feature space geometries.

2.2 Multi-metric Anomaly Score

DiMMAD is a semi-supervised AD method. It consists of two steps – the first step (training) involves calculating the centroids of 'seen classes' (knowns/inliers) in a given labeled feature set. Then, during test time, it looks at an unseen unlabeled dataset and sorts all of them based on an anomaly score –

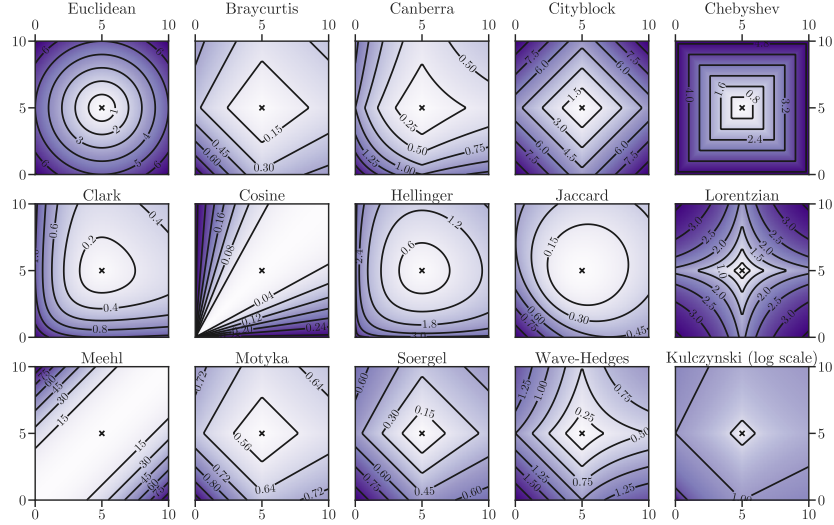[2]Full list of distance metrics is available online: DistClassiPy docs.

Figure 1: A visualization of 15 (of total 16) distance metrics used in our anomaly detector, DiMMAD. In each panel, the intensity of the color denotes the distance value from a central point $(5, 5)$ for that metric in a 2-dimensional feature space. Contours have been plotted for clarity. To aid readability, Kulczynski has been plotted in a log scale. Each distance metric has a distinct geometry, and helps in making DiMMAD more robust. The correlation metric is omitted from this plot, as it cannot be meaningfully represented within a two-dimensional space.

giving us a ranking of the most anomalous to least anomalous objects. This semi-supervised approach allows us to incorporate our knowledge of known classes into the anomaly hunt, unlike some other unsupervised AD approaches (for *e.g.*, iForest, LOF, *etc.*)

Our central hypothesis is that a 'true novelty' will be distant from the centroids of all known classes across a majority of distance metrics. Thus, at test time, we calculate the distances of all test objects to all known classes for all 16 distance metrics. We then calculate the anomaly score in two aggregation stages: (1) For every distance metric, aggregate the distance of the test object to every class centroid by a statistic (class agg: `min/median`) to get a single-metric score for every metric. (2) Then, aggregate these single-metric scores across all distance metrics by another statistic (metric agg: `median`) to get the final multi-metric score. Through this, we get a multi-metric anomaly score for all test objects, and rank them from highest (most anomalous) to lowest (least anomalous).

3 Experiments & Results

We test DiMMAD against a suite of AD algorithms, including the state-of-the-art methods in time-domain astronomy. This set was chosen to represent a variety of approaches common in literature: tree-based methods like the Isolation Forest [iForest, 6, 25, 26], density-based methods like the Local Outlier Factor [LOF, 5, 27], boundary-based methods like the One-Class Support Vector Machines [OC-SVM, 25, 26], as well as modern deep learning approaches like Autoencoders [8, 28], and the Multiclass Deep support vector data description [MCSVDD, 11, 29].

We test our method on the two key discovery tasks described earlier: finding entirely new phenomena (OOD detection) and finding specific rare subtypes (rare in-distribution detection). A key challenge at the start of a survey like LSST is that labeled data will be scarce. Our experimental setup aims to mimic this 'data-poor' scenario: we train only on a small number of well-understood object classes. Evaluating performance for the discovery of astrophysical unknowns is non-trivial. Metrics used to assess the power of classifiers are not suited to assess the performance of an AD method. Standard metrics like the AUROC [30] evaluate performance across all thresholds, and can be misleading for highly imbalanced datasets. An AUROC score would contaminate the result with the distribution of values at low anomaly scores, which is by definition not an interesting region of the space where to find anomalies. Instead, we evaluate performance using a metric directly tied to the scientific goal: purity *vs.* follow-up budget. Purity (also known as precision) measures the fraction of 'true' out-of-class anomalies discovered within the top N ranked candidates, reflecting the reality of limited

telescope resources in order to prioritize true novelties for follow-up. Our approach focuses on the crucial high-confidence regime relevant for effective astronomical follow-up [*e.g.* 13].

For our experiments, we use two datasets of astronomical light curves. First, we use the Extended LSST Astronomical Time-series Classification Challenge [ELAsTiCC, 31], a simulated dataset designed to mimic the six-band ($ugrizy$) data from the upcoming Vera C. Rubin Observatory. Second, we use real data from the Zwicky Transient Facility [32], a precursor survey to LSST that provides public data in two optical bands (gr).

We run our models on features derived from the Supernova Parametric Model [33] fit to light curves with the ALeRCE pipeline software [34]. These features (*e.g.*, amplitude, rise/fall times) are easy and computationally inexpensive to calculate, and crucially, have a low failure rate ($< 5\%$), making them robustly available for nearly all objects as compared to the other features which can often have much higher failure rates ($\sim 45\%$), making it a pragmatic choice for real-time anomaly searches with surveys like ZTF and LSST. For the ELAsTiCC data, we use 25 features derived from model fits across the $grizy$ bands, and for the ZTF data, we use 14 features from the g and r bands. Details of this are available on our code repository.

For each run, we first create a balanced training set by downsampling all 'known' inlier classes to the number of samples in the smallest class. The test set for each run is constructed from unseen inliers and a diverse, stratified sample of 'unknown' outlier objects. While not astrophysically representative, the test set's high anomaly fraction ($\sim 47\%$, spread equally over many anomalous unknown classes) is a deliberate choice to create a robust benchmark, ensuring a statistically stable comparison of each algorithm's ranking performance. However, we also tested on a test dataset composed of $> 98\%$ inliers to confirm the robustness of the results we present here (described in Appendix A). All models are evaluated on the exact same train/test splits. We perform Monte Carlo cross-validation [35], and run each experiment 20 times with different random seeds to ensure statistical robustness.

3.1 Out-of-Distribution (Inter-Class) Anomaly Detection

We first test DiMMAD at finding 'unknown unknowns'. We simulate this by training on well-understood classes of objects (knowns) and treating physically distinct classes as unseen, OOD anomalies (unknowns).

Simulated LSST Data (ELAsTiCC) To create a controlled environment, we train the models on four classes of variable stars (cepheids, RR lyrae, delta scutis and eclipsing binaries) and task them with identifying 34 unseen transient classes (supernovae, TDEs, *etc.*) as anomalies. The results for this are shown in Figure 2 for two variants of DiMMAD (cluster aggregators (`min`/`median`) with `median` as the metric agg. The purity plot (Figure 2 left panel) shows that both DiMMAD methods are the best performing in the lower budget regime, and consistently maintain a purity above 60% within a realistic follow-up budget of the first few hundred objects. Furthermore, Figure 2 right panel shows that DiMMAD identifies a diverse range of new phenomena more rapidly than any other method, maximizing the potential for broad scientific discovery.

Real ZTF Data (ALeRCE) We validate these findings on real ZTF data, from the light curve feature set calculated by ALeRCE [36]. We train on 11 common variable and stochastic classes and treat four supernova classes as the unseen OOD sample. We see that one variant of DiMMAD (`min` cluster agg, `median` metric agg) is again the best performing. The other DiMMAD variant (`med-med`) performs comparably to other methods like iForest and OC-SVM (Figure 3).

3.2 Rare In-Distribution (Intra-Class) Anomaly Detection

We also tested a more challenging scenario: finding rare subtypes within a broader family of objects. For this task, we trained on 16 known supernova subtypes and defined 3 rare pair-instability supernova models as the 'unknown' class. While DiMMAD remains competitive, we find that iForest often shows superior performance. This suggests that while our multi-metric consensus excels at finding OOD objects, other approaches may be better suited for finding rare examples within a known population. Our hypothesis for this difference is as follows - the iForest finds isolated points located in the sparse regions of the manifold, while DiMMAD's distance based approach explicitly finds objects that are robustly distant from all known centroids across multiple geometries, making it
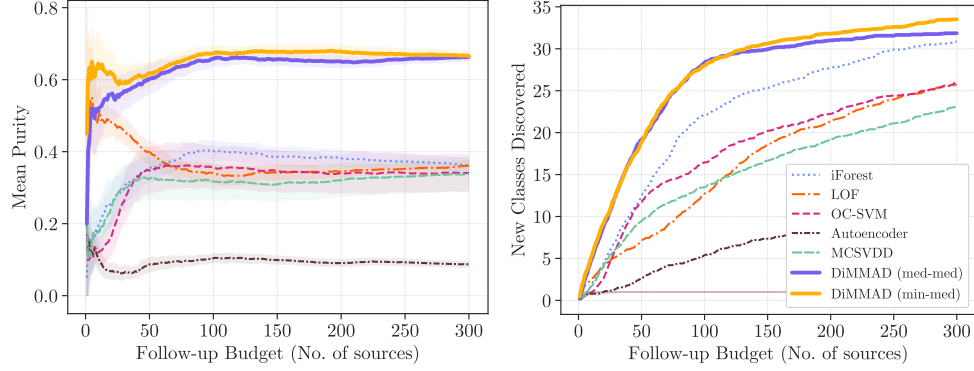
Figure 2: Anomaly Detection Performance of our method (DiMMAD), compared to other standard AD methods. The shaded region denotes 1σ errors from Monte Carlo cross-validation. *Left*: the mean purity of OOD anomalies within the top N candidates on the ELAsTiCC data. The DiMMAD methods (solid lines) maintain significantly higher purity. *Right*: the cumulative number of new, unique OOD classes discovered on the ELAsTiCC data. DiMMAD discovers a greater, more diverse set of anomalies more efficiently.
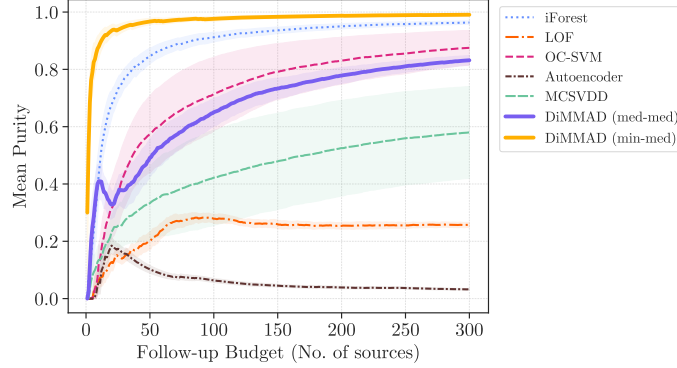

Figure 3: As Figure 2 (*left*) but for the case of real ZTF (ALeRCE) data.

exceptionally effective for off-manifold (OOD) detection but less sensitive to subtle, on-manifold variations. More details for this experiment are provided in Appendix B.

4   Conclusion

We introduced DiMMAD, a novel, semi-supervised anomaly detection method that leverages an ensemble of distance metrics to identify novel astronomical phenomena. Our work demonstrates that DiMMAD is a powerful tool for Out-of-Distribution detection, a crucial task for maximizing serendipitous discoveries in large surveys like LSST. The strength of our approach lies in its demand for a geometric consensus. While standard algorithms may flag extreme in-sample objects located in sparse regions of the feature space, DiMMAD requires a candidate to be robustly distant from all known classes from the perspective of multiple, diverse geometries. This makes it particularly effective at identifying truly novel, off-manifold objects. A key advantage of our semi-supervised approach is its ability to dynamically improve. Once a new class of anomaly is discovered and characterized through follow-up, its centroid can be added to the set of 'knowns'.

For the Rubin LSST community, DiMMAD offers an interpretable and computationally efficient method to filter through vast alert streams, enriching the top of ranked candidate lists with true novelties and optimizing follow-up resources. Our implementation is open-source, and available within the `DistClassiPy` package,[3] along with all code to reproduce the results of this paper.[4]

---

[3]https://github.com/sidchaini/distclassipy/

[4]https://github.com/sidchaini/dimmad/

Acknowledgments and Disclosure of Funding

References

[1] Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, LSST: From Science Drivers to Reference Design and Anticipated Data Products, The Astrophysical Journal, 873, 111, doi: 10.3847/1538-4357/ab042c

[2] Li, X., Ragosta, F., Clarkson, W. I., & Bianco, F. B. 2021, Preparing to discover the unknown with rubin lsst: Time domain, The Astrophysical Journal Supplement Series, 258, 2

[3] Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., & Collins, R. A. 1968, Observation of a Rapidly Pulsating Radio Source, Nature, 217, 709, doi: 10.1038/217709a0

[4] Meech, K. J., Weryk, R., Micheli, M., et al. 2017, A brief visit from a red and extremely elongated interstellar asteroid, Nature, 552, 378, doi: 10.1038/nature25020

[5] Lochner, M., & Bassett, B. 2021, Astronomaly: Personalised active anomaly detection in astronomical data, Astronomy and Computing, 36, 100481, doi: 10.1016/j.ascom.2021.100481

[6] Malanchev, K. L., Pruzhinskaya, M. V., Korolev, V. S., et al. 2021, Anomaly detection in the Zwicky Transient Facility DR3, Monthly Notices of the Royal Astronomical Society, 502, 5147, doi: 10.1093/mnras/stab316

[7] Martínez-Galarza, J. R., Bianco, F. B., Crake, D., et al. 2021, A method for finding anomalous astronomical light curves and their analogues, Monthly Notices of the Royal Astronomical Society, 508, 5734, doi: 10.1093/mnras/stab2588

[8] Villar, V. A., Cranmer, M., Berger, E., et al. 2021, A Deep-learning Approach for Live Anomaly Detection of Extragalactic Transients, The Astrophysical Journal Supplement Series, 255, 24, doi: 10.3847/1538-4365/ac0893

[9] Muthukrishna, D., Mandel, K. S., Lochner, M., Webb, S., & Narayan, G. 2022, Real-time detection of anomalies in large-scale transient surveys, Monthly Notices of the Royal Astronomical Society, 517, 393, doi: 10.1093/mnras/stac2582

[10] Volnova, A. A., Aleo, P. D., Lavrukhina, A., et al. 2024, Exploring the Universe with SNAD: Anomaly Detection in Astronomy, in Data Analytics and Management in Data Intensive Domains, ed. J. Baixeries, D. I. Ignatov, S. O. Kuznetsov, & S. Stupnikov (Cham: Springer Nature Switzerland), 195–208

[11] Pérez-Carrasco, M., Cabrera-Vives, G., Hernández-García, L., et al. 2023, Multi-Class Deep SVDD: Anomaly Detection Approach in Astronomy with Distinct Inlier Categories, arXiv, doi: 10.48550/arXiv.2308.05011

[12] Gagliano, A., & Villar, V. A. 2023, A Physics-Informed Variational Autoencoder for Rapid Galaxy Inference and Anomaly Detection, arXiv, doi: 10.48550/arXiv.2312.16687

[13] Aleo, P. D., Engel, A. W., Narayan, G., et al. 2024, Anomaly Detection and Approximate Similarity Searches of Transients in Real-time Data Streams, The Astrophysical Journal, 974, 172, doi: 10.3847/1538-4357/ad6869

[14] Gupta, R., Muthukrishna, D., & Lochner, M. 2025, A Classifier-Based Approach to Multi-Class Anomaly Detection for Astronomical Transients, RAS Techniques and Instruments, 4, doi: 10.1093/rasti/rzae054

[15] Kornilov, M., Korolev, V., Malanchev, K., et al. 2025, Coniferest: A complete active anomaly detection framework, Astronomy and Computing, 52, 100960, doi: 10.1016/j.ascom.2025.100960

[16] Romao, M. C., Croon, D., & Godines, D. 2025, Anomaly Detection to identify Transients in LSST Time Series Data, arXiv, doi: 10.48550/arXiv.2503.09699

[17] Pruzhinskaya, M. V., Lavrukhina, A. D., Semenikhi, T. A., et al. 2025, What ZTF Saw Where Rubin Looked: Anomaly Hunting in DR23, arXiv, doi: 10.48550/arXiv.2507.06217

[18] Yang, J., Zhou, K., Li, Y., & Liu, Z. 2024, Generalized Out-of-Distribution Detection: A Survey, arXiv, doi: 10.48550/arXiv.2110.11334

[19] Sánchez-Sáez, P., Reyes, I., Valenzuela, C., et al. 2021, Alert Classification for the ALeRCE Broker System: The Light Curve Classifier, The Astronomical Journal, 161, 141, doi: 10.3847/1538-3881/abd5c1

[20] Malanchev, K. 2021, light-curve: Light curve analysis toolbox, Astrophysics Source Code Library, ascl:2107.001. https://ui.adsabs.harvard.edu/abs/2021ascl.soft07001M

[21] Parker, L., Lanusse, F., Golkar, S., et al. 2024, AstroCLIP: a cross-modal foundation model for galaxies, Monthly Notices of the Royal Astronomical Society, 531, 4990, doi: 10.1093/mnras/stae1450

[22] Chaini, S., Mahabal, A., Kembhavi, A., & Bianco, F. B. 2024, Light curve classification with DistClassiPy: A new distance-based classifier, Astronomy and Computing, 48, 100850, doi: 10.1016/j.ascom.2024.100850

[23] Churová, V., Vyškovský, R., Maršálová, K., Kudláček, D., & Schwarz, D. 2021, Anomaly Detection Algorithm for Real-World Data and Evidence in Clinical Research: Implementation, Evaluation, and Validation Study, JMIR Medical Informatics, 9, e27172, doi: 10.2196/27172

[24] Souto Arias, L. A., Oosterlee, C. W., & Cirillo, P. 2023, AIDA: Analytic isolation and distance-based anomaly detection algorithm, Pattern Recognition, 141, 109607, doi: 10.1016/j.patcog.2023.109607

[25] Liu, F. T., Ting, K. M., & Zhou, Z.-H. 2008, Isolation Forest, in 2008 Eighth IEEE International Conference on Data Mining (Pisa, Italy: IEEE), 413–422, doi: 10.1109/ICDM.2008.17

[26] Liu, F. T., Ting, K. M., & Zhou, Z.-H. 2012, Isolation-Based Anomaly Detection, ACM Transactions on Knowledge Discovery from Data, 6, 1, doi: 10.1145/2133360.2133363

[27] Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. 2000, LOF: identifying density-based local outliers, ACM SIGMOD Record, 29, 93, doi: 10.1145/335191.335388

[28] Rumelhart, D. E., McClelland, J. L., & AU. 1986, Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations (The MIT Press), doi: 10.7551/mitpress/5236.001.0001

[29] Ruff, L., Vandermeulen, R., Goernitz, N., et al. 2018, Deep One-Class Classification, in Proceedings of Machine Learning Research, Vol. 80, Proceedings of the 35th International Conference on Machine Learning, ed. J. Dy & A. Krause (PMLR), 4393–4402. https://proceedings.mlr.press/v80/ruff18a.html

[30] McClish, D. K. 1989, Analyzing a Portion of the ROC Curve, Medical Decision Making, 9, 190, doi: 10.1177/0272989X8900900307

[31] Narayan, G., & ELAsTiCC Team. 2023, The Extended LSST Astronomical Time-series Classification Challenge (ELAsTiCC), in American Astronomical Society Meeting Abstracts, Vol. 241, American Astronomical Society Meeting Abstracts #241, 117.01

[32] Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, The Zwicky Transient Facility: System Overview, Performance, and First Results, Publications of the Astronomical Society of the Pacific, 131, 018002, doi: 10.1088/1538-3873/aaecbe

[33] Villar, V. A., Berger, E., Miller, G., et al. 2019, Supernova Photometric Classification Pipelines Trained on Spectroscopically Classified Supernovae from the Pan-STARRS1 Medium-deep Survey, The Astrophysical Journal, 884, 83, doi: 10.3847/1538-4357/ab418c

[34] Jainaga, I. R., Arredondo, J., Valenzuela, C., et al. 2021, Alercebroker/Lc_classifier: Release 1.2.3-P,, Zenodo doi: 10.5281/ZENODO.5275453

[35] Xu, Q.-S., & Liang, Y.-Z. 2001, Monte Carlo cross validation, Chemometrics and Intelligent Laboratory Systems, 56, 1, doi: 10.1016/S0169-7439(00)00122-2

[36] Perez-Carrasco, M., Cabrera-Vives, G., Hernandez-García, L., et al. 2023, Alert Classification for the ALeRCE Broker System: The Anomaly Detector, The Astronomical Journal, 166, 151, doi: 10.3847/1538-3881/ace0c1

A Performance on Individual Out-of-Distribution Classes

To understand how DiMMAD performs on specific types of novelties, we repeated the OOD experiment from section 3.1 on ELAsTiCC with the same 'known' set of 4 variable stars (CEP, RR, DSCT, EB), but restricted the test set to contain only one 'unknown' transient class at a time, comprising <2% of the total test set. This tests the ability of each algorithm to recover a single, rare type of OOD signal.

Figure 4 shows the mean cumulative number of objects discovered for a selection of physically diverse and rare anomalous classes over 20 runs of Monte Carlo cross validation. The discovery rate is much lower for all models, but the DiMMAD methods consistently rank among the top performers, demonstrating particular strength in identifying rare explosive transients like Kilonovae (KN), Superluminous Supernovae (SLSN) and Calcium Rich Transients (CART) as well as other rare events like Strongly Lensed Type Ia Supernovae (SL-SN1a) and Microlensing events (uLens). For some classes, other methods like MCSVDD, iForest and LOF perform as well (Pair-Instability Supernovae; PISN) or better (Intermediate Luminous Optical Transients; ILOT), highlighting that no single algorithm is universally optimal. However, across the full spectrum of anomalies, the DiMMAD ensemble approach provides a consistently high rate of discovery, reinforcing its utility as an efficient, interpretable general-purpose method for finding the unknown.

B DiMMAD Performance on Rare In-Distribution Anomalies

To test DiMMAD's performance on subtle, intra-class anomalies in ELAsTiCC, we trained our models on 16 known classes of supernova models (for Supernova Types Ia, Iax, Ib, Ic, II, IIb, IIn, and SLSN-I), and tried finding 3 predicted classes of rare pair-instability supernova models (PISN) as the 'unknown' class. The results (Figure 5) show that while DiMMAD is competitive, iForest provides a higher and more stable purity for this on-manifold detection task. This highlights that the optimal anomaly detection method is task-dependent. While our multi-metric ensemble excels at finding 'unknown unknowns' (OOD), other approaches may be better suited for finding the 'known unknowns' (rare in-distribution).
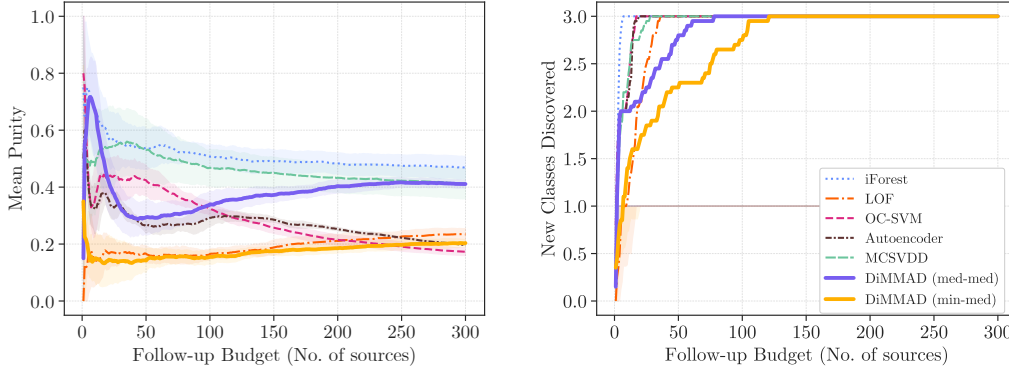

Figure 5: Anomaly Detection Performance of our method (DiMMAD), compared to other standard AD methods on the ELAsTiCC data as in Figure 2, but for Rare In-Distribution Anomalies. The shaded region denotes the 1σ errors from Monte Carlo cross-validation.
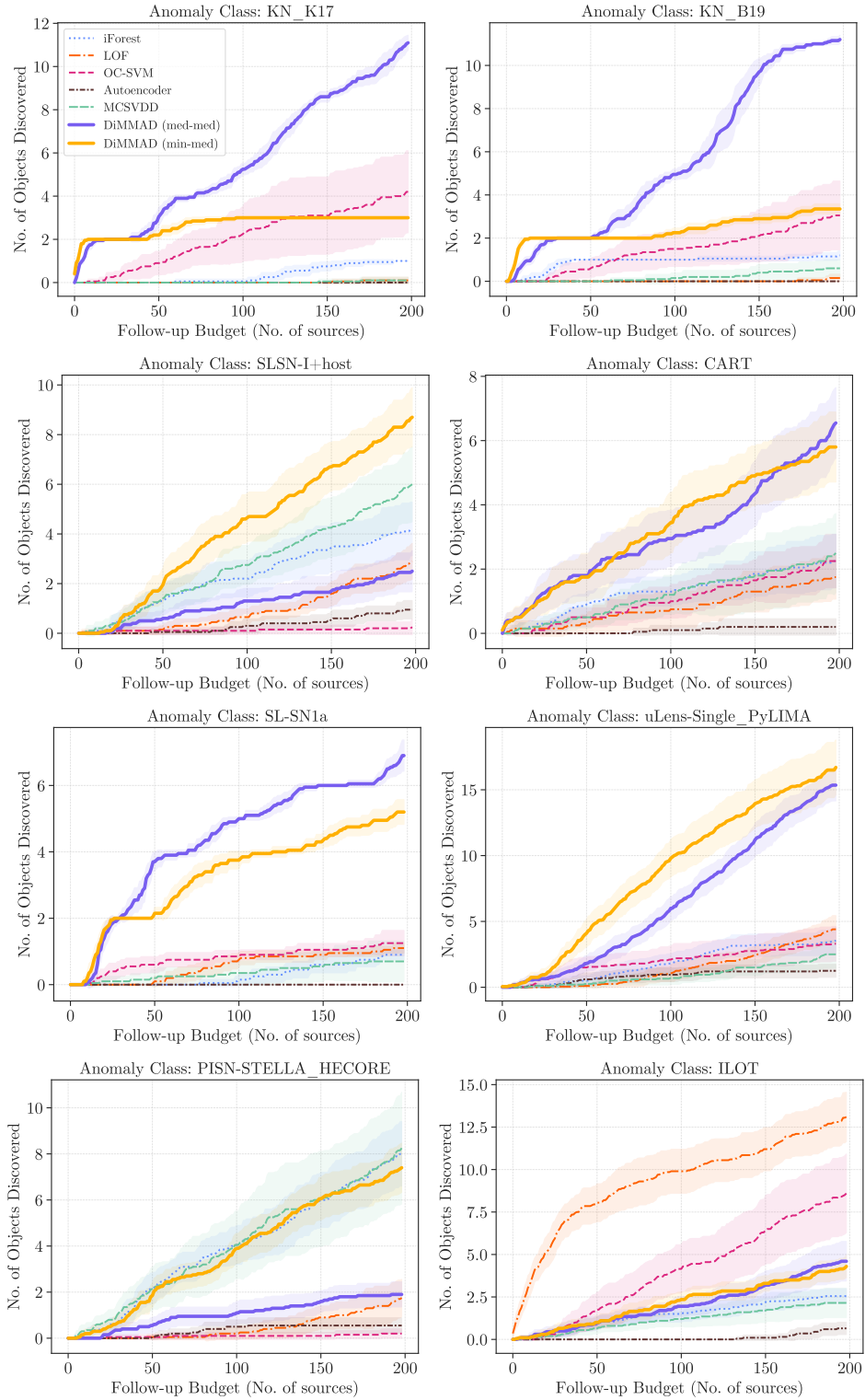
Figure 4: Cumulative number of objects discovered vs. follow-up budget for individual OOD classes from the ELAsTiCC dataset. Each panel represents a separate experiment where the titled class was the sole anomaly type. DiMMAD (bold orange/purple) consistently performs among the best methods for most classes of OOD anomalies across a diverse range of physical phenomena