



Alert Classification for the ALerCE Broker System: The Anomaly Detector

Manuel Perez-Carrasco^{1,2}, Guillermo Cabrera-Vives^{1,2,3}, Lorena Hernandez-García^{2,4}, F. Förster^{2,5,6,7}, Paula Sanchez-Saez^{2,8}, Alejandra M. Muñoz Arancibia^{2,6}, Javier Arredondo^{2,6}, Nicolás Astorga⁶, Franz E. Bauer^{2,9,10}, Amelia Bayo^{4,8,11}, M. Catelan^{2,9,10}, Raya Dastidar^{2,12}, P. A. Estévez^{2,13}, Paulina Lira^{14,15}, and Giuliano Pignata^{2,12}

¹Data Science Unit, Universidad de Concepción, Edmundo Larenas 310, Concepción, Chile; maperezc@inf.udec.cl

²Millennium Institute of Astrophysics (MAS), Nuncio Monseñor Sótero Sanz 100, Providencia, Santiago, Chile

³Department of Computer Science, Universidad de Concepción, Edmundo Larenas 219, Concepcion, Chile

⁴Instituto de Física y Astronomía, Facultad de Ciencias, Universidad de Valparaíso, Av. Gran Bretaña 1111, Playa Ancha, Valparaíso, Chile

⁵Data and Artificial Intelligence Initiative (ID&IA), Universidad de Chile, Chile

⁶Center for Mathematical Modeling, Universidad de Chile, Beauchef 851, North Bldg., 7th Flr., Santiago 8320000, Chile

⁷Departamento de Astronomía, Universidad de Chile, Casilla 36D, Santiago, Chile

⁸European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany

⁹Instituto de Astrofísica, Facultad de Física, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile

¹⁰Centro de Astroingeniería, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile

¹¹Núcleo Milenio de Formación Planetaria (NPF), Chile

¹²Departamento de Ciencias Físicas, Universidad Andres Bello, Av. Republica 252, Santiago, Chile

¹³Department of Electrical Engineering, Universidad de Chile, Av. Tupper 2007, Santiago 8320000, Chile

¹⁴Departamento de Astronomía, Universidad de Chile, Camino el Observatorio 1515, Santiago, Chile

¹⁵Millennium Nucleus on Transversal Research and Technology to Explore Supermassive Black Holes (TITANS), Chile

Received 2022 June 13; revised 2023 June 20; accepted 2023 June 20; published 2023 September 7

Abstract

Astronomical broker systems, such as Automatic Learning for the Rapid Classification of Events (ALerCE), are currently analyzing hundreds of thousands of alerts per night, opening up an opportunity to automatically detect anomalous unknown sources. In this work, we present the ALerCE anomaly detector, composed of three outlier detection algorithms that aim to find transient, periodic, and stochastic anomalous sources within the Zwicky Transient Facility data stream. Our experimental framework consists of cross-validating six anomaly detection algorithms for each of these three classes using the ALerCE light-curve features. Following the ALerCE taxonomy, we consider four transient subclasses, five stochastic subclasses, and six periodic subclasses. We evaluate each algorithm by considering each subclass as the anomaly class. For transient and periodic sources the best performance is obtained by a modified version of the deep support vector data description neural network, while for stochastic sources the best results are obtained by calculating the reconstruction error of an autoencoder neural network. Including a visual inspection step for the 10 most promising candidates for each of the 15 ALerCE subclasses, we detect 31 bogus candidates (i.e., those with photometry or processing issues) and seven potential astrophysical outliers that require follow-up observations for further analysis.¹⁶

Unified Astronomy Thesaurus concepts: [Astronomy data analysis \(1858\)](#); [Surveys \(1671\)](#); [Interdisciplinary astronomy \(804\)](#)

1. Introduction

Modern survey telescopes are producing unprecedented volumes of data that make analysis by human inspection unfeasible. Therefore, automated pipelines are needed to produce knowledge in a data-driven fashion. One interesting and challenging task is anomaly detection, which refers to finding abnormal or unexpected patterns that do not conform to our knowledge about the data (Chandola et al. 2009).

A broad array of methods have been developed to find outlier events, focusing on specific scientific objectives and/or specific data sets. Examples of anomaly detection in the literature are listed in the following: Xiong et al. (2010) used a hierarchical probabilistic model, while Baron & Poznanski (2016) used an unsupervised random forest to find outliers

among the galaxy spectra from the Sloan Digital Sky Survey. More recently Sánchez-Sáez et al. (2021a) used a variational recurrent autoencoder (VRAE) architecture on the active galactic nucleus (AGN) light curves in Zwicky Transient Facility (ZTF) Data Release 5. Using the light curves in the Massive Compact Halo Object (MACHO) catalog, Nun et al. (2014) used a supervised random forest, whereas using the light curves of periodic variable stars, Twomey et al. (2019) applied a hierarchical Gaussian process to the Optical Gravitational Lensing Experiment (OGLE) data set. Tsang & Schultz (2019) developed a recurrent neural network autoencoder (AE) with a Gaussian mixture model in the latent space to detect outliers in the All Sky Automated Survey for Supernovae (ASAS-SN) variable star database. Pruzhinskaya et al. (2019) used the photometric data of the Open Supernova Catalog (OSC) with an isolation forest (IForest) algorithm. Villar et al. (2020) also applied an IForest algorithm in the latent space of a VRAE to a simulated data set of supernovae (SNe), and Ishida et al. (2021) applied an IForest algorithm to features of the OSC data set and the Photometric LSST Astronomical Time-series Classification Challenge (PLATiCC). Reyes & Estévez (2020) used images from the High Cadence Transient Survey (HiTS) and ZTF data

¹⁶ The code and the data needed to reproduce our results are publicly available at <https://github.com/mperezcarrasco/AnomalyALerCE>.

sets and applied the geometrical transformation method named Geotransform (Golan & El-Yaniv 2018) to find anomalies in these massive data sets. Ultimately, Muthukrishna et al. (2019) detected anomalies using a probabilistic neural network approach built upon Temporal Convolutional Networks.

The detection of anomalous astronomical sources in a systematic fashion is crucial to discovering new astronomical phenomena in massive data sets. In particular, the development of algorithms for fast detection and classification of events is critical for short-lived phenomena, for the early phases of the evolution of longer-lived processes, and for the follow-up of events that require additional observations to uncover their true nature. Some potential anomalies to be detected include new families of explosive events, including neutron star–neutron star mergers (kilonovae; Abbott et al. 2016), neutron star–black hole mergers, optical counterparts of high-energy neutrino events, SN events that include significant interaction with a companion star, and pair instability SNe, among others (IceCube Collaboration et al. 2018; Abbott et al. 2020; Graham et al. 2020); new families of stochastic objects, such as so-called changing-state AGNs (LaMassa et al. 2015; MacLeod et al. 2019), extremely variable AGNs (Graham et al. 2017), and the new family of transient events detected in narrow-line Seyfert 1 galaxies (Frederick et al. 2021); and new families of periodic objects, such as the recently discovered BLAPs (blue large-amplitude pulsators; Pietrukowicz et al. 2017; Kupfer et al. 2019), which have been suggested to be the elusive surviving companions of Type Ia SNe (Meng et al. 2020).

The challenge of fast detection and classification of events is being addressed by a new generation of astronomical alert brokers that read, annotate, classify, and redistribute data from large survey telescopes in real time. Several brokers were selected as community brokers for the Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST; Ivezić et al. 2019). The LSST is expected to revolutionize the time domain in astronomy. Among these brokers are Automatic Learning for the Rapid Classification of Events¹⁷ (ALeRCE; Förster et al. 2021), Alert Management Photometry and Evaluation of Lightcurves (AMPEL; Nordin et al. 2019), the Arizona–NOIRLab Temporal Analysis and Response to Events System (ANTARES; Narayan et al. 2018), Babamul, Fink (Möller et al. 2021), Lasair (Smith et al. 2019), and Pitt-Google.¹⁸

In this work, we present a systematic study of methods that led to the online anomaly detector implemented in the ALeRCE broker. ALeRCE is currently processing alerts from ZTF (Bellm et al. 2018), in preparation for the LSST (Ivezić et al. 2019). ALeRCE uses two real-time classifiers, a *stamp classifier* (Carrasco-Davis et al. 2021) and a *light-curve classifier* (Sánchez-Sáez et al. 2021b), and has implemented outlier detection methods for AGN light curves (Sánchez-Sáez et al. 2021a) in an offline manner.

Our proposed methodology for anomaly detection builds on a hierarchical principle similar to that adopted in ALeRCE’s light-curve classifier. At the top level, we divide the light curves into three main classes: transient, stochastic, and periodic. For each class, we build a different anomaly detection model that uses only information about the known objects (i.e., inliers) for training. At test time, in order to assign the light curve to one of the anomaly detectors and compute the anomaly

score, we use the probabilities, as given by ALeRCE’s light-curve classifier (Sánchez-Sáez et al. 2021b), that the light curve corresponds to an object of transient, stochastic, or periodic nature. By selecting the top 10 light curves with the highest anomaly score in each of the ALeRCE subclasses, we are able to find seven potential outlier events and 31 bogus candidates explained by errors in the data, such as wrong period estimations for periodic data or SNe that appear in the template images.

This paper is structured as follows: In Section 2 we describe the data used in this work and the procedure for building the light curves, as well as the taxonomy, the labeled training set, and the features used to perform anomaly detection. In Section 3 we describe the methodology. In Section 4 we describe several algorithms to test the anomaly detector. In Section 5 we describe the metrics used to compare the performance of the different models, and report some interesting outlier candidates. Finally, in Section 6 we draw our conclusions, and discuss challenges for future work.

2. Data

2.1. Input Data

ALeRCE has been processing alerts from the ZTF data stream since 2019 May. In this data stream, an alert is triggered by an object in the sky whose current (science) image has a significant difference with respect to a template (reference) image (Maschi et al. 2018). An alert is a *data packet* in the form of an Avro file¹⁹ that contains image cutouts, features, and metadata for an alert event²⁰ (Maschi et al. 2018). For alerts to be streamed by ZTF, they need to pass the cutoff criteria defined by the real/bogus detection system designed by the ZTF Collaboration. These criteria include signal-to-noise ratios, near-edge image positioning, negative and bad pixels, and morphological and photometric features (Duev et al. 2019; Mahabal et al. 2019).

ALeRCE uses the information contained in the Avro files to construct the light curve of every object, as described in Section 4.4 and Appendix A of Förster et al. (2021). Similar to the ALeRCE light-curve classifier (Sánchez-Sáez et al. 2021b), we perform a crossmatch with the AllWISE²¹ public source catalog (Wright et al. 2010; Mainzer et al. 2011), using a matching radius of 2″, obtaining W1, W2, and W3 photometry. Then, 152 features are calculated for every object with at least six detections in either the *g* or *r* band as defined in Sánchez-Sáez et al. (2021b).²² We use data up to 2021 July 19.

2.2. Taxonomy and Data Filtering

The main idea behind our anomaly detection algorithm is to learn the feature distribution of known objects, in order to look for anomalous sources that deviate from such distribution. To characterize known objects, we adopt the ALeRCE light-curve classifier taxonomy (Sánchez-Sáez et al. 2021b). We start by dividing the sources into three main classes (hereafter, the top-level classes): transient, stochastic, and periodic. Each of these categories is then subdivided into the following subclasses:

¹⁹ The documentation of the ZTF Avro alert files can be found at <https://zwickytransientfacility.github.io/ztf-avro-alert/>.

²⁰ A complete schema of the metadata included in the alert packet can be found at <https://zwickytransientfacility.github.io/ztf-avro-alert/schema.html>.

²¹ The AllWISE Data Release can be found at <http://wise2.ipac.caltech.edu/docs/release/allwise/>.

²² <http://alerce.science/features/>

¹⁷ <http://alerce.science>

¹⁸ <https://pitt-broker.readthedocs.io/en/latest/>

1. Transient: Type Ia SN (SNIa), Type Ibc SN (SNIbc), Type II SN (S II), and superluminous SN (SLSN)
2. Stochastic: Type 1 Seyfert galaxy (AGN; i.e., host-dominated AGN), Type 1 quasar (QSO; i.e., nucleus-dominated AGN), blazar (Blazar; i.e., beamed jet dominated AGN), young stellar object (YSO; including bursters, dippers, and purely rotation-modulated light curves of pre-main-sequence stars), and cataclysmic variable/nova (CV/Nova)
3. Periodic: long-period variable (LPV; includes regular, semi-regular, and irregular variable stars), RR Lyrae (RRL), Cepheid (CEP), eclipsing binary (E), δ Scuti (DSCT), and other periodic variable stars (Periodic-Other; this includes classes of variable stars that are not well represented in the labeled set, e.g., sources classified as miscellaneous and RS CVn stars)

Given that the Periodic-Other subclass included in the taxonomy presented in Sánchez-Sáez et al. (2021b) serves as a catchall category for underrepresented subclasses, it is not used for training, but is included for testing purposes in this work.

In Section 4 we explain in detail how the abovementioned categories are used to evaluate our anomaly detection algorithms by hiding the light curves from a given subclass and considering all others as inliers. This allows for a rigorous comparison of algorithms that helps us to select the most promising method to apply in a real-world scenario.

In order to avoid spurious data, we define several selection criteria. First, we remove from our data set all light curves with fewer than six detections in one of the r and g bands. Also, depending on the top-level class, we remove from our data set any transient sources that meet either of the following two criteria: (1) there are two or more reference images associated with the source in a specific band or (2) there is an SN in the reference image (detected as a negative difference between the science and reference images).²³ We also eliminate stochastic and periodic sources for which the apparent magnitude is not computed, or computed in only one band due to our not finding a source in the template within a radius of $1.4''$, as discussed in Förster et al. (2021). It is important to mention that for the labeled data set, the top-level class used to apply the criterion is given by the labels, while for the unlabeled data set, the top-level class is given by the ALerCE light-curve classifier predictions.

3. Anomaly Detection Algorithms

Anomaly detection is the task of finding outlier sources that deviate from the distribution of a specific data set (inliers; Edgeworth 1887; Chandola et al. 2009). Semi-supervised anomaly detection considers a training set made up of known inliers only. In practice, the test set contains both inliers and outliers and the method is expected to be able to discriminate between them.

In this work, six anomaly detection algorithms are examined in order to compare their performance in finding outliers. These algorithms are explained below:

IForest. An isolation tree is a model in which features are randomly selected to be divided into distinct nonoverlapping regions, based on a randomly selected threshold criterion, and where the output is an anomaly score. Anomaly scores are

proportional to the number of splits required to isolate each object in the sample. Intuitively, anomalous objects should require fewer splits to be isolated. As each feature may have many different randomly selected threshold criteria (even infinite, in the continuous feature case), this method is prone to overfitting. IForest (Liu et al. 2012) is a method based on ensembles of isolation trees for anomaly detection. IForest models have been demonstrated to avoid overfitting, by averaging the anomalous scores of samples within the different isolation trees in an IForest.

We use the `IsolationForest` implementation provided by `scikit-learn` (Pedregosa et al. 2011). The hyperparameters are set as follows: number of trees $t = 100$; number of samples to draw from the data to train each base estimator $\psi = 256$; and contamination parameter $c = 0.1$, as recommended in the original work of Liu et al. (2012). Notice that as outliers are not used for training, we do not select the hyperparameters via cross-validation.

One-class support vector machine.

One-class support vector machine (OCSVM; Schölkopf et al. 1999) is an anomaly detection method based on support vector machines (Cortes & Vapnik 1995). The OCSVM method maps the data into a new feature space such that the inner product between two objects can be represented with a kernel (e.g., a Gaussian kernel). A hyperplane is then learned in the new feature space such that it delimits the region where most of the data lie. At test time, the anomaly score is given by evaluating the distance of the data points with respect to the hyperplane. The anomaly score is assigned depending on which side of the hyperplane each data point falls on.

We use the `OneClassSVM` implementation provided by `scikit-learn` with the *radial basis function* kernel, the hyperparameter $\nu = 0.01$, and the contamination parameter $c = 0.1$. Note that the hyperparameters are selected as default, since finding them would require a sample of anomalous objects, while we assume anomalous samples to be unknown.

AE.

AEs (Rumelhart & McClelland 1987) are unsupervised neural network-based algorithms aiming at generating a reconstruction of the input using a lower-dimensional representation of the data called latent space. AEs are composed of an encoder function $E(\cdot)$, which maps the input data x into the lower-dimensional version $z = E(x; \theta)$ (where θ represents the parameters of the encoder), and a decoder function $D(\cdot)$, which takes the lower-dimensional representation z and reconstructs the original data $\hat{x} = D(z; \phi)$, where ϕ denotes the parameters of the decoder.

To encourage the reconstruction $x = \hat{x}$, the mean squared error is used as the loss function. AEs have been proven effective in anomaly detection (Sakurada & Yairi 2014; Chen et al. 2017; Zhou & Paffenroth 2017). Intuitively, these methods assume that anomalies are incompressible, and therefore they cannot be effectively reconstructed from low-dimensional projections. Thus, it is possible to use the reconstruction error (computed as the mean squared error) as the anomaly score.

We implement an AE using `PyTorch` 1.0.0, and we select the hyperparameters considering the reconstruction error over a validation set composed only of inliers.

Variational AE.

Variational AEs (VAEs; Kingma & Welling 2014) are deep generative models. Similar to an AE, a VAE uses an encoder-

²³ This procedure removes most of these errors in the data, but not all of them. Some are later caught by our anomaly detection algorithm.

decoder architecture to map the data into a lower-dimensional representation, but adds an extra regularization term that forces the data to follow a known distribution in latent space (e.g., a normal distribution). In this way, it is possible to generate multiple reconstructions for each sample and obtain an averaged reconstruction error that defines the anomaly score.

We implement a VAE using `PyTorch` 1.0.0, and select the hyperparameters using the unsupervised loss function over a validation set composed only of inliers.

Deep support vector data description.

Deep support vector data description (SVDD; Ruff et al. 2018) is a neural network-based approach related to OCSVM (Schölkopf et al. 1999) where a hypersphere (instead of a hyperplane) is used to separate normal samples from abnormal ones. The idea is to learn a new feature representation using neural networks, such that this representation lies in a hypersphere of minimum volume.

In practice, an AE is trained until the loss function converges. Then, the decoder is removed and the center \mathbf{c} of the hypersphere is estimated as

$$\mathbf{c} = \frac{1}{N} \sum_{i=1}^N E(\mathbf{x}_i; \theta_{\text{AE}}^*), \quad (1)$$

where \mathbf{x}_i for $i = 1, \dots, N$ represents each of the N data points in the training set and θ_{AE}^* denotes the parameters of the trained AE. Finally, the vector of the parameters θ of the encoder is reoptimized using θ_{AE}^* as pretrained parameters, by minimizing the following objective function:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \|E(\mathbf{x}_i; \theta) - \mathbf{c}\|^2 + \frac{\lambda}{2} \theta^T \theta, \quad (2)$$

where λ is a hyperparameter that controls the weight decay regularizer on the network parameters θ .

At test time, we define the anomaly score $A(\cdot)$ as the distance between each data point and the center of the hypersphere as follows:

$$A(\mathbf{x}_i) = \|E(\mathbf{x}_i; \theta^*) - \mathbf{c}\|^2, \quad (3)$$

where θ^* denotes the parameters of the trained neural network. We implement deep SVDD using `PyTorch` 1.0.0. The hyperparameter $\lambda = 0.5 \times 10^{-6}$ is selected by measuring the unsupervised loss function over a validation set.

Multiclass deep SVDD.

In this work we extend deep SVDD taking account the information contained in class labels. Instead of modeling one hypersphere, we model multiple hyperspheres, each corresponding to a given class (Pérez-Carrasco et al. 2023). Objects from the same class should be close to each other and far from objects from different classes. As abnormal samples come from unseen classes, their distances to each hypersphere should be larger than those of normal data points. We name this method multiclass deep SVDD (MCDSVDD). By following this approach, it is possible to define an anomaly score based on the distance of the data points to the centers of the hyperspheres.

As in deep SVDD, an AE is trained until convergence, and the decoder is removed. Assuming normal data pairs coming from M different classes $y \in \{1, \dots, M\}$, the center of each

hypersphere is estimated as

$$\mathbf{c}_j = \frac{1}{M_j} \sum_{i=1}^N \mathbf{1}_{(y_i=j)} E(\mathbf{x}_i; \theta_{\text{AE}}^*), \quad (4)$$

where θ_{AE}^* denotes the parameters of the trained AE, $\mathbf{1}_{(y_i=j)}$ is an indicator function that becomes 1 if $y_i = j$ and 0 otherwise, and M_j is the number of data points that belong to class j .

Using θ_{AE}^* as pretrained parameters, the parameter vector θ of the encoder is reoptimized following the objective

$$\min_{\theta} \sum_{j=1}^M \frac{1}{M_j} \sum_{i=1}^N \mathbf{1}_{(y_i=j)} \|E(\mathbf{x}_i; \theta) - \mathbf{c}_j\|^2 + \frac{\lambda}{2} \theta^T \theta, \quad (5)$$

where λ is a hyperparameter that controls the weight decay regularizer on the network parameters θ .

At test time, the anomalous score $A(\cdot)$ is determined by measuring the distance of each data point to the center of its closest hypersphere as follows:

$$A(x_i) = \min_j \|E(\mathbf{x}_i; \theta^*) - \mathbf{c}_j\|^2, \quad (6)$$

where θ^* denotes the parameters of the trained neural network. We implement MCDSVDD using `PyTorch` 1.0.0, and we set the hyperparameter $\lambda = 0.5 \times 10^{-6}$ by cross-validating the unsupervised loss function over a validation set composed of inliers.

4. Methodology

The same procedure is used to train and evaluate all the algorithms described in Section 3.

We randomly split the data into a training set (80%) and a test set (20%) in a stratified fashion in order to preserve the proportion of samples per class. The training set is divided into five stratified subsets in order to perform fivefold cross-validation for model selection. Figure 1 shows a scheme of our training and evaluation methodology.

4.1. Training

Our anomaly detection approaches are based on the hierarchical structure of the ALerCE light-curve classifier (Sánchez-Sáez et al. 2021b). Following the ALerCE taxonomy, we split our training data set into three main classes: transient, stochastic, and periodic. We further divide the data into the 14 subclasses described in Section 2.2. We use three anomaly detectors: one for each of the main classes. As in real-world scenarios anomalies are unknown, so when performing cross-validation we choose a subclass of each main class as the anomalous class and we remove it from the training set. The model is trained on the remaining subclasses and the removed subclass is used for evaluation purposes. Therefore, the model does not use data of the chosen anomalous subclass while training. The process is repeated for each subclass. The overall performance of the anomaly detectors is obtained by evaluating how good the models are at finding the removed subclasses, as is common in machine-learning literature (Ruff et al. 2018).

4.2. Evaluation

Although in practice real outlier events are not available for evaluation, we construct a realistic scenario to select the most promising models for anomaly detection. As explained above, 20% of the data set is kept as a test set (TS1) so no model ever

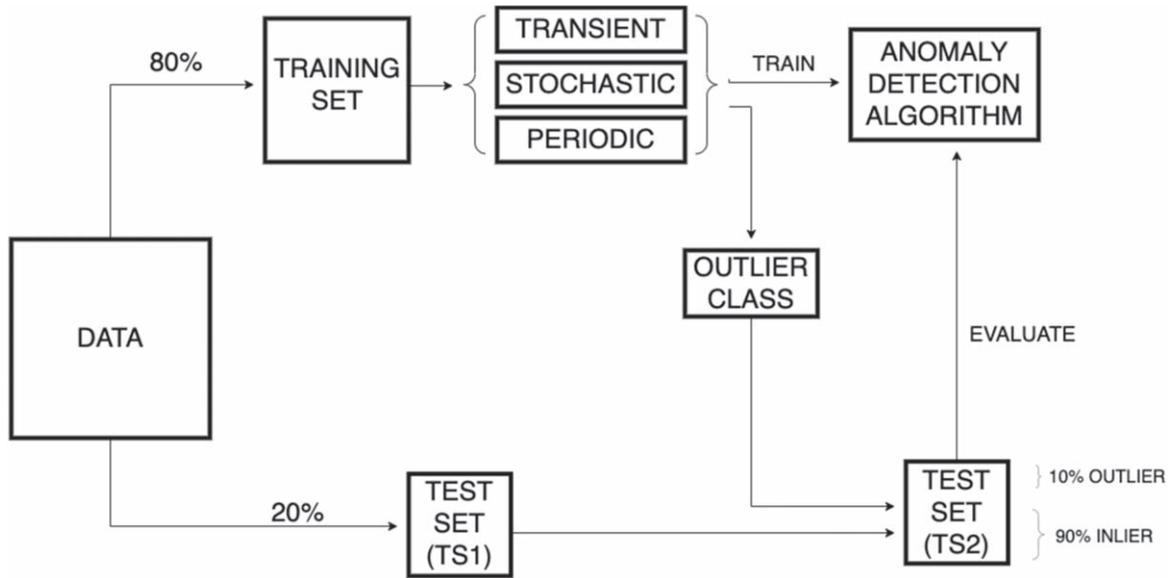


Figure 1. Methodology for training and evaluation of the anomaly detection algorithms. We split the data into a training set and a test set, composed of 80% and 20% of the data, respectively. The training set is subdivided into transient, stochastic, and periodic data. For each of these classes, we choose each subclass as the outlier class. The outlier class is removed from the training set and added to the test set (TS2). Then, an anomaly detection algorithm is trained using the remaining objects of each of the classes, and is evaluated using TS2.

uses it for training. When training, the models use only the inlier subclasses of the remaining 80% of the data set. As described in Section 4.1, the outlier class is iteratively selected from the ALerCE taxonomy and removed from the training set. To construct a realistic evaluation scenario, we define a second test set (TS2) that includes all the objects from TS1 that belong to the inlier subclasses, and also objects belonging to the outlier class from both TS1 and the training set. We make sure that TS2 is composed of 10% outliers and 90% inliers. In this way, we evaluate how good the models are at finding the chosen outlier class from TS2. Figure 1 shows a diagram of the proposed evaluation methodology.

In order to evaluate the model performance when a subclass is selected and treated as an outlier, we use the area under the receiver operating characteristic curve (AUROC; Davis & Goadrich 2006). The receiver operating characteristic curve is a graph that shows the true-positive rate tpr ²⁴ and the false-positive rate fpr ²⁵ at different discrimination threshold values. Computing the area under the generated curve we obtain the AUROC value, which represents the probability that a randomly selected positive sample (outlier) has an anomaly score greater than that of a negative sample (inlier; Hanley & Mcneil 1982). Consequently, a random positive example detector achieves a 0.5 AUROC value, and a perfect classifier achieves a 1.0 AUROC value.

5. Results and Discussion

5.1. Defining the Best Outlier Model for Each Class

Table 1 shows the cross-validated AUROC values for each of the anomaly detection models described in Section 3. This table shows the performance when each of the 14 subclasses is considered as an outlier in a different trial (see

²⁴ $tpr = tp/(tp + fn)$, where tp is the number of true positives and fn is the number of false negatives.

²⁵ $fpr = fp/(fp + m)$, where fp is the number of false positives and m is the number of true negatives.

Section 2.2). The highest AUROC values for each subclass are marked in boldface, showing the best evaluated model for such tasks.

As can be seen, our proposed method MCDSVDD consistently outperforms all the other methods we consider for transient and periodic objects. Our method is able to detect SNII, SNIa, and SNIbc as outliers with higher AUROC values than the rest. For the subclass SLSN, the AE’s performance shows no difference from MCDSVDD’s (i.e., the difference is not statistically significant; p -value = 0.0789). For the periodic classes, MCDSVDD shows a better performance at detecting CEP, DSCT, E, and RRL as outliers, while the AE shows a better performance at detecting LPV. For stochastic sources, the AE method outperforms the other methods for four of the five subclasses (AGN, Blazar, CV/Nova, and YSO), while for QSO, the highest performance is obtained by OCSVM. It is important to mention that QSO light curves usually show slow and smooth temporal variations that make them harder to detect when treated as anomalies. Therefore, the best outlier detection algorithm for stochastic sources is the AE, and the best one for transient and periodic sources is MCDSVDD. We use these models for the evaluation hereafter.

Figure 2 shows the cumulative distribution of the anomaly score for each of the top-level classes. As can be seen, within the stochastic objects, the CV/Nova and YSO subclasses generally result in larger anomaly scores compared to the other stochastic subclasses (AGN/QSO/Blazar). This means that if we select any threshold value in the stochastic class, we will have a large number of CV/Nova sources, and only a few AGN sources. For that reason, we decide to select a homogeneous sample including the same number of sources per subclass (see Section 5.2) to evaluate the outlier candidates.

5.2. Evaluation of the Outlier Detectors in a Real-world Scenario

Developing and integrating an anomaly detector into a brokering system such as ALerCE that processes hundreds of

Table 1
Evaluation of the Performance of Each Model When Applied to Each Class of the ALerCE Top-level Taxonomy

Method	Transient				Stochastic					Periodic				
	SLSN	SNI	SNIa	SNIbc	AGN	Blazar	CV/Nova	QSO	YSO	CEP	DSCT	E	RRL	LPV
IForest (Liu et al. 2012)	0.640 ±0.014	0.721 ±0.021	0.428 ±0.032	0.490 ±0.038	0.573 ±0.017	0.710 ±0.009	0.975 ±0.001	0.468 ±0.016	0.913 ±0.003	0.359 ±0.007	0.295 ±0.012	0.469 ±0.021	0.549 ±0.033	0.971 ±0.007
OCSVM (Schölkopf et al. 1999)	0.577 ±0.014	0.587 ±0.014	0.434 ±0.021	0.492 ±0.011	0.532 ±0.008	0.443 ±0.002	0.909 ±0.001	0.517 ±0.005	0.792 ±0.005	0.432 ±0.004	0.557 ±0.005	0.555 ±0.003	0.539 ±0.004	0.943 ±0.001
AE (Rumelhart & McClelland 1987)	0.736 ±0.022	0.807 ±0.021	0.438 ±0.015	0.537 ±0.019	0.701 ±0.010	0.762 ±0.006	0.980 ±0.016	0.443 ±0.004	0.990 ±0.001	0.564 ±0.024	0.367 ±0.015	0.864 ±0.009	0.907 ±0.015	0.996 ±0.000
VAE (Kingma & Welling 2014)	0.669 ±0.015	0.690 ±0.023	0.404 ±0.018	0.522 ±0.025	0.596 ±0.007	0.597 ±0.010	0.849 ±0.028	0.500 ±0.009	0.795 ±0.009	0.442 ±0.010	0.417 ±0.007	0.561 ±0.007	0.451 ±0.006	0.936 ±0.007
Deep SVDD (Ruff et al. 2018)	0.644 ±0.043	0.731 ±0.043	0.475 ±0.040	0.507 ±0.040	0.496 ±0.025	0.607 ±0.044	0.932 ±0.015	0.411 ±0.008	0.901 ±0.022	0.707 ±0.027	0.482 ±0.054	0.636 ±0.055	0.774 ±0.068	0.785 ±0.025
MCDSVDD (Ours)	0.686 ±0.051	0.828 ±0.024	0.624 ±0.039	0.584 ±0.032	0.706 ±0.069	0.512 ±0.113	0.770 ±0.127	0.483 ±0.080	0.854 ±0.041	0.858 ±0.025	0.819 ±0.015	0.945 ±0.006	0.953 ±0.003	0.953 ±0.008

Notes. Each row represents a different outlier detection algorithm, and each column represents the subclass considered as an outlier. The performance is evaluated using the cross-validation AUROC scores.

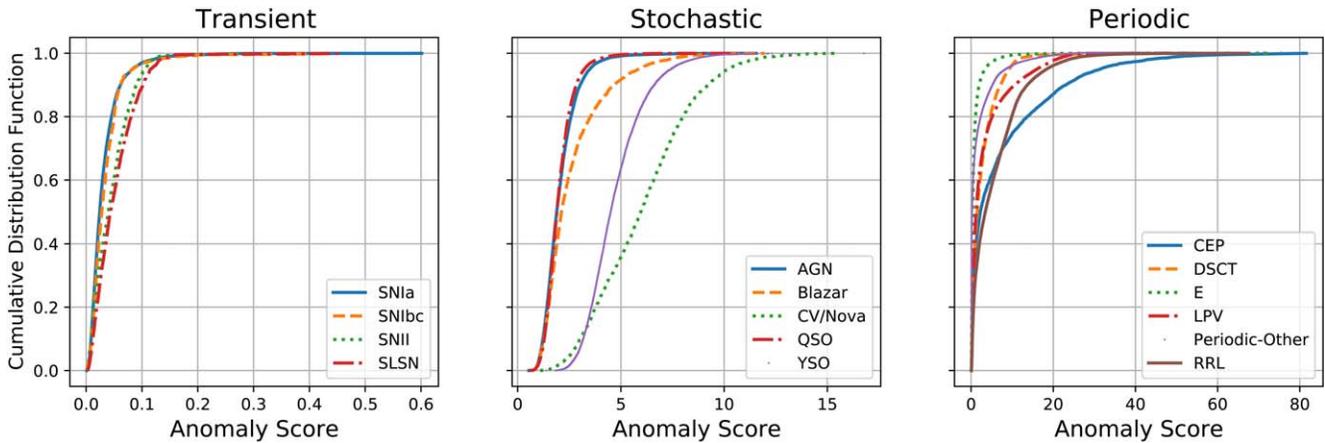


Figure 2. Cumulative distribution function of the anomaly scores for the transient, stochastic, and periodic objects. It is important to note that the anomaly score range displayed in each plot corresponds to the score range covered on the respective top-level class.

thousands of alerts per night online is a challenging task. We cannot assume that the top-level class, namely transient, periodic, or stochastic, is known. For example, a priori we do not know which sources are periodic (Tsang & Schultz 2019; Twomey et al. 2019) or SNe (Villar et al. 2020; Ishida et al. 2021). In this section, we analyze the performance of the anomaly detection algorithm on a real unlabeled alert data set received by ALERCE.

In order to further analyze the results obtained from the anomaly detector, we feed the outlier models with 506,451 alert light curves from ZTF.²⁶ Each source is assigned an anomaly score from the outlier model corresponding to the class given to it by the top-level ALERCE light-curve classifier. A list of 150 sources with the highest outlier scores for each subclass is selected and discussed by a team of 12 astronomers (four experts in each of the three categories, namely transients, stochastic sources, and periodic stars). In what follows, this team is called the “inspection team.” The list of candidates to be inspected by the experts is selected through the following procedure: First, we use the ALERCE light-curve classifier to estimate the most probable class between transient, periodic, and stochastic. For each of the predicted 15 subclasses within the three top-level classes (see Section 2.2), we select the top 10 objects with the highest anomaly scores, using the corresponding anomaly detection algorithm (see Section 3). In order to avoid selecting misclassified objects as outliers, we select only objects whose final probability is consistent with their top-level probability. For instance, we exclude objects classified as transient and CV/Nova at the same time. It is worth noting that in general the outlier candidates are not homogeneously distributed among the subclasses, whereas if we selected some threshold value of the anomaly score to select outliers we would obtain a larger number of events classified into some of the subclasses.

The list of 150 objects is visually inspected by the inspection team. They check the light curves derived from the alert stream, as well as the light curves contained in ZTF Data Release 6, the image stamps, and any other external catalog information to group the outlier candidates into four categories:

1. Outlier (OL): An interesting object that could be considered an outlier. This includes subclasses that are not part of the ALERCE taxonomy.
2. Wrong class (WC): The predicted class given by the top-level light-curve classifier is not correct. An example of this would be a stochastic source classified as a transient or periodic source.
3. Nothing special (NS): The light curve is correctly classified by the top-level light-curve classifier and should not be considered an outlier. This includes correctly classified sources and misclassified sources that are correctly classified by the top-level light-curve classifier (e.g., an AGN classified as a QSO).
4. Bogus (B): There is some error in the object’s data (e.g., there is an SN in the template image) or in the processing of the light curve (e.g., wrong periods) that produces an anomalous behavior.

We summarize below the results obtained, and we discuss the most interesting candidates obtained from this analysis. It is worth remarking that some of these sources can be considered to be confirmed outliers, for instance those sources that are not included in the training set because they are quite peculiar, while others are outlier candidates that would require additional research in order for their true nature to be unveiled.

5.2.1. Transient

After removing transient objects with two or more reference images associated and objects with an SN in the template (see Section 2.2), 40 transient sources are selected, 10 for each of the transient subclasses (SNIa, SNIbc, SNI, and SLSN). Among these sources, the inspection team finds four objects that are classified as outliers (OL), 12 stochastic sources wrongly classified as transients (WC), and 24 objects that do not have any particular anomalous pattern (NS). Notice that in this case we do not find bogus candidates because most of them are filtered using the cutoff criteria defined in Section 2.2. Figure 3 (left) shows the fraction of objects classified as OL, WC, and NS as a function of the number of candidates ordered by descending anomaly score. The fraction of sources defined as NS goes down to lower than 40% and then reaches a value of 60%.

²⁶ These correspond to all unlabeled light curves by 2021 July 19 that have at least six detections in either the g or r band and pass the flag criteria defined in Section 2.2.

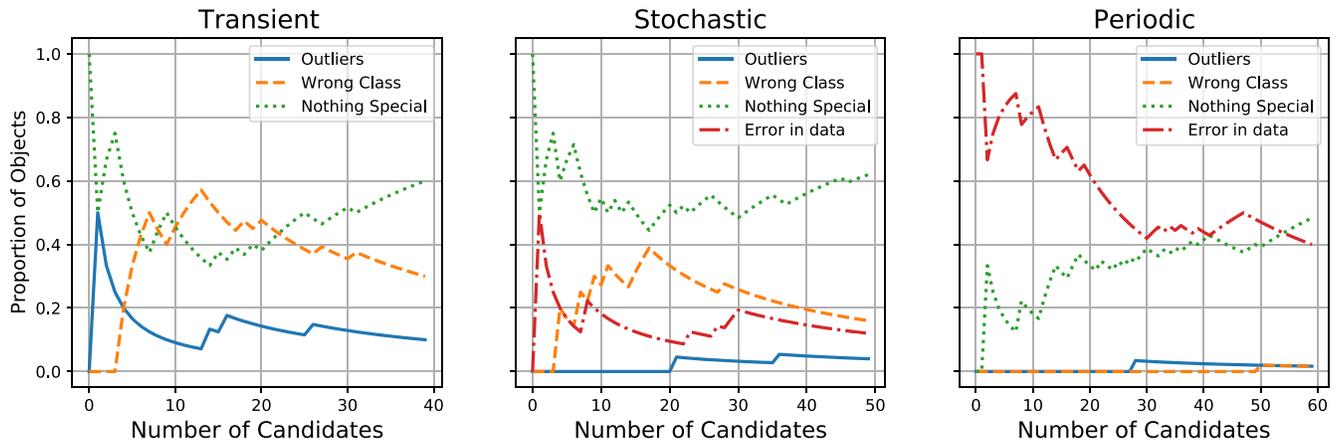


Figure 3. Proportion of objects in terms of the number of selected outlier objects. Notice that as the number of outlier candidates increases, objects with lower anomaly scores are selected. Therefore, the outlier candidates are plotted in decreasing order. Outliers are represented by blue lines, wrong classes by orange lines, “nothing special” sources by green lines, and sources showing errors in the data by red lines. Left: Transient sources. Center: Stochastic sources. Right: Periodic sources.

5.2.2. Stochastic

A total of 50 stochastic sources are selected, 10 for each subclass (AGN, QSO, Blazar, YSO, and CV/Nova). Among these sources, we find two objects that are classified as outliers (OL), eight transients wrongly classified as stochastic objects (WC), 33 objects that do not have any particular anomalous pattern (NS), and seven with SNe in the template (B). Figure 3 (center) shows the fraction of objects classified as OL, WC, B, and NS as a function of the number of outlier candidates ordered by descending anomaly score. The proportion of NS sources stays below 60% between the 7th and 45th ordered candidates. In this case, there is a larger number of sources with errors in the data compared to real outliers. Although they are not astrophysically interesting outlier candidates, they are indeed outliers as they do not follow the stochastic sources’ parameter distribution.

5.2.3. Periodic

We obtain 60 periodic outlier candidates in total, 10 for each of the periodic subclasses (LPV, RRL, CEP, E, DSCT, and Periodic-Other). These candidates are grouped into the four categories defined above (OL, WC, NS, and B). Specifically, we find one outlier source, 34 sources that do not have any particular behavior (NS), 24 objects with errors in data (B; corresponding to objects with wrong periods), and one stochastic source incorrectly classified as periodic (WC). Figure 3 (right) shows the fraction of objects classified as OL, WC, B, and NS as a function of the number of outlier candidates for the periodic objects ordered by descending anomaly score. In this case, most of the 55 sources with the highest anomaly scores show errors in the data such as incorrectly calculated periods. This shows the potential of our model for finding such errors in order to improve preprocessing algorithms.

5.3. Astrophysical Analysis of Selected Outliers

In what follows, we discuss the potential outliers that are selected by the inspection team. External information such as from ZTF Data Release 6, the SIMBAD Astronomical Database,²⁷ the NASA/IPAC Extragalactic Database,²⁸ the Gaia Archive,²⁹ the

Transient Name Server (TNS³⁰), and the International Variable Star Index (VSX³¹) is used to further inspect these sources.

5.3.1. Transient

The light curves of the OL sources are presented in Figure 4. In what follows, we provide notes on each of the corresponding sources.

1. ZTF21aajmdui is classified as an SN Ia by the ALerCE light-curve classifier (note that the ALerCE light-curve classifier uses only the alerts and the alert stream contains only a few points after the peak of the light curve). This however is a confirmed tidal disruption event (TDE; TNS#73970, TNS-Classification#7306; Forster et al. 2020; Hosseinzadeh et al. 2020), a class that is not included in the ALerCE taxonomy. Thus, this is a confirmed outlier.³²
2. ZTF21aanfcmk is classified as an SN Ibc by the light-curve classifier. This is a microlensing event (ATel #14575; Tagchi et al. 2021), a class that is not included in the ALerCE taxonomy. Thus, this is a confirmed outlier.³³
3. ZTF20aamtitiw is classified as an SN II by the light-curve classifier, which is consistent with its spectroscopic classification (Dahiwalé & Fremling 2020). However, the light-curve evolution is not typical of “normal” Type II SNe, which is the class on which the classifier is trained. The strong evolution between the initial peak of the light curve over the first ~ 10 days and the subsequent bump in both the g and r bands points to a probable ejecta–circumstellar material interaction. The absolute magnitudes at maximum brightness (assuming zero extinction, -18.7 and -18.5 in the g and r bands, respectively) are consistent with this interaction scenario. Unfortunately, the classification spectrum is taken with relatively low spectral resolution and 16 days after discovery, making it difficult to see any residual signs of

²⁷ <http://simbad.u-strasbg.fr/simbad/>

²⁸ <https://ned.ipac.caltech.edu/>

²⁹ <https://gea.esac.esa.int/archive/>

³⁰ <https://www.wis-tns.org/>

³¹ <https://www.aavso.org/vsx/index.php?view=search.top>

³² <https://alerce.online/object/ZTF21aajmdui>

³³ <https://alerce.online/object/ZTF21aanfcmk>

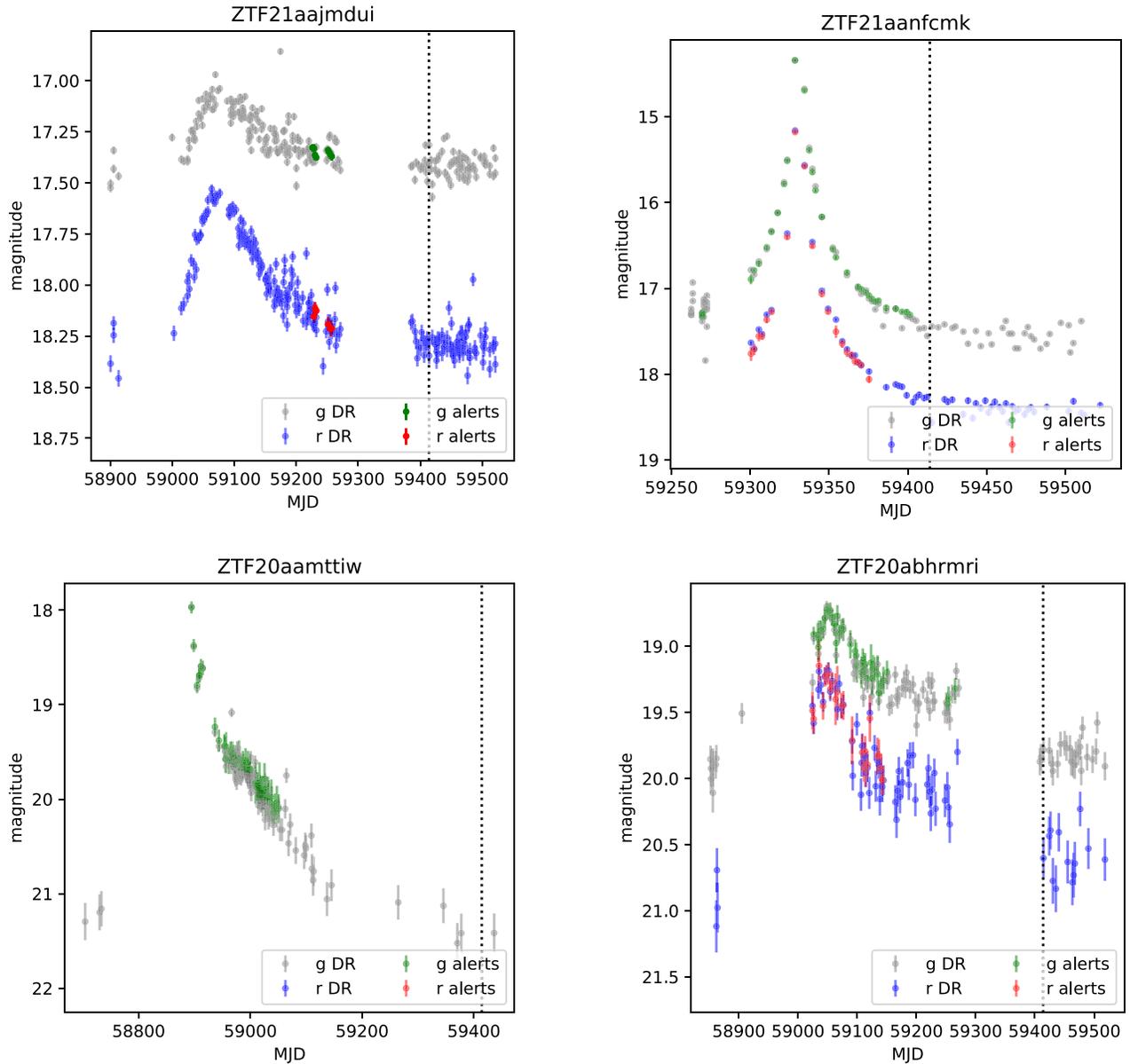


Figure 4. Light curves for outliers classified as transients by the ALerCE light-curve classifier. The light curves include data from the ZTF alert stream in the g (green data points) and r (red data points) filters, as well as from ZTF Data Release 6 in the g (gray data points) and r (blue data points) filters. The gray vertical dotted line represents the last time step used to compute the features and assign the anomaly score.

this early interaction. Therefore, it is a confirmed outlier.³⁴

- ZTF20abhmrri is classified as an SLSN by the light-curve classifier. The light curve shows a slow rise to peak about 25 days after initial detection, and then a slower, shallow decay for ~ 200 days (~ 1 mag over ~ 400 days if we include the data release photometry), which is unusual, even for SLSNe. Notably, the host appears to be compact in Pan-STARRS imaging and infrared-bright in Wide-field Infrared Survey Explorer imaging. The alert position is very close to the galaxy center, leaving open the possibility of this being a TDE or flaring AGN. There are no available follow-up data in the literature. Therefore, this is an outlier candidate.³⁵

5.3.2. Stochastic

The light curves of the OL sources are presented in Figure 5 and we describe them in more detail below:

- ZTF18abgpdfy is classified as a blazar by the light-curve classifier. This source is classified as an apparent R Coronae Borealis (RCB) star in VSX³⁶ (Watson et al. 2006). RCB stars are C-rich, H-deficient red supergiants that undergo dramatic dimming episodes at irregular intervals, caused by mass-loss events and subsequent dust condensation (Nikzat & Catelan 2016, and references therein). DY Per stars have been suggested as a subclass of RCBs (Bhowmick et al. 2018), presenting more symmetrical declines, slower decline rates, and redder colors. The light curve of ZTF 18abgpdfy bears some resemblance to

³⁴ <https://alerce.online/object/ZTF20aamtiiw>

³⁵ <https://alerce.online/object/ZTF20abhmrri>

³⁶ <https://www.aavso.org/vsx/index.php?view=detail.top&oid=1260101>

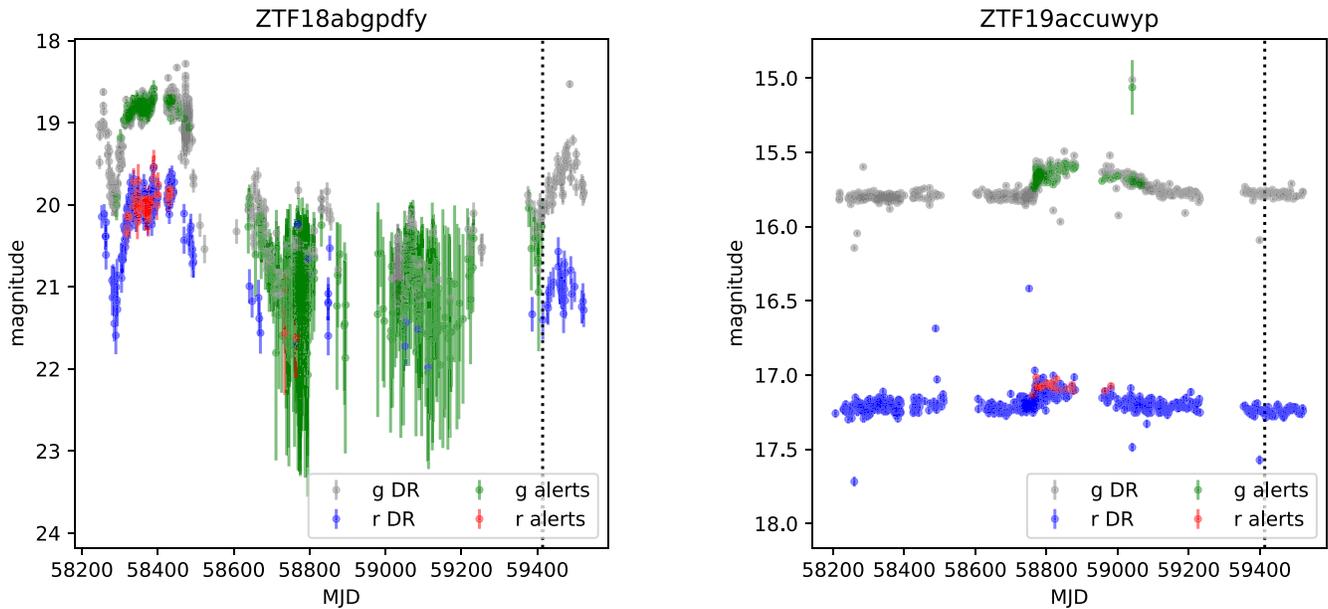


Figure 5. Light curves for outliers classified as stochastic sources by the ALerCE light-curve classifier. Color code as in Figure 4.

that typically seen in DY Per stars, including the prototype (see, e.g., Začs et al. 2007; Shields et al. 2019). Further analysis of ZTF 18abgpdfy’s colors and spectra is thus required to properly establish its variability class (Tisserand et al. 2020, and references therein).³⁷

- ZTF19accuwyp is classified as a YSO by the light-curve classifier. Brightening in YSOs tends to be related to accretion processes and it is expected to occur in two ways (Hillenbrand & Findeisen 2015): as short-duration bursts (lasting days or weeks and with amplitudes $\leq 3-4$ mag in the optical, i.e., EXors) or as much longer timescale outbursts (months to several years or even decades and very large optical amplitudes of ~ 5 mag, i.e., FUors). The timescale of the rise observed in the light curve of ZTF 19accuwyp (~ 1 month), its amplitude (0.2 mag), and the duration of the brightening (~ 1 yr) do not yield a clear correspondence with either of these classes. On the other hand, the Gaia eDR3 distance agrees with the distance to the cluster NGC 7419, which is known to host several Be stars, and this star is very likely a member (Dias et al. 2014; Sampedro et al. 2017; Dias et al. 2018; Cantat-Gaudin & Anders 2020). While the turnoff age of the cluster has previously been estimated at ~ 25 Myr (when accreting protoplanetary disks have already dissipated; Fedele et al. 2010), Subramaniam et al. (2006) reported a much younger turn-on age of 0.3–3 Myr that would be compatible with its YSOs still actively accreting. An optical spectrum taken in 2021 October displayed none of the emission lines expected from accretion (e.g., $H\alpha$, [S II], and [O I]), revealing a hot reddened photosphere, with foreground reddening estimated as $E(B - V) = 2.87$ mag, based on the Schlegel et al. (1998) dust map. Based on this spectrum and the ZTF light curve, this object is classified as an irregular eruptive variable of the γ Cassiopeiae (or Be) type in VSX³⁸ (Watson et al. 2006). On the other hand, the near-

infrared (JHK) colors of the source (Subramaniam et al. 2006) are inconsistent with the color range for both the known Be and Herbig Ae/Be stars in the cluster, and the spectral energy distribution compiled from the optical to the mid-infrared can be qualitatively better explained with a highly reddened pure photosphere, based on a TLUSTY (Lanz & Hubeny 2007) model with a temperature of ~ 3700 K and $A_V \sim 5.3$ mag (extinction in the visible), than with a Herbig Ae/Be–like, disk-bearing system. Further data are needed to confirm the true nature of this source, and thus this is an outlier candidate.³⁹

5.3.3. Periodic

One periodic object is classified as an outlier (OL). Its light curve and periodogram are shown in Figures 6 and 7, respectively.

- ZTF19aaxooyz. A first hypothesis for this star is that it is a type II Cepheid. Indeed, its reported period, light-curve shape, and amplitudes are consistent with what is typically observed in type II Cepheids, when they pulsate in the fundamental radial mode (see, e.g., Soszyński et al. 2008b). However, type II Cepheids with periods in the range of $\sim 5-10$ days are relatively uncommon, in comparison with type II Cepheids with both longer and shorter periods (Matsunaga et al. 2011). In addition, the light curves of type II Cepheids with periods ~ 8 days are typically more sinusoidal and have significantly smaller amplitudes, compared to those of both longer- and shorter-period ones. If such intermediate-period type II Cepheids are not included in sufficient numbers in the training set, they could plausibly be considered outliers by the classifier.⁴⁰

Another possibility is that ZTF 19aaxooyz is a classical Cepheid, in which case its period would imply

³⁷ <https://alerce.online/object/ZTF18abgpdfy>

³⁸ <https://www.aavso.org/vsx/index.php?view=detail.top&oid=2225967>

³⁹ <https://alerce.online/object/ZTF19accuwyp>

⁴⁰ <https://alerce.online/object/ZTF19aaxooyz>

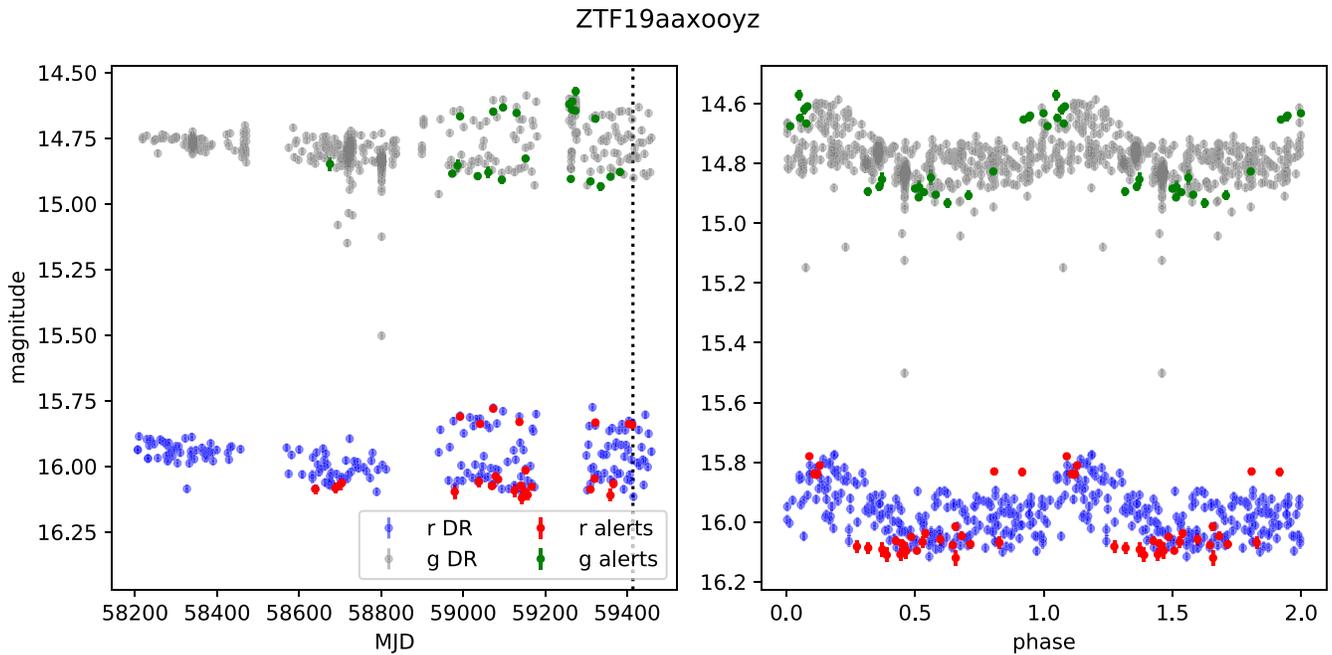


Figure 6. Light curve for the outlier source classified as periodic by the ALerCE light-curve classifier, namely ZTF 19aaxooyz. The left panel shows the original light curve, whereas the right panel shows the folded light curve using a period of 8.071 days. Color code as in Figure 4.

that it pulsates in the fundamental mode. This classification would seem unlikely at first, in view of the fact that ZTF 19aaxooyz does not present bumps in its light curve, contrary to what would be expected for stars (like this one) within the “bump Cepheid” period regime of the so-called Hertzsprung progression (e.g., Soszyński et al. 2008a). However, a number of classical Cepheids that deviate from this trend are also known: at least in the LMC, they have periods in the range of 6–12 days, small amplitudes, and bumps that are either inconspicuous or not present.⁴¹ It is thus also possible that the anomaly detector identifies this as an anomaly in view of the fact that there may be too few, if any, classical Cepheids with this kind of behavior in the training set.

Be that as it may, the (type II/classical) Cepheid hypothesis is untenable, in view of Gaia results, which imply that the star is much too close, with a parallax of 0.750 ± 0.026 mas according to Gaia’s eDR3 (Lindgren et al. 2021), corresponding to a distance of only ~ 1.3 kpc. Thus, the star is most likely not a pulsator, but rather a different type of variable star, possibly a rotational variable whose light output is modulated by spots (e.g., Iwanek et al. 2019). If so, the fact that the amplitude of the light curve increases over the timespan of the available observations could suggest that we are witnessing the evolution of the spot pattern on the star’s surface in the course of a longer-term magnetic cycle. Thus, further analysis is still required in order to properly establish the variability status of ZTF 19aaxooyz. Therefore, this is an outlier candidate.

6. Conclusions and Future Work

Broker systems, such as ALerCE, analyze hundreds of thousands of alerts per night, and aim at analyzing millions of alerts when the Vera Rubin Observatory starts operating. This massive data stream provides an opportunity to discover both rare and unknown astrophysical sources. Doing this in a real alert stream is a challenging task given all the different types of objects appearing in the sky. In this paper, we present the ALerCE anomaly detection framework, which aims at automatically discovering transient, stochastic, and periodic sources that do not belong to the ALerCE taxonomy, as well as at identifying anomalous/peculiar sources within the known classes. This is a general-purpose anomaly detection algorithm, as opposed to those of previous work found in the literature, which aim at detecting outliers within specific classes of objects (e.g., transients).

Our methods are trained using the ZTF alert stream and benefit from the ALerCE light-curve classifier. Specifically, we divide sources into transient, periodic, and stochastic classes, and train six anomaly detection algorithms for each class. Then, these three top-level classes are expanded into 15 subclasses. Each is treated as the anomaly class, which is removed from the training set.

We report the performance of every anomaly detection algorithm in finding each of the chosen subclasses. For stochastic sources, the best results are obtained by calculating the reconstruction error of an AE neural network, while for transient and periodic sources the best results are obtained by a proposed modification of the deep SVDD neural network.

We validate our framework in a real-world scenario by selecting the 10 sources with the highest outlier scores for each of the 15 classes predicted by the ALerCE light-curve classifier. The light curves of these outlier candidates are then inspected, along with information from the literature when available, by a team of 12 astronomers. We impose consistency between the subclass classification of the source and its top-

⁴¹ A sample of light curves for objects of this type can be found in the OGLE Atlas of Variable Star Light Curves (http://ogle.astrouw.edu.pl/atlas/classical_Cepheids.html).

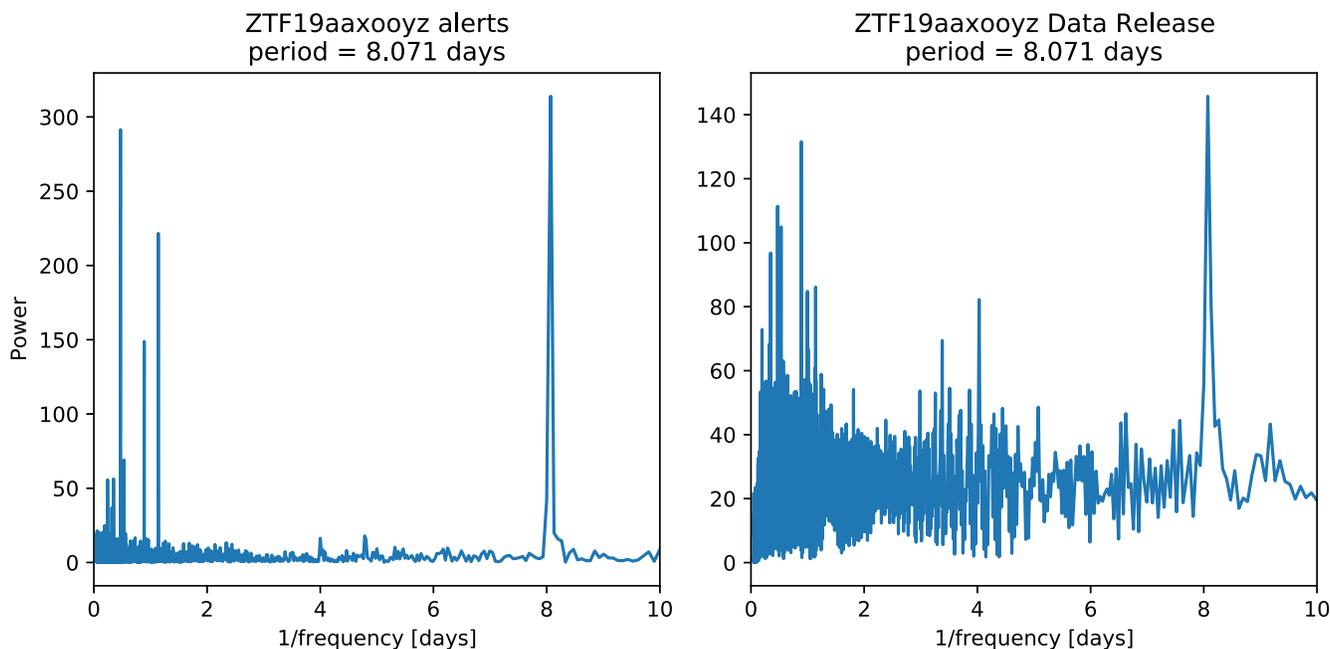


Figure 7. Periodogram for ZTF 19aaxooyz using ZTF alert (left) and ZTF Data Release 6 (right) data. Both periodograms are calculated using P4J (Huijse et al. 2018).

level classification (transient, stochastic, periodic). A total of (40, 50, 60) candidates are inspected, (12, 8, 1) of them being wrongly classified, (24, 33, 34) being nothing special, (0, 7, 24) showing errors in the data, and (4, 2, 1) being potential outliers. Thus, based on these selection criteria, we find seven outliers that are further analyzed, among them a confirmed TDE and a microlensing event, both of which are not included in the ALerCE taxonomy.

We recall and emphasize the importance of human expertise in confirming potential outliers, as these sources are indeed rare. In particular, in this work we define very simple and homogeneous selection criteria for a large diversity of objects, but in the future, additional feature cuts can be implemented in order to specifically select outlier candidates that may be of particular interest for a given research field. For instance, in addition to the outlier score, one could select a minimum number of detections, a minimum length of the light curve, a period range criterion, and other variability features (or any other metric related to the physical processes to be analyzed) to select sources whose measured parameters fall outside the typical values for a given class.

Our methodological framework will be beneficial for the detection of unusual astrophysical phenomenon sources in surveys employing the next generation of telescopes such as the Vera C. Rubin Observatory’s LSST. Our anomaly detection algorithms are currently being implemented within the ALerCE brokering system.⁴²

Acknowledgments

This work gratefully acknowledges funding from ANID through Millennium Science Initiative Programs ICN12_009 (G.C.-V., L.H.-G., F.F., A.M.M.A., F.E.B., M.C., P.A.E., G.P.) and NCN19_171 (A.B.); CATA-Basal AFB-170002 (F.E.B., M.C.), ACE210002 (F.E.B., M.C.), and FB210003 (F.E.B., M.C., F.F.); FONDECYT Regular 1200710 (F.F.), 1190818 (F.E.

B.), and 1200495 (F.E.B.); FONDECYT Postdoctorado 3200250 (P.S.-S.); and infrastructure funds QUIMAL140003 and QUIMAL190012. We acknowledge support from REUNA Chile, which hosts and maintains some of our infrastructure. This work has been possible thanks to AWS–University of Chile–NLHPC credits and computational resources provided by the Data Observatory. Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). This research makes use of the community-built resource VSX available at the AAVSO. We additionally thank G. Hajdu for useful discussions.

ORCID iDs

Manuel Perez-Carrasco <https://orcid.org/0000-0003-4644-8698>
 Guillermo Cabrera-Vives <https://orcid.org/0000-0002-2720-7218>
 Lorena Hernandez-García <https://orcid.org/0000-0002-8606-6961>
 F. Förster <https://orcid.org/0000-0003-3459-2270>
 Paula Sanchez-Saez <https://orcid.org/0000-0003-0820-4692>
 Alejandra M. Muñoz Arancibia <https://orcid.org/0000-0002-8722-516X>
 Javier Arredondo <https://orcid.org/0000-0002-2045-7134>
 Nicolás Astorga <https://orcid.org/0000-0001-6172-9362>
 Franz E. Bauer <https://orcid.org/0000-0002-8686-8737>
 Amelia Bayo <https://orcid.org/0000-0001-7868-7031>
 M. Catelan <https://orcid.org/0000-0001-6003-8877>
 Raya Dastidar <https://orcid.org/0000-0001-6191-7160>
 P. A. Estévez <https://orcid.org/0000-0001-9164-4722>
 Paulina Lira <https://orcid.org/0000-0003-1523-9164>
 Giuliano Pignata <https://orcid.org/0000-0003-0006-0188>

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, *PhRvL*, 116, 241102
 Abbott, R., Abbott, T. D., Abraham, S., et al. 2020, *PhRvL*, 125, 101102
 Baron, D., & Poznanski, D. 2016, *MNRAS*, 465, 4530

⁴² <https://alerce.online/>

- Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2018, *PASP*, **131**, 018002
- Bhowmick, A., Pandey, G., Joshi, V., & Ashok, N. M. 2018, *ApJ*, **854**, 140
- Cantat-Gaudin, T., & Anders, F. 2020, *A&A*, **633**, A99
- Carrasco-Davis, R., Reyes, E., Valenzuela, C., et al. 2021, *AJ*, **162**, 231
- Chandola, V., Banerjee, A., & Kumar, V. 2009, *ACM Comput. Surv.*, **41**
- Chen, J., Sathe, S., Aggarwal, C. C., & Turaga, D. S. 2017, in Proc. 2017 SIAM International Conference on Data Mining (SDM) (Philadelphia, PA: SIAM), 90
- Cortes, C., & Vapnik, V. 1995, *Mach. Learn.*, **20**, 273
- Dahiwal, A., & Fremling, C. 2020, Transient Name Server Classification Report, 2020-667, 1
- Davis, J., & Goadrich, M. 2006, in Int. Conf. on Machine Learning (ICML 2006) (New York: ACM), 233
- Dias, W. S., Monteiro, H., & Assafin, M. 2018, *MNRAS*, **478**, 5184
- Dias, W. S., Monteiro, H., Caetano, T. C., et al. 2014, *A&A*, **564**, A79
- Duev, D. A., Mahabal, A., Masci, F. J., et al. 2019, *MNRAS*, **489**, 3582
- Edgeworth, M. A. 1887, *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, **23**, 364
- Fedele, D., van den Ancker, M. E., Henning, T., Jayawardhana, R., & Oliveira, J. M. 2010, *A&A*, **510**, A72
- Forster, F., Bauer, F. E., Galbany, L., et al. 2020, Transient Name Server Discovery Report, 2020-1791
- Förster, F., Cabrera-Vives, G., Castillo-Navarrete, E., et al. 2021, *AJ*, **161**, 242
- Frederick, S., Gezari, S., Graham, M. J., et al. 2021, *ApJ*, **920**, 56
- Golan, I., & El-Yaniv, R. 2018, in Advances in Neural Information Processing Systems, ed. S Bengio et al., Vol. 31 (Red Hook, NY: Curran Associates, Inc.), <https://proceedings.neurips.cc/paper/2018/file/5e62d03aec0d17facfc5355dd90d441c-Paper.pdf>
- Graham, M. J., Djorgovski, S. G., Drake, A. J., et al. 2017, *MNRAS*, **470**, 4112
- Graham, M. J., Ford, K. E. S., McKernan, B., et al. 2020, *PhRvL*, **124**, 251102
- Hanley, J., & Mcneil, B. 1982, *Radiology*, **143**, 29
- Hillenbrand, L. A., & Findeisen, K. P. 2015, *ApJ*, **808**, 68
- Hosseinzadeh, G., Hiramatsu, D., DeMarchi, L., et al. 2020, Transient Name Server Classification Report, 2020-2478
- Huijse, P., Estévez, P. A., Förster, F., et al. 2018, *ApJS*, **236**, 12
- IceCube Collaboration, Aartsen, M. G., Ackermann, M., et al. 2018, *Sci*, **361**, eaat1378
- Ishida, E. E. O., Kornilov, M. V., Malanchev, K. L., et al. 2021, *A&A*, **650**, A195
- Iwanek, P., Soszyński, I., Skowron, J., et al. 2019, *ApJ*, **879**, 114
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, **873**, 111
- Kingma, D. P., & Welling, M. 2014, in 2nd Int. Conf. on Learning Representations (ICLR 2014), ed. Yoshua Bengio & Yann LeCun (Banff: DBLP)
- Kupfer, T., Bauer, E. B., Burdge, K. B., et al. 2019, *ApJL*, **878**, L35
- LaMassa, S. M., Cales, S., Moran, E. C., et al. 2015, *ApJ*, **800**, 144
- Lanz, T., & Hubeny, I. 2007, *ApJS*, **169**, 83
- Lindgren, L., Klioner, S. A., Hernández, J., et al. 2021, *A&A*, **649**, A2
- Liu, F. T., Ting, K. M., & Zhou, Z.-H. 2012, *ACM Trans. Knowl. Discov. Data*, **6**, 1
- MacLeod, C. L., Green, P. J., Anderson, S. F., et al. 2019, *ApJ*, **874**, 8
- Mahabal, A., Rebbapragada, U., Walters, R., et al. 2019, *PASP*, **131**, 038002
- Mainzer, A., Bauer, J., Grav, T., et al. 2011, *ApJ*, **731**, 53
- Masci, F. J., Laher, R. R., Rusholme, B., et al. 2018, *PASP*, **131**, 018003
- Matsunaga, N., Feast, M. W., & Soszyński, I. 2011, *MNRAS*, **413**, 223
- Meng, X.-C., Han, Z.-W., Podsiadlowski, P., & Li, J. 2020, *ApJ*, **903**, 100
- Möller, A., Peloton, J., Ishida, E. E., et al. 2021, *MNRAS*, **501**, 3272
- Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, *PASP*, **131**, 118002
- Narayan, G., Zaidi, T., Soraisam, M. D., et al. 2018, *ApJS*, **236**, 9
- Nikzat, F., & Catelan, M. 2016, *IBVS*, **6190**, 1
- Nordin, J., Brinnel, V., Van Santen, J., et al. 2019, *A&A*, **631**, A147
- Nun, I., Pichara, K., Protopapas, P., & Kim, D.-W. 2014, *ApJ*, **793**, 23
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *JMLR*, **12**, 2825
- Pérez-Carrasco, M., Cabrera-Vives, G., Hernández-García, L., et al. 2023, arXiv:2308.05011
- Pietrukowicz, P., Dziembowski, W. A., Latour, M., et al. 2017, *NatAs*, **1**, 0166
- Pruzhinskaya, M. V., Malanchev, K. L., Kornilov, M. V., et al. 2019, *MNRAS*, **489**, 3591
- Reyes, E., & Estévez, P. A. 2020, CoRR, arXiv:2005.07779
- Ruff, L., Vandermeulen, R., Goernitz, N., et al. 2018, in Proc. of Machine Learning Research, Vol. 80, ed. J. Dy & A. Krause (Stockholm: PMLR), 4393, <http://proceedings.mlr.press/v80/ruff18a.html>
- Rumelhart, D. E., & McClelland, J. L. 1987, Learning Internal Representations by Error Propagation (Cambridge, MA: MIT Press), 318
- Sakurada, M., & Yairi, T. 2014, in Proc. MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis—MLSDA (New York: ACM)
- Sampedro, L., Dias, W. S., Alfaro, E. J., Monteiro, H., & Molino, A. 2017, *MNRAS*, **470**, 3937
- Sánchez-Sáez, P., Lira, H., Martí, L., et al. 2021a, *AJ*, **162**, 206
- Sánchez-Sáez, P., Reyes, I., Valenzuela, C., et al. 2021b, *AJ*, **161**, 141
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, **500**, 525
- Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., & Platt, J. 1999, in Advances in Neural Information Processing Systems 12 (NIPS 1999), ed. S. Solla, T. Leen, & K. Muller (Cambridge, MA: MIT Press), 582
- Shields, J. V., Jayasinghe, T., Stanek, K. Z., et al. 2019, *MNRAS*, **483**, 4470
- Smith, K. W., Williams, R. D., Young, D. R., et al. 2019, *RAAS*, **3**, 26
- Soszyński, I., Poleski, R., Udalski, A., et al. 2008a, *AcA*, **58**, 163
- Soszyński, I., Udalski, A., Szymański, M. K., et al. 2008b, *AcA*, **58**, 293
- Subramaniam, A., Mathew, B., Bhatt, B. C., & Ramya, S. 2006, *MNRAS*, **370**, 743
- Tagchi, K., Isogai, K., Tampo, Y., & Kojiguchi, N. 2021, Transient Name Server Classification Report, 2021-1363
- Tisserand, P., Clayton, G. C., Bessell, M. S., et al. 2020, *A&A*, **635**, A14
- Tsang, B. T.-H., & Schultz, W. C. 2019, *ApJL*, **877**, L14
- Twomey, N., Chen, H., Diethe, T., & Flach, P. 2019, *Neurocomputing*, **342**, 152
- Villar, V. A., Cranmer, M., Contardo, G., Ho, S., & Lin, J. Y.-Y. 2020, arXiv:2010.11194
- Watson, C. L., Henden, A. A., & Price, A. 2006, *SASS*, **25**, 47
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, **140**, 1868
- Xiong, L., Poczós, B., Connolly, A., & Schneider, J. 2010, Anomaly Detection for Astronomical Data (Pittsburg, PA: Carnegie Mellon Univ.)
- Začs, L., Mondal, S., Chen, W. P., et al. 2007, *A&A*, **472**, 247
- Zhou, C., & Paffenroth, R. C. 2017, in Proc. of the 23rd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, KDD'17 (New York: ACM), 665