The Astrophysical Journal Letters, 965:L14 (10pp), 2024 April 20 https://doi.org/10.3847/2041-8213/ad3337

tdescore: An Accurate Photometric Classifier for Tidal Disruption Events

Robert Stein[1], Ashish Mahabal[1,2], Simeon Reusch[3,4], Matthew Graham[1], Mansi M. Kasliwal[1], Marek Kowalski[3,4], Suvi Gezari[5,6], Erica Hammerstein[7,8,9], Szymon J. Nakoneczny[1], Matt Nicholl[10], Jesper Sollerman[11], Sjoert van Velzen[12], Yuhan Yao[13,14], Russ R. Laher[15], and Ben Rusholme[15]
[1] Division of Physics, Mathematics, and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA; rdstein@caltech.edu
[2] Center for Data Driven Discovery, California Institute of Technology, Pasadena, CA 91125, USA
[3] Deutsches Elektronen-Synchrotron (DESY), Platanenallee 6, D-15378 Zeuthen, Germany
[4] Institut für Physik, Humboldt-Universität zu Berlin, D-12489 Berlin, Germany
[5] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[6] Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA
[7] Department of Astronomy, University of Maryland, College Park, MD 20742, USA
[8] Astrophysics Science Division, NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, MD 20771, USA
[9] Center for Research and Exploration in Space Science and Technology, NASA/GSFC, Greenbelt, MD 20771, USA
[10] Astrophysics Research Centre, School of Mathematics and Physics, Queens University Belfast, Belfast BT7 1NN, UK
[11] Department of Astronomy, The Oskar Klein Center, Stockholm University, AlbaNova, 10691 Stockholm, Sweden
[12] Leiden Observatory, Leiden University, Postbus 9513, 2300 RA Leiden, The Netherlands
[13] Miller Institute for Basic Research in Science, 468 Donner Lab, Berkeley, CA 94720, USA
[14] Department of Astronomy, University of California, Berkeley, CA 94720, USA
[15] IPAC, California Institute of Technology, 1200 E. California Boulevard, Pasadena, CA 91125, USA
Received 2023 November 22; revised 2024 March 5; accepted 2024 March 11; published 2024 April 10

Abstract

Optical surveys have become increasingly adept at identifying candidate tidal disruption events (TDEs) in large numbers, but classifying these generally requires extensive spectroscopic resources. Here we present `tdescore`, a simple binary photometric classifier that is trained using a systematic census of ∼3000 nuclear transients from the Zwicky Transient Facility (ZTF). The sample is highly imbalanced, with TDEs representing ∼2% of the total. `tdescore` is nonetheless able to reject non-TDEs with 99.6% accuracy, yielding a sample of probable TDEs with recall of 77.5% for a precision of 80.2%. `tdescore` is thus substantially better than any available TDE photometric classifier scheme in the literature, with performance not far from spectroscopy as a method for classifying ZTF nuclear transients, despite relying solely on ZTF data and multiwavelength catalog cross matching. In a novel extension, we use "Shapley additive explanations" to provide a human-readable justification for each individual `tdescore` classification, enabling users to understand and form opinions about the underlying classifier reasoning. `tdescore` can serve as a model for photometric identification of TDEs with time-domain surveys, such as the upcoming Rubin observatory.

Unified Astronomy Thesaurus concepts: Tidal disruption (1696); Time domain astronomy (2109); Black holes (162); Galaxy nuclei (609); Sky surveys (1464)

1. Introduction

Tidal disruption events occur when stars pass too close to supermassive black holes (SMBHs). The tidal force exerted by the SMBH exceeds the self-gravity holding the star together, and the star disintegrates (Rees 1988). Much of the resulting stellar debris remains gravitationally bound to the SMBH and is ultimately accreted onto the black hole. These TDEs can generate luminous emission across the entire electromagnetic spectrum, from radio to soft gamma rays, and in recent years, all-sky surveys have become increasingly adept at finding the previously elusive class of transients (see Gezari 2021 for a recent review). TDEs offer a unique probe of otherwise-quiescent SMBHs residing in galaxies and can be used to study a variety of areas such as astrophysical jet launching, SMBH demographics, and accretion disk formation.

There are now $\gtrsim$100 TDEs in the literature, the vast majority of which are identified by optical surveys. In particular, the Zwicky Transient Facility (ZTF; Bellm et al. 2019; Dekany et al. 2020) at Palomar Observatory conducts an all-sky survey that has detected ∼90 TDEs since 2018 (see, e.g., van Velzen et al. 2021; Hammerstein et al. 2023; Yao et al. 2023). With this large sample, we now know that at least some TDEs emit quasithermal optical flares with high apparent temperature that rise on a timescale of weeks and fade more slowly over a timescale of months with little apparent temperature evolution (Gezari 2021). These optical TDEs appear to have a marked preference for "green-valley" galaxies (see, e.g., Arcavi et al. 2014; French et al. 2016; Graur et al. 2018; Hammerstein et al. 2021a).

Despite a nominal survey depth of 20.5 mag (Graham et al. 2019), the ZTF TDE program remains incomplete below a magnitude of ≈19.1 mag due to limited spectroscopic resources (Yao et al. 2023). This spectroscopic bottleneck will become even more severe with upcoming instruments and observatories such as the Vera C. Rubin Observatory (Ivezić et al. 2019) and ULTRASAT (Shvartzvald et al. 2024), which are expected to detect thousands of TDEs each year (see, e.g., Bricman & Gomboc 2020; Shvartzvald et al. 2024).

There is thus increasing need for the development of TDE selection methods that do not rely on expensive spectroscopic follow-up. However, photometric classification of nuclear transients remains in its infancy. Although some effort has

been devoted to finding TDEs as part of generic multimodal transient classifiers (see, e.g., Muthukrishna et al. 2019; Graham et al. 2023), the only effort in the literature that was specifically tailored to TDEs was Gomez et al. (2023).

In this Letter, we introduce a novel binary machine-learning photometric classifier, `tdescore`, trained with the sample of ZTF nuclear transients to identify TDEs. The code itself is already available on GitHub[16] and Zenodo (Stein 2024), while the corresponding training data will be released in a dedicated future publication (S. Reusch et al. 2024, in preparation). In Section 2 we introduce this ZTF Nuclear Sample, and in Section 3 we describe the process of generating high-level "features" from the available data. We then outline the `tdescore` classifier itself (Section 4), and explore the reasoning behind the corresponding classifications (Section 5). Finally, in Section 6, we highlight the relevance of `tdescore` to both existing and future surveys.

2. The ZTF Nuclear Transient Sample

The first photometric optical search for TDEs was conducted by van Velzen et al. (2011) using archival searches of Sloan Digital Sky Survey data (York et al. 2000), finding that TDEs can be differentiated from supernovae using light-curve evolution. Photometric identification of TDEs at Palomar began with the predecessor survey to ZTF, the intermediate Palomar Transient Factory (iPTF) survey (Kulkarni 2013). A systematic census of nuclear transients in 4800 deg^2 of iPTF data was used to develop simple algorithmic cuts yielding candidate TDEs with a precision of 20%, which was sufficiently high to serve as a model for spectroscopic surveys (Hung et al. 2018). For the ZTF survey, looser cuts were paired with light-curve analysis for the nuclear transient filter (van Velzen et al. 2019), which has been used to identify dozens of TDEs over the course of the survey (van Velzen et al. 2021; Hammerstein et al. 2023; Yao et al. 2023). The filter was implemented in AMPEL, a real-time data analysis framework and ZTF alert broker (Nordin et al. 2019). The nuclear transient filter itself is an open-source Python script,[17] which broadly selects candidates based on

1. estimated "nuclearity" of the flux-weighted ZTF transient position using proximity to sources detected by the deeper Pan-STARRS1 (PS1) survey (Chambers et al. 2016);
2. probability of detection being "real" based on machine-learning `RealBogus/DeepRealBogus` classification of images (Duev et al. 2019; Mahabal et al. 2019) and algorithmic cuts on image detection parameters;
3. rejection of stellar sources via the machine-learning `sgscore` classification (Tachibana & Miller 2018) of underlying PS1 sources (Chambers et al. 2016), measured parallax in GAIA DR2 (Gaia Collaboration et al. 2018), and cuts on bright hosts ($m < 12$);
4. rejection of Galactic sources by requiring Galactic latitude $|b| > 5$; and
5. rejection of moving objects by requiring multiple time-separated detections of a source.

These cuts are designed to be loose and inclusive, prioritizing recall over precision. As part of the ongoing ZTF TDE program, additional light-curve analysis and ranking are performed to highlight potential TDE candidates (van Velzen et al. 2021), which are then vetted by humans and assigned additional follow-up observations for classification. In many cases, a spectrum is required to resolve ambiguity. With `tdescore`, we aim to develop an alternative to this resource-intensive process using a machine-learning approach.

The nuclear transient filter has been iteratively modified over the course of the survey to improve the false-positive or false-negative rate. To develop `tdescore`, we start with the latest version of the filter, which was developed and applied to all archival ZTF alert data, yielding a uniform sample of 11699 nuclear transients discovered in ZTF-I, from 2018 April 1 to 2020 September 30, and in ZTF-II from 2020 October 1 to 2022 April 30 (S. Reusch et al. 2024, in preparation).

We extract any available classifications for these transients from the `ZTF Fritz Marshal`[18] (van der Walt et al. 2019; Coughlin et al. 2023) and the predecessor `ZTF GROWTH Marshal` (Kasliwal et al. 2019). In general, these are accumulated human-assigned classifications that can be based on spectra (including public ones taken from, e.g., the Transient Name Server, and host spectroscopy from the Sloan Digital Sky Survey, York et al. 2000), light-curve evaluation, or other contextual information. We verify each of these human classifications (see Appendix A for details) and recover 5264 classified sources, of which 86 are classified as TDEs. This includes 30 sources from ZTF-I presented in van Velzen et al. (2021) and Hammerstein et al. (2023), 17 additional bright ZTF-II TDEs from Yao et al. (2023), as well as 39 additional faint or recent TDEs from ZTF-II that have not yet been published.

3. Feature Extraction

3.1. Light-curve Analysis

To develop a flexible framework that could be easily generalized to other surveys, we use a Gaussian Process to convert the extensive photometry from ZTF into more survey-independent, high-level physical features such as peak magnitude and fade rate. We specifically design a multistep fitting procedure tailored to the known characteristics of TDEs, namely that they are blue, long-lived transients with little apparent color evolution. Beyond this, the fitting procedure is agnostic about any underlying physical model for TDE emission and can therefore capture the full diversity of TDE optical emission, including observed TDE outlier behavior such as multiple peaks or long plateaus.

We use the alert photometry provided directly by ZTF as the basis of the analysis. No K-correction is applied to the data, but we do correct for Galactic extinction using results from Schlafly & Finkbeiner (2011) and the extinction law from Fitzpatrick (1999). We perform a series of cuts similar to those in the nuclear filter to remove detections that are not well subtracted, returning a subset of "clean" photometry for each source. We specifically require an FWHM $< 5''$, no bad pixels, a Real/Bogus score >0.3, a pixel distance to host $< 1''$, and a difference image depth of at least 19.0 mag to reject images taken under poor conditions. Though ZTF provides some

[16] https://github.com/robertdstein/tdescore

[17] https://github.com/AmpelAstro/Ampel-nuclear/blob/main/ampel/nuclear/t0/NuclearFilter.py

[18] https://fritz.science/

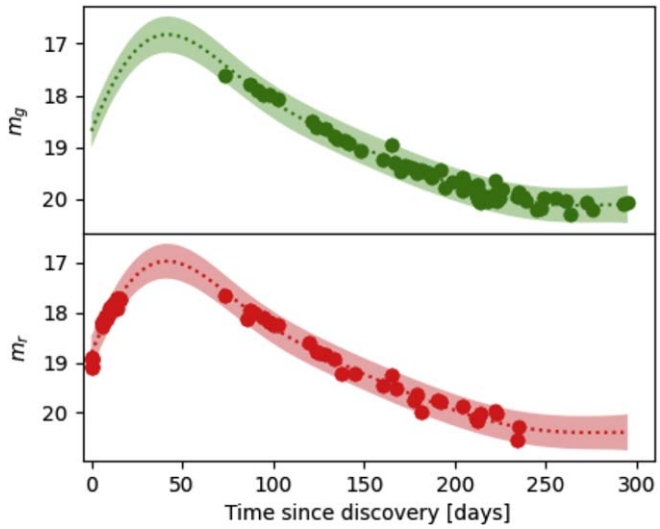

Figure 1. An example of the light-curve fitting procedure on a real TDE, ZTF20achpcvt/AT2020vwl (Hodgkin et al. 2020; Hammerstein et al. 2021b), for which limited data were available at peak. Using the two-step fit, the approximate g-band peak time and the color at peak can be inferred for use in classification. AT2020vwl is relatively red with $(g-r)\approx 0$, but bluer TDEs with $(g-r)>0$ are detected more frequently in g-band. Nonetheless, the fitting procedure still works well for this TDE.

(sporadic) i-band coverage as part of the partnership surveys, we only consider the g-band and r-band data, which are primarily provided in a uniform 2 day cadence by the ZTF MSIP public survey.

Our data set is three dimensional (detections have a flux, wavelength, and time), so we cannot directly apply a simple univariate Gaussian process. While multivariate Gaussian processes have been applied for astronomical data sets, they require a customized covariance matrix to balance variation between bands against variation in time (see, e.g., Aigrain & Foreman-Mackey 2023 for a recent review). Moreover, multivariate Gaussian processes have more associated uncertainty in cases such as here, where the bands are not sampled uniformly.

Instead, we simplify the problem and fit flux in one band as a function of time with a univariate Gaussian process. Given that TDEs are generally blue, we first fit the g-band data with a univariate Gaussian process model implemented in `scikit-learn` (Pedregosa et al. 2011), using a "radial basis function" (RBF) kernel restricted to timescales of 50–500 days, and an additional white-noise component equal to at least 0.1 mag to account for systematic uncertainty and prevent overfitting. After obtaining a model for the g-band data, we then perform a least-squares minimization fit of the r-band data to this g-band light-curve model, under the assumption that the data follow a linear color evolution of the form

$$m_r(t) = m_g(t) + C_0 + (C_1 \times t), \quad (1)$$

where C_0 and C_1 are fit parameters derived for each source and t is the observer time in days. After obtaining these coefficients, we estimate the g-band magnitude of the source for each r-band detection.

We then fit the combined (g-band and converted r-band) light curve with the same univariate Gaussian process procedure. This provides our final model for each source light curve. An example of this fitting is shown in Figure 1 for a real TDE with sparse early data where the joint fit is required to constrain the g-band rise and fade.

With these light-curve fits, we can extract high-level parameters for each source. We specifically extract

1. the peak magnitude in g-band.
2. the time of peak in g-band.
3. the color at g-band peak.
4. the color change rate (C_1).
5. fade time (defined as the time in for the g-band light curve to return from peak to 0.5 mag below peak).
6. the RBF length scale from the Gaussian process fit.
7. the RBF amplitude from the Gaussian process fit.
8. the Gaussian process "score," which quantifies how well the model describes the data.
9. the number of inflection points in the light-curve fit that occur prepeak, and the number of postpeak inflection points, to fully capture the multiple peaks that can be exhibited by many active galactic nuclei (AGN) and some transients.
10. the mean detection cadence (total number of detections divided by time in days between first and last detection).

Entirely independently of the above procedure, we also try to fit the light curves with SALT2 supernova Type Ia (SN Ia) models (Guy et al. 2007) using `sncosmo` (Barbary et al. 2016) and retrieve the underlying c/x_1 parameters, as well as the χ^2, to serve as a proxy for the "Ia"-ness of the light curve.

When run on a standard MacBook Pro without any parallelization, the Gaussian process analysis requires $\sim$3 s per transient on average. The time varies somewhat between individual transients, with more light-curve detections leading to longer process times. `sncosmo` is faster, requiring $\sim$1 s on average per source. The light-curve analysis procedure is thus fast enough to scale to deeper surveys such as Rubin. For surveys with more than two bands, the model in Equation (1) could be generalized to a thermal model with a temperature and linear temperature evolution.

3.2. Additional Features

In addition to parameters directly extracted from the ZTF photometry, additional contextual information is extracted for each source. The ZTF alerts themselves (Masci et al. 2019; Patterson et al. 2019) provide the cataloged "`sgscore`" value for the source host (a binary machine-learning classification score based on morphology to distinguish stars from galaxies, Tachibana & Miller 2018). Each individual detection also contains

1. *distpsnr1*—distance of detection to PS1 host in arcseconds, from which we calculate a median.
2. *distnr*—pixel distance to nearest source in reference image, from which we calculate a median.
3. *sumrat*—the ratio of summed pixels values in a detection to the sum of absolute pixel values, serving as a proxy for yin-yang subtraction artifacts. We calculate a median sumrat for each source.
4. *classtar*—star/galaxy classification score from `SourceExtractor` (Bertin & Arnouts 1996).
5. *isdiffpos*—Boolean value for whether the detection is positive or negative, from which we calculate an overall fraction of positive detections.

We also crossmatch the sources to their underlying PS1 hosts (Chambers et al. 2016), yielding $g-r$, $r-i$, $i-z$, and $z-y$ host colors. By construction, all sources will be close to a

source with at least one PS1 detection. We also crossmatch to mid-infrared host colors (W1–W2, W3–W4) from WISE (Wright et al. 2010) and to underlying W1 variability using WISE+NEOWISE (Mainzer et al. 2014), similar to Yao et al. (2023). We also crossmatch to the `Milliquas` catalog to known radio/X-ray-selected AGN (Flesch 2023), yielding a boolean has_milliquas flag.

4. `tdescore`

4.1. ZTF Nuclear Machine Learning (ML) Data Set

From the nuclear sample, we have 5264 sources with classifications that could in principle be used for analysis. The sample is dominated by the 4218 AGN (80.1%) but also includes 213 core-collapse supernovae (CCSNe; 4.0%), 708 SNe Ia (13.4%), 39 variable stars (0.7%), and 86 TDEs (1.6%). Additional quality cuts are then applied to select a sample of nuclear transients with uniformly derived properties. In particular, we restrict ourselves to sources that passed the light-curve fitting described in Section 3 and had a significantly measured fade time (i.e, were detected at least 0.5 mag below peak). In practice, this requires sources to be detected multiple times in both g and r bands and to have a detection at least 0.5 mag below g-band peak. Of the initial 5264 classified nuclear sources, only 3040 pass this additional "fade and color change" cut. All sources passing this step also have the other relevant light-curve parameters such as score, color at peak, etc.

From these 3040 sources with high-quality light curves, we additionally select those for which all WISE host colors and PS1 host colors were available and for which `sncosmo` successfully ran. Overall, half of the AGN (2153) and core-collapse supernovae (106) pass all cuts, along with 60% of SNe Ia (427) and 64% of TDEs (55). However, only ∼8% variable stars (3) pass, due primarily to their erratic light curves. This ultimately leaves 2744 sources in our final "nuclear ML sample," of which 55 are TDEs and the remaining 2689 are non-TDEs. The share of TDEs thus increases slightly from 1.6% of classified sources to 2.0% of the "nuclear ML sample." These steps are illustrated in Figure 2.

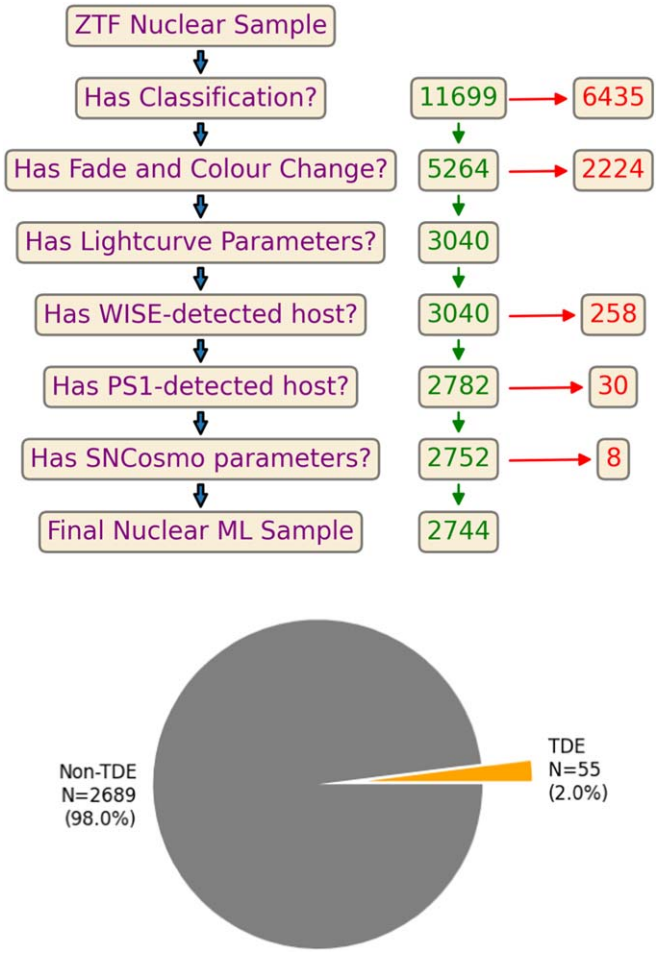

Figure 2. Top: breakdown of the various cuts applied to the ZTF nuclear sample. Of 11,699 ZTF sources; 5264 have a secure classification; 3040 also have a well-measured fade; and 2744 sources pass all cuts. Bottom: of these 2744 sources used to train `tdescore`, 55 (2.0%) are TDEs.

4.2. Training and Testing Sets

Given the small number of TDEs (55) in the data set, it would not be possible to measure classifier performance with reasonable accuracy using a simple division into separate training and testing sets. Even if 20% of the sources were reserved for testing, this would correspond to just ∼11 TDEs, with consequently high uncertainty for metrics such as recall. Furthermore, given the small number of TDEs, the performance of a classifier on the test sample will be strongly influenced by the randomly varying composition of the sample. If "atypical TDEs" were randomly to be allocated to the training set, classifier performance would be much better than if they were allocated to the test set.

Instead, to maximize the number of TDEs available for training and to minimize stochasticity, we employ the "leave one out" k-fold stratified cross validation to create testing and training sets (see, e.g., Hastie et al. 2009). We randomly divide our sample into 55 different, equally sized groups, each containing one TDE. The non-TDEs are randomly sorted and then allocated evenly to one of these groups. As 2689 is not exactly divisible by 55, some groups have 48 non-TDEs, while others have 49 non-TDEs. We select one group to be our test data set and use the remaining 54 groups as a training set. After training, we can derive performance metrics on the test data set.

We can then repeat the process on a second group, again using the other 54 groups as a training set. This process is repeated for every single group in the data set, meaning that 55 different classifiers are trained, with each source being tested once and used for training 55 times. To further reduce the variance in metrics, we repeat the process ten times, each with a different random sorting of the data. By using the average performance of classifiers across groups and iterations, we can obtain more robust estimates of performance and be certain that any outlier sources are fairly represented.

4.3. Data Set Augmentation

Given the severe class imbalance in nuclear transients, where TDEs represent a tiny minority (∼2% of the total), any classifier that simply rejected all candidate TDEs would already have an accuracy of ∼98%. To mitigate this effect, we employ synthetic minority oversampling technique (SMOTE) to generate a balanced training set (Chawla et al. 2002). With SMOTE, for each of the k-fold training sets, we randomly select pairs of TDEs and generate new pseudo-TDEs with properties lying a random distance between the two real TDEs. This process is repeated until the training set contains as many

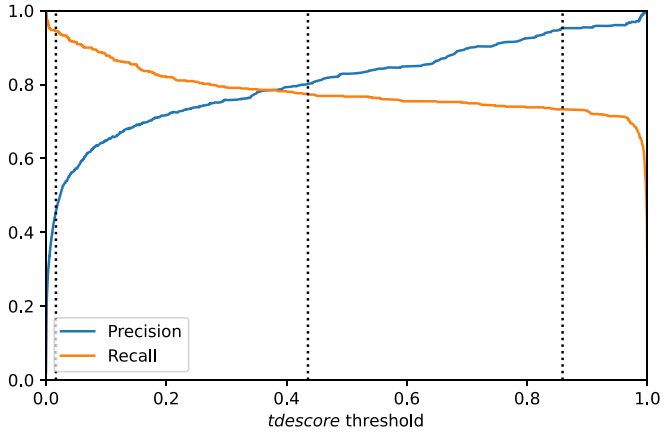

Figure 3. Precision and recall as a function of tdescore threshold. The balanced threshold (chosen such that precision is at least 80%) is illustrated by the central dotted vertical line. The inclusive (>95% recall) and clean (>95% precision) thresholds are illustrated by the left and right dotted vertical lines, respectively. The corresponding confusion matrices for these three scenarios are shown in Figure 4.

TDEs as non-TDEs (and is thus composed of 50% non-TDEs, ∼2% real TDEs, and ∼48% pseudo-TDEs). Once trained on a fold, a classifer can then be tested on the test data, which contains only non-TDEs and real TDEs, to assess its performance. The process of generating pseudo-TDEs via SMOTE is repeated from scratch for each k-fold permutation on the train set, excluding the sources in the test set, so there is no contamination from test data in the training sample.

4.4. Classifier Architecture and Performance

With the balanced training sets built in Sections 4.1–4.3, we can train the tdescore classifier. tdescore is built with the XGBoost algorithm (Chen & Guestrin 2016), which employs a gradient-boosted decision tree architecture to build a classifier. For tdescore, we use the python implementation with 27 features. Given the risk of overfitting on our relatively small data set and the lack of an independent validation set to measure performance, we generally do not modify the default settings in XGBoost.[19] We use 100 estimators, and to mitigate overtraining, we further adopt a subsampling rate of 70% for XGBoost to employ in each iteration of the boosting procedure. During training, we use the area under the precision-recall curve as the optimization metric. Use of this metric ensures that both false positives and false negatives are minimized. The augmentation, training, and testing are rapid, requiring approximately 5 s for a single iteration on a typical MacBook Pro.

Having trained our classifier and applied it to the entire nuclear ML sample, we then require a threshold score to determine to which class each source is assigned. The precision and recall as a function of possible threshold is illustrated in Figure 3. As our base case, we adopt a threshold at which >80% precision[20] is achieved, with the corresponding confusion matrices shown in Figure 4. With this cut, 77.5% of TDEs are successfully recovered (∼43 TDEs). The classifier efficiently rejects non-TDEs, with 99.6% being correctly classified, while just 0.4% are misclassified as TDEs (∼11 non-TDEs). Given the unbalanced sample, this results in 80.2% of tdescore-selected candidates being real TDEs, with 19.8% being non-TDEs.

The appropriate threshold for classifiers such as tdescore ultimately depends on the intended scientific application. A high-precision sample with lower recall[21] may be preferable for rate studies or other population analysis, whereas a high recall might be desired to generate a complete, spectroscopically classified TDE sample where some contamination is acceptable. We consider an alternative stricter threshold, chosen such that at least 95% of tdescore-selected TDEs would be genuine. Applying this higher threshold produces a very clean sample of probable TDEs, which nonetheless retains a recall of 73.3% (∼40 TDEs and ∼2 non-TDEs). This confirms that nearly three-quarters of genuine TDEs are confidently identified, receiving very high classifier scores. We also consider a loose threshold that is nearly complete, chosen such that a recall of at least 95% is achieved. With this loose cut, only ∼5% of TDEs are lost (∼3 TDEs), but the background is rejected with such efficiency (97.4%) that the share of TDEs in the sample reaches 45.7% (∼52 TDEs and ∼62 non-TDEs), versus just 2.0% in the parent training sample. tdescore is thus able to reject most of the background at very little cost to completeness. Further tests of tdescore using subsets of the parameters are detailed in Appendix B, which confirm that much of the background can be rejected even before light-curve information is available.

As a cross-check, we repeat the tdescore training without using the SMOTE augmentation described in Section 4.3. For the balanced threshold (defined as >80% precision), recall slightly increases from 77.5% to 79.5%, but for the clean threshold, recall falls from 73.3% to 72.0%. For the inclusive threshold (defined as >95% recall), precision falls substantially from 45.7% to 29.5%. Overall, the area under the precision/recall curve decreases from 0.893 to 0.882. The data augmentation step thus provides clear performance improvements for cases prioritizing either high recall or precision.

5. Understanding Classifier Reasoning

To have confidence in the results of tdescore, it is important to understand whether classifications are based on sound reasoning. The global importance of different features are listed in Table 1. In agreement with Gomez et al. (2023), we find that color at peak is an important discriminator, confirming the well-known property that TDEs are atypically blue relatively to most other transients. However, given the overwhelming dominance of AGN as contaminant nuclear sources, we find that WISE W1–W2 color is by the far the most important feature in identifying TDEs. This is not unexpected, given the ubiquity of WISE color cuts as a method of selecting AGN (Stern et al. 2012). We also find that sncosmo analysis can be a useful tool, with the resultant χ^2 values being useful proxies for both SNe Ia (with good fits) and AGN (with poor fits).

tdescore also attempts to overcome the "black box problem" by incorporating *explainable AI*. We analyze the tdescore classifier using Shapley additive explanations (SHAP) Python package (Lundberg & Lee 2017). SHAP explains the output of ML classifiers for individual objects by estimating the local importance of each feature for a given

[19] For a full explanation of available settings, see https://xgboost.readthedocs.io/en/stable/python/python_api.html#xgboost.XGBClassifier.

[20] "Precision" is often called "purity" in astronomical contexts.

[21] "Recall" is synonymous with "completeness."

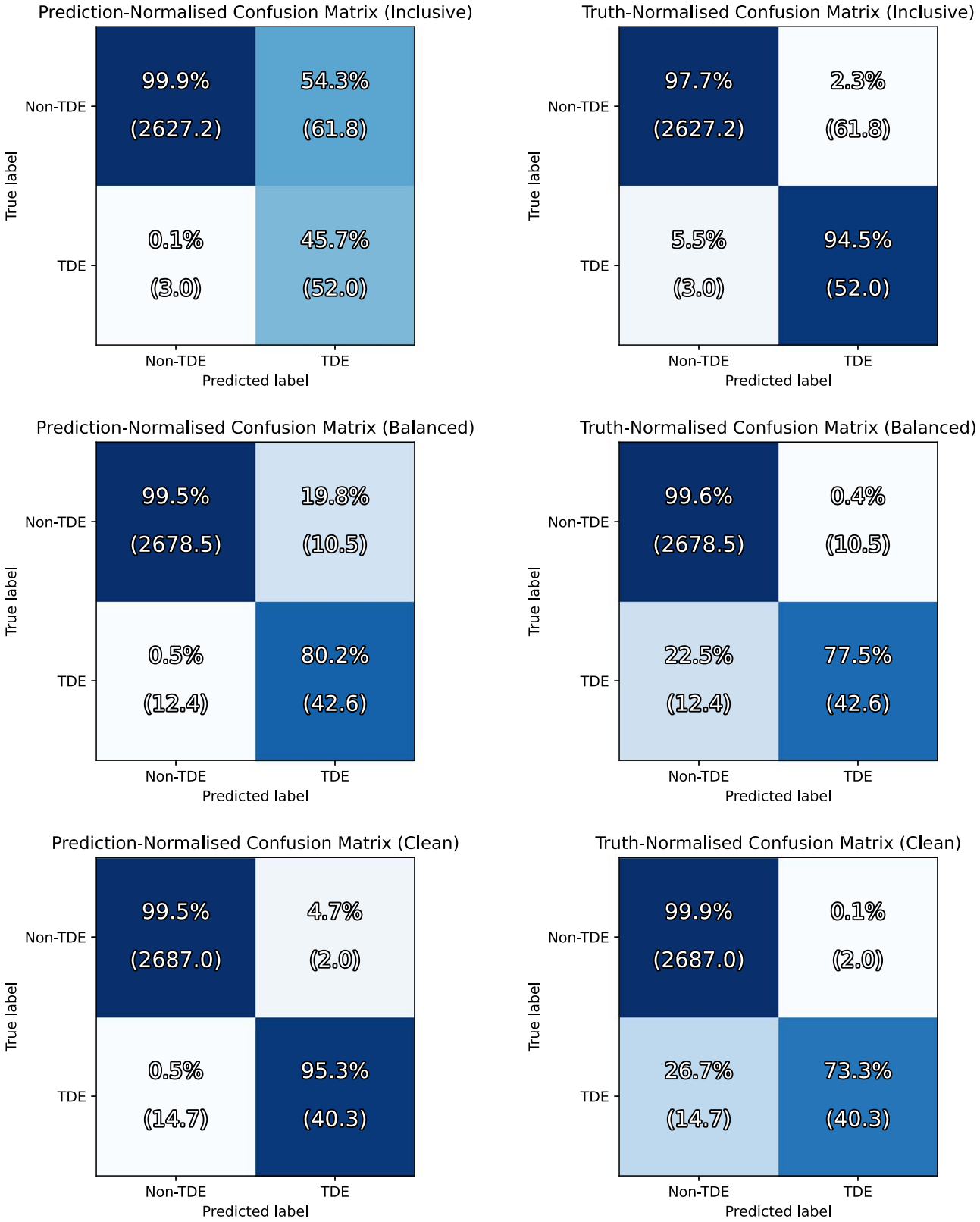

Figure 4. Prediction-normalized confusion matrices (left) and truth-normalized confusion matrices (right), showing the performance of `tdescore` on the real data for different thresholds. The data set is highly imbalanced, as seen in Figure 2. The source shuffling is performed ten times, yielding averaged performance across the iterations, with the average expected number of sources for each category shown in parentheses. Top: an inclusive threshold, optimized for recall. At the cost of 5% loss of TDEs, a sample is produced with a TDE fraction increased from ∼2% to ∼46%. Center: an intermediate threshold, chosen to achieve >80% precision. It achieves relatively high recall (77.5%). Bottom: a strict threshold, optimized for precision. >70% of TDEs pass this requirement, yielding a clean sample with <5% contamination rate.

Table 1
Relative Importance of all 27 Features in `tdescore`

Feature	Description	Importance (%)
w1_m_w2	WISE W1–W2 host color	32.4
peak_color	Color at g-band peak	16.0
has_milliquas	Has milliquas crossmatch?	9.3
color_grad	Rate of color change	7.8
sncosmo_chisq	sncosmo χ^2	5.9
sncosmo_c	sncosmo c parameter	4.6
fade	Fade from G.P.	3.8
det_cadence	Mean detection cadence	2.4
pre_ inflection	Number of prepeak inflections	2.3
distpsnr1	Distance to PS1 host	1.8
length_scale	Length scale from G.P.	1.6
y_scale	Y scale from G.P.	1.6
sncosmo_x1	sncosmo X1 parameter	1.6
w3_m_w4	WISE W3–W4 host color	1.2
post_inflection	Number of postpeak inflections	1.0
g-r_MeanPSFMag	PS1 host $g-r$ color	0.9
sumrat	"Sum ratio"	0.9
score	Score from G.P	0.8
classtar	`SourceExtractor` variable	0.8
positive_fraction	Fraction of positive detections	0.6
w1_chi2	WISE W1 χ^2	0.6
distnr	Pixel distance to nearest source	0.6
z-y_MeanPSFMag	PS1 host $z-y$ color	0.4
sncosmo_chi2pdof	sncosmo χ^2 per d.o.f	0.4
i-z_MeanPSFMag	PS1 host $i-z$ color	0.2
r-i_ MeanPSFMag	PS1 host $r-i$ color	0.1
sgscore1	Star/galaxy score for PS1 host	0.1

Note. Calculated by `XGBoost` (Chen & Guestrin 2016) using the standard averaging of importance across all decision trees in the final model (see, e.g., Hastie et al. 2009).

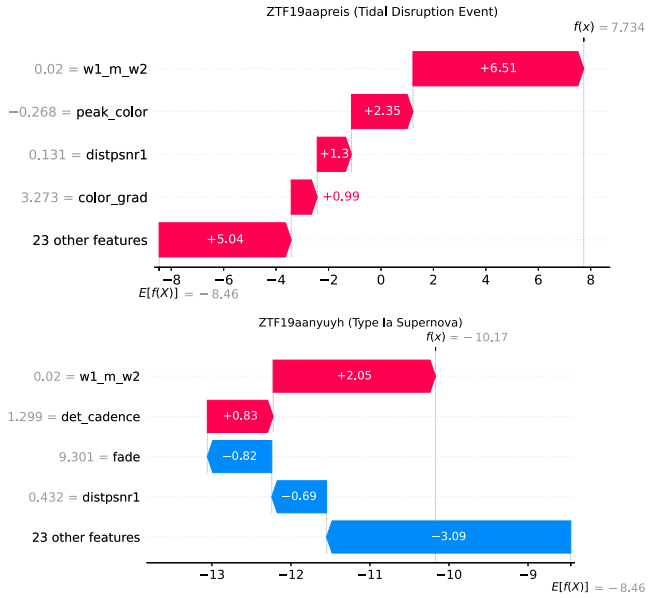

Figure 5. "Waterfall plots" produced by SHAP for a TDE (top) and an SN Ia (bottom), demonstrating the thinking behind the `tdescore` classifications. In both plots, red/right is more TDE-like, while blue/left is less TDE-like. The four most salient features for each source are shown, with the actual value for each parameter given in the leftmost column. Top: the TDE (ZTF19aapreis) has WISE colors inconsistent with an AGN host (W1–W2 = 0.0), has a blue color at peak ($g-r=-0.3$), is very nuclear ($0\farcs1$ offset to PS1), and has very little cooling (0.003 mag per day). All these variables lead to a TDE classification. Bottom: the SN (ZTF19aanyuyh) also has WISE colors inconsistent with an AGN host (W1–W2 = 0.0) and a high detection rate (one data point per 1.3 days), supporting a possible TDE classification. However, the source also fades very rapidly (9.3 days), and is somewhat offset from its PS1 host ($0\farcs4$). In combination, these other variables lead to a non-TDE classification. In both cases, the `tdescore` use of features closely approximates the reasoning that would be employed by an astronomer.

source. This means that every individual `tdescore` classification can readily be understood and sanity checked by humans. An illustration of `tdescore` reasoning for classifying a TDE and an SN Ia, are shown in Figure 5. In these cases, `tdescore` follows a decision-making process very similar to that employed by human scanners in ZTF.

6. Discussion and Conclusion

`tdescore` is a novel photometric classifier developed with the explicit aim of approximating the human scanning employed in ZTF. Our ZTF sample provides the largest homogeneous sample of nuclear transients by far (Yao et al. 2023) and thus presently serves as the best template for developing techniques to detect TDEs. `tdescore` combines well-tested algorithmic cuts to robustly identify nuclear transients, an agnostic light-curve analysis technique using Gaussian processes, and a simple binary tree-based classifier using physically motivated features.

The sole other dedicated TDE classifier in the literature, `fleet` (Gomez et al. 2023), is based on an adapted SN classifier. Gomez et al. (2023) began with a sample of spectroscopically classified transients from the Transient Name Server, rather than a dedicated sample of nuclear transients as presented here. In other respects Gomez et al. (2023) followed a similar procedure to the one presented here, with an imbalanced sample of transients that are first analyzed for light-curve and host galaxy properties, augmentation via SMOTE, and then performance assessment via k-fold cross validation. `fleet` achieved just ≈40% recall with ≈50% precision for a loose selection, or alternatively ≈30% recall with ≈80% precision for a stricter selection, in contrast to the ∼80% recall and ∼80% precision in the `tdescore` balanced case. However, the performance is not directly comparable because `fleet` was applied to only 40 days of photometry, rather than the full light-curve history employed here. For a TDE such as that in Figure 1, 40 days would be insufficient to adequately measure fade or color evolution. As detailed in the Appendix B, the performance of `tdescore` is closer to `fleet` if late-time data is ignored.

Looking further ahead, `tdescore` can serve as a template for obtaining a photometrically selected sample of TDEs from surveys such as the Legacy Survey of Space and Time (LSST) with the Vera C. Rubin Observatory (Ivezić et al. 2019). In combination with photometric redshifts, an ML-based approach like `tdescore` could enable us to perform large-sample TDE demographic studies for the first time without use of any spectroscopic observations. In particular, Bricman & Gomboc (2020) estimated that LSST should detect >3000 TDEs per year, under the assumption of a conservative detection requirement of 2 mag above the median 5σ limit. Pushing 1 magnitude deeper, to match the cuts employed by this work, would increase this number even further. The performance of `tdescore` suggests such a depth would be plausible using

photometric selection, with the slow evolution of TDEs being well suited to the expected LSST cadence.

The performance of `tdescore` with real-time ZTF data will be the subject of a future publication. There are many other possible uses of photometrically selected TDEs, for example, to build a much larger sample of probable TDEs to test possible multimessenger correlations between neutrinos and TDEs (see, e.g., Stein 2019), for which there is growing evidence (Stein et al. 2021; Reusch et al. 2022; Jiang et al. 2023; van Velzen et al. 2024). Another use is to quickly identify candidate TDEs among transients detected by surveys at other wavelengths through crossmatching to probable ZTF TDEs found by `tdescore`. We will use this method to aid searches for dust-obscured TDEs with the Wide-Field Infrared Transient Explorer (Lourie et al. 2020), a newly commissioned near-infrared survey telescope at Palomar Observatory.

Building broader TDE samples is important because by construction, `tdescore` will not find TDEs that differ substantially from the existing ZTF TDE sample. In particular, given the importance of the W1–W2 color, `tdescore` is likely to be heavily biased against finding TDEs in AGN. This is a direct consequence of the parent sample of ZTF TDEs, none of which occur in AGN-like hosts with W1–W2 > 0.7 (Stern et al. 2012). To find such "AGN−TDEs" (or other outliers such as red TDEs or fast TDEs), we would first require a handful of spectroscopically confirmed ZTF examples. As our understanding of TDE diversity improves, `tdescore` can be retrained to find a broader selection of TDEs.

Applying `tdescore` directly to future optical surveys should be relatively straightforward because the classifier is trained almost exclusively on light-curve features that are generic and do not encode any specific ZTF survey information. However, there is also substantial scope for improvement in performance. While all ZTF light curves were analyzed here in observer frame units, with no correction for redshift, ongoing industrial spectroscopic surveys such as DESI (DESI Collaboration et al. 2016) mean that spectroscopic redshifts will be available systematically for much of the local Universe. Even in the LSST/Rubin era, widespread adoption of photometric redshifts would enable intrinsic rest-frame properties such as peak luminosity to be employed for classification. Additionally, TDEs are generally characterized by luminous UV emission, and u-band color is an excellent discriminator to find TDEs (see, e.g., van Velzen et al. 2011). While no UV observations were used for `tdescore`, due to a lack of systematic coverage, Rubin will have u-band coverage of all transients on a ∼weekly cadence. At higher redshifts, much of the TDE rest-frame emission at UV wavelengths will also be detectable with optical LSST filters. There are thus many reasons to be optimistic that future iterations of `tdescore` will be able to outperform the classifier presented here.

Acknowledgments

We thank Adam Stein, Ludwig Rauch, and Niharika Sravan for fruitful discussions about machine-learning classification.

R.S. and M.M.K acknowledge support from grants by the National Science Foundation (AST 2206730) and the David and Lucille Packard Foundation (PI Kasliwal). M.N. is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 948381) and by UK Space Agency grant No. ST/Y000692/1. E.K.H. acknowledges support by NASA under award number 80GSFC21M0002. S. J.N. is supported by the US National Science Foundation (NSF) through grant AST-2108402.

Based on observations obtained with the Samuel Oschin Telescope 48-inch and the 60-inch Telescope at the Palomar Observatory as part of the Zwicky Transient Facility project. ZTF is supported by the National Science Foundation under grants Nos. AST-1440341 and AST-2034437 and a collaboration including current partners Caltech, IPAC, the Weizmann Institute of Science, the Oskar Klein Center at Stockholm University, the University of Maryland, Deutsches Elektronen-Synchrotron and Humboldt University, the TANGO Consortium of Taiwan, the University of Wisconsin at Milwaukee, Trinity College Dublin, Lawrence Livermore National Laboratories, IN2P3, University of Warwick, Ruhr University Bochum, Northwestern University and former partners the University of Washington, Los Alamos National Laboratories, and Lawrence Berkeley National Laboratories. Operations are conducted by COO, IPAC, and UW. SED Machine is based upon work supported by the National Science Foundation under grant No. 1106171. The Gordon and Betty Moore Foundation, through both the Data-Driven Investigator Program and a dedicated grant, provided critical funding for SkyPortal.

Facility: PO:1.2m (ZTF; Bellm et al. 2019), PO:1.5m (SEDm; Blagorodnova et al. 2018; Rigault et al. 2019; Kim et al. 2022).

Software: `AMPEL` (Nordin et al. 2019), `astropy` (Astropy Collaboration et al. 2022), astroquery (Ginsburg et al. 2019), `numpy` (Harris et al. 2020), `pandas` (Wes McKinney 2010), `scikit-learn` (Pedregosa et al. 2011), `scipy` (Virtanen et al. 2020), `SHAP` (Lundberg & Lee 2017), `sncosmo` (Barbary et al. 2016), `tdescore` (Stein 2024), `XGBoost` (Chen & Guestrin 2016).

Appendix A
Classification of ZTF Nuclear Transients

With the 11,699 transients in our sample, we employ a two-step process to assign classifications to each source. We extract available classifications for these transients from the `ZTF Fritz Marshal` (van der Walt et al. 2019; Coughlin et al. 2023) and the predecessor `ZTF GROWTH Marshal` (Kasliwal et al. 2019). However, these classifications have been assigned by human scanners in a variety of heterogeneous ways. While our sample of TDEs has been thoroughly vetted, we cannot vet the justification and reliability of each individual classification assigned for the >5000 non-TDEs in this way. Instead, we independently classify this sample of nuclear sources using objective contextual data, and cross-check to confirm that the classifications agree with the human-assigned labels.

We follow a hierarchical approach for the non-TDEs, and first consider classifications based directly on spectra of transients:

1. *Transient Name Server (TNS) Spectra*. Transients that are classified on TNS require an accompanying spectrum. We assume that these TNS classifications are accurate. A total of 953 transients have a TNS classification.
2. *ZTF Spectra*. We assume that internal ZTF classifications are reliable if at least one spectrum of the source is available. A total of 328 transients are not classified on TNS but do have an internal ZTF classification and spectrum.

We further consider sources that have archival spectroscopy confirming a variable host:

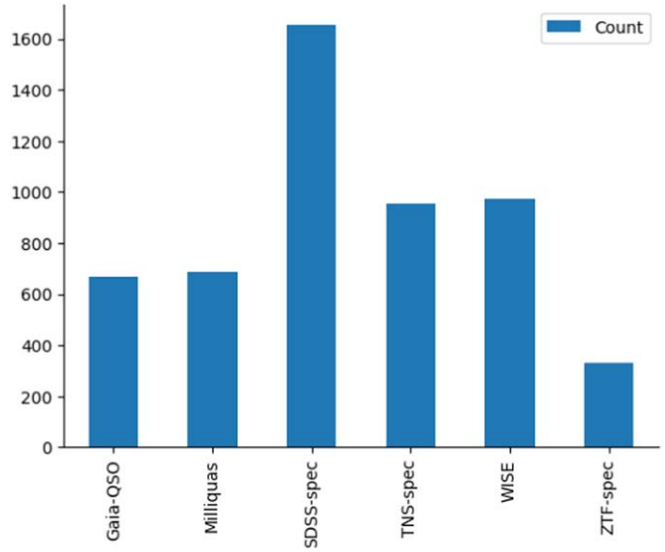

Figure 6. Breakdown of the validation method for classifications, as described in Section A. Each source requires both a human-assigned classification and a second piece of confirmatory evidence to be considered reliably classified.

1. *SDSS Spectra*. We crossmatch every source to the catalog of SDSS spectra (York et al. 2000). We assume that the classifications are reliable if a human scanner has classified a source as variable and the source also has a spectroscopic classification confirming it as variable (AGN or stellar) from SDSS. There are 1653 such sources.

Finally, we consider sources that are very likely to be variables:

1. *Milliquas AGN*. We consider a classification to be reliable if a source is listed in of the "Million Quasar Catalog" (Milliquas V8) of known AGN (Flesch 2023) and has been also classified by a human as an AGN. Six hundred eighty-eight sources meet this criteria.
2. *WISE AGN*. We consider sources with very AGN-like WISE colors (W1–W2) > 0.8 to be reliably classified (Stern et al. 2012). There are 973 such sources classified via WISE colors.
3. *Gaia QSOs*. We consider Gaia-Data Release 3 cataloged QSOs (Gaia Collaboration et al. 2023) as AGN if this agrees with a human classification. Six hundred sixty-nine sources meet this criteria.

After following this procedure, a total of 5264 sources have a verified classification. We omit the remaining ambiguous sources, and treat these 5264 verified-classification sources as our final sample of "classified sources." The breakdown in classification origin is given in Figure 6.

Appendix B
Performance of `tdescore` with Different Parameter Subsets

We tested the performance of `tdescore` using subsets of parameters listed in Table 2. For consistency, we measure performance on the same 2744 sources for which all information is available. The parameters were grouped into "Host Only," "Early," "At Peak," and "Full," where "Full" corresponds to the complete parameter set described in Section 3. The parameter sets were cumulative, with, e.g., the "Early" set including all "Host" parameters to provide an estimate of performance over time as available data increased.

Table 2
Performance of `tdescore` for Four Parameter Sets

Parameter Set	New Parameters	Total Parameters	ROC Area	Precision/ Recall Area
Host Only	sgscore1 w1_m_w2 w3_m_w4 w1_chi2 has_milliquas g-r_MeanPSFMag r-i_MeanPSFMag i-z_MeanPSFMag z-y_MeanPSFMag	9	0.91	0.20
Early	distpsnr1 classtar sumrat distnr	13	0.95	0.30
At Peak	peak_color pre_inflection positive_fraction det_cadence y_scale	18	0.96	0.63
Full	color_grad fade length_scale post_inflection score sncosmo_chisq sncosmo_chi2pdof sncosmo_x1 sncosmo_c	27	0.99	0.89

Note. Information only about the host, information available shortly after discovery, information available by the time of peak, and the full parameter set. The performance of `tdescore` improves substantially with more data, but high performance is only achieved for the full data set.

We measure the performance of `tdescore` using each of the four data sets, with the precision/recall area and "receiver operating characteristic curve" (ROC) area listed in Table 2. We also show the full precision and recall of each classifier in Figure 7. As expected, performance improves with increased parameter numbers. However, we note that using only data available after first detections, a recall of $\sim$50% and precision of $\sim$30% is still achieved. `tdescore` can thus identify candidate TDEs with a moderate false-positive rate at this stage, making triggered follow-up of "infant TDEs" feasible.

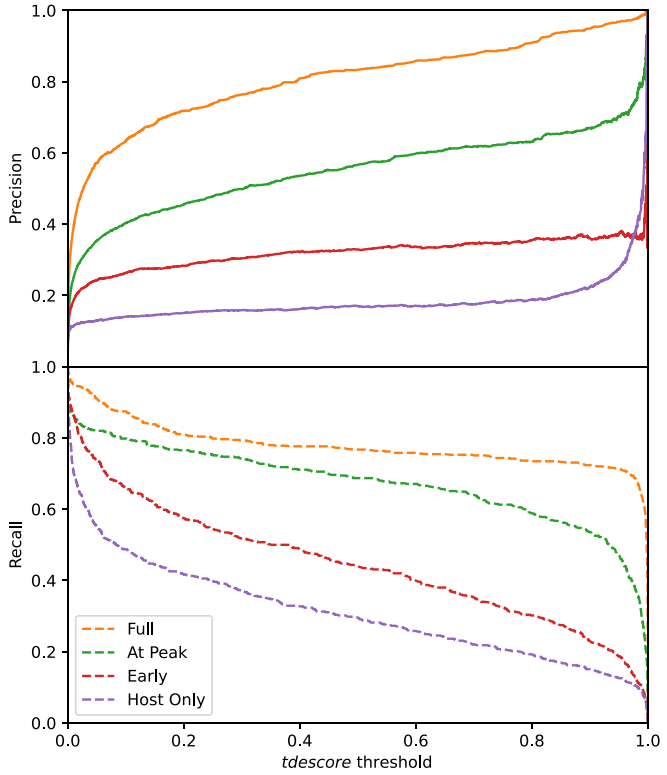

Figure 7. Precision (top, solid) and recall (bottom, dashed) curves for the four parameter sets listed in Table 2, as a function of threshold. Both precision and recall increase substantially as more data are added.

ORCID iDs

Robert Stein https://orcid.org/0000-0003-2434-0387
Ashish Mahabal https://orcid.org/0000-0003-2242-0244
Simeon Reusch https://orcid.org/0000-0002-7788-628X
Matthew Graham https://orcid.org/0000-0002-3168-0139
Mansi M. Kasliwal https://orcid.org/0000-0002-5619-4938
Marek Kowalski https://orcid.org/0000-0001-8594-8666
Suvi Gezari https://orcid.org/0000-0003-3703-5154
Erica Hammerstein https://orcid.org/0000-0002-5698-8703
Matt Nicholl https://orcid.org/0000-0002-2555-3192
Jesper Sollerman https://orcid.org/0000-0003-1546-6615
Sjoert van Velzen https://orcid.org/0000-0002-3859-8074
Yuhan Yao https://orcid.org/0000-0001-6747-8509
Russ R. Laher https://orcid.org/0000-0003-2451-5482
Ben Rusholme https://orcid.org/0000-0001-7648-4142

References

Aigrain, S., & Foreman-Mackey, D. 2023, ARA&A, 61, 329
Arcavi, I., Gal-Yam, A., Sullivan, M., et al. 2014, ApJ, 793, 38
Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167
Barbary, K., Barclay, T., Biswas, R., et al. 2016, SNCosmo: Python Library for Supernova Cosmology, Astrophysics Source Code Library, ascl:1611.017
Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Blagorodnova, N., Neill, J. D., Walters, R., et al. 2018, PASP, 130, 035003
Bricman, K., & Gomboc, A. 2020, ApJ, 890, 73
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2002, Journal of Artificial Intelligence Research, 16, 321
Chen, T., & Guestrin, C. 2016, in Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, KDD'16 (New York: Association for Computing Machinery), 785
Coughlin, M. W., Bloom, J. S., Nir, G., et al. 2023, ApJS, 267, 31
Dekany, R., Smith, R. M., Riddle, R., et al. 2020, PASP, 132, 038001
DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
Duev, D. A., Mahabal, A., Masci, F. J., et al. 2019, MNRAS, 489, 3582
Fitzpatrick, E. L. 1999, PASP, 111, 63
Flesch, E. W. 2023, OJAp, 6, 49
French, K. D., Arcavi, I., & Zabludoff, A. 2016, ApJL, 818, L21
Gaia Collaboration, Bailer-Jones, C. A. L., Teyssier, D., et al. 2023, A&A, 674, A41
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gezari, S. 2021, ARA&A, 59, 21
Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, AJ, 157, 98
Gomez, S., Villar, V. A., Berger, E., et al. 2023, ApJ, 949, 113
Graham, M. J., Kulkarni, S. R., Bellm, E. C., et al. 2019, PASP, 131, 078001
Graham, M. J., McKernan, B., Ford, K. E. S., et al. 2023, ApJ, 942, 99
Graur, O., French, K. D., Zahid, H. J., et al. 2018, ApJ, 853, 39
Guy, J., Astier, P., Baumont, S., et al. 2007, A&A, 466, 11
Hammerstein, E., Gezari, S., van Velzen, S., et al. 2021a, ApJL, 908, L20
Hammerstein, E., Gezari, S., Velzen, S. V., et al. 2021b, TNSCR, 2021-159, 1
Hammerstein, E., van Velzen, S., Gezari, S., et al. 2023, ApJ, 942, 9
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Hastie, T., Tibshirani, R., & Friedman, J. 2009, The Elements of Statistical Learning, Data Mining, Inference, and Prediction (Berlin: Springer)
Hodgkin, S. T., Breedt, E., Delgado, A., et al. 2020, TNSTR, 2020-3089, 1
Hung, T., Gezari, S., Cenko, S. B., et al. 2018, ApJS, 238, 15
Ivezić, ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Jiang, N., Zhou, Z., Zhu, J., Wang, Y., & Wang, T. 2023, ApJL, 953, L12
Kasliwal, M. M., Cannella, C., Bagdasaryan, A., et al. 2019, PASP, 131, 038003
Kim, Y. L., Rigault, M., Neill, J. D., et al. 2022, PASP, 134, 024505
Kulkarni, S. R. 2013, ATel, 4807, 1
Lourie, N. P., Baker, J. W., Burruss, R. S., et al. 2020, Proc. SPIE, 11447, 114479K
Lundberg, S. M., & Lee, S.-I. 2017, in 31st Conf. on Neural Information Processing Systems (NIPS 2017), Advances in Neural Information Processing Systems 30, ed. I. Guyon et al. (Red Hook, NY: Curran Associates, Inc.), http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf
Mahabal, A., Rebbapragada, U., Walters, R., et al. 2019, PASP, 131, 038002
Mainzer, A., Bauer, J., Cutri, R. M., et al. 2014, ApJ, 792, 30
Masci, F. J., Laher, R. R., Rusholme, B., et al. 2019, PASP, 131, 018003
Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, PASP, 131, 118002
Nordin, J., Brinnel, V., van Santen, J., et al. 2019, A&A, 631, A147
Patterson, M. T., Bellm, E. C., Rusholme, B., et al. 2019, PASP, 131, 018001
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Rees, M. J. 1988, Natur, 333, 523
Reusch, S., Stein, R., Kowalski, M., et al. 2022, PhRvL, 128, 221101
Rigault, M., Neill, J. D., Blagorodnova, N., et al. 2019, A&A, 627, A115
Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103
Shvartzvald, Y., Waxman, E., Gal-Yam, A., et al. 2024, ApJ, 964, 74
Stein, R. 2019, ICRC, 36, 1016
Stein, R., van Velzen, S., Kowalski, M., et al. 2021, NatAs, 5, 510
Stein, R. D. 2024, robertdstein/tdescore: v1.0.0, Zenodo, doi:10.5281/zenodo.10784037
Stern, D., Assef, R. J., Benford, D. J., et al. 2012, ApJ, 753, 30
Tachibana, Y., & Miller, A. A. 2018, PASP, 130, 128001
van der Walt, S. J., Crellin-Quick, A., & Bloom, J. S. 2019, JOSS, 4, 1247
van Velzen, S., Farrar, G. R., Gezari, S., et al. 2011, ApJ, 741, 73
van Velzen, S., Gezari, S., Cenko, S. B., et al. 2019, ApJ, 872, 198
van Velzen, S., Gezari, S., Hammerstein, E., et al. 2021, ApJ, 908, 4
van Velzen, S., Stein, R., Gilfanov, M., et al. 2024, MNRAS, 529, 2559
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Wes McKinney 2010, in Proc. of the 9th Python in Science Conf., ed. S. van der Walt & J. Millman, 61
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Yao, Y., Ravi, V., Gezari, S., et al. 2023, ApJL, 955, L6
York, D. G., Adelman, J., Anderson, J. E. J., et al. 2000, AJ, 120, 1579