
Phenomenological classification of the Zwicky Transient Facility astronomical event alerts

Dmitry A. Duev*

Weights & Biases, Inc.
1479 Folsom St, San Francisco, CA 94103

Division of Physics, Mathematics, and Astronomy
California Institute of Technology
Pasadena, CA 91125
duev@caltech.edu

Stéfan J. van der Walt

Berkeley Institute for Data Science
University of California, Berkeley
Berkeley, CA 94720, USA
stefanv@berkeley.edu

Abstract

The Zwicky Transient Facility (ZTF), a state-of-the-art optical robotic sky survey, registers on the order of a million transient events — such as supernova explosions, changes in brightness of variable sources, or moving object detections — every clear night, and generates associated real-time alerts. We present Alert-Classifying Artificial Intelligence (ACAI), an open-source deep-learning framework for the phenomenological classification of ZTF alerts. ACAI uses a set of five binary classifiers to characterize objects which, in combination with the auxiliary/contextual event information available from alert brokers, provides a powerful tool for alert stream filtering tailored to different science cases, including early identification of supernova-like and anomalous transient events. We report on the performance of ACAI during the first months of deployment in a production setting.

1 Introduction

Astronomical sky surveys observe a myriad of transient events in the dynamic sky originating from a wide range of astrophysical objects and processes. When detection of such events is performed in the image domain, an epochal image of a patch of the sky is compared to a reference image, usually adopting an image subtraction algorithm.

Current and future large-scale astronomical sky surveys have the ability to detect millions of transient events, manifesting the need for automated separation of genuine astrophysical events from bogus detections and further categorization of real events. Events, real and bogus, are caused by a wide variety of phenomena, some of which are very hard to model.

To address this, sky surveys typically rely on binary real/bogus (RB) classifiers [2, 1, 5, 6, 20, 18, 12, 21, 7, 19, 11] or multi-class classifiers [8].

1.1 The Zwicky Transient Facility

The Zwicky Transient Facility (ZTF)² is a state-of-the-art robotic time-domain sky survey capable of visiting the entire visible sky north of -30° declination every night. ZTF observes the sky in the g , r , and i bands at different cadences depending on the scientific program and sky region [3, 13]. The 576 megapixel camera with a 47 deg^2 field of view, installed on the Samuel Oschin 48-inch (1.2-m)

*<https://duev.space>

²<https://ztf.caltech.edu>

Schmidt Telescope, can scan more than 3750 deg^2 per hour, to a 5σ detection limit of 20.7 mag in the r band with a 30-second exposure during new moon [17, 10].

The raw data are processed in real time at IPAC, the California Institute of Technology (Caltech). The ZTF Science Data System (ZSDS) housed at IPAC consists of data processing pipelines, data archives, infrastructure for long-term curation, and services for data retrieval and visualization [17].

The part of ZSDS responsible for transient event detection and extraction performs image differencing of the science and reference images using the ZOGY algorithm [22]. If the resulting difference image is of sufficient quality, the pipeline then detects events from the point-source match-filtered images, performing detection on both the positive (science minus reference) and negative (reference minus science) images. Events are extracted with both aperture and point spread function (PSF) fit photometry, and additional source features are computed. The events are then lightly filtered to remove obvious false positives and image cutouts are generated [17].

Events may have been triggered by a transient, a variable, or a moving object; or caused by, e.g., an optical or processing artifact. Metadata and contextual information, including image cutouts, are put into alert packets in the Apache AVRO format that are then distributed by the ZTF Alert Distribution System (ZADS) using Apache Kafka. The number of events detected nightly typically ranges from $10^5 - 10^6$.

2 ACAI: a deep learning framework for the phenomenological classification of ZTF alerts

In this work, we present Alert-Classifying Artificial Intelligence (ACAI), an open-source deep-learning framework for the classification of ZTF alerts³. ACAI uses a set of *independent* binary classifiers to categorize objects into five phenomenological classes:

- “Hosted” (`acai_h`) – genuine transients in the vicinity of a (host) galaxy with detectable morphology. Objects such as supernovae (SNe), novae, or cataclysmic variables occurring near galaxy-like objects get high `acai_h` scores.
- “Orphan” (`acai_o`) – genuine orphan transients, i.e. when there are no identifiable “hosts” in their vicinity. This category catches asteroids and host-less (or with hosts that are too faint) transients.
- “Nuclear” (`acai_n`) – genuine transients occurring in the galaxy/quasar nucleus. Active galactic nuclei (AGN) and tidal disruption events (TDEs) fall into this category.
- “Variable star” (`acai_v`) – genuine brightness changes of variable stars detectable in the reference image or in a reference catalog.
- “Bogus” (`acai_b`) – bogus events such as cosmic ray hits, optical reflections, or data processing artifacts, get high `acai_b` scores.

The main advantage of our approach with independent binary classifiers is significantly greater flexibility compared to typically used multi-class classifiers, where an object is assumed to have a single correct label of many, or multi-label classifiers, where a single system outputs probabilistic predictions of object class membership for multiple classes at once. If the performance on a particular class is deemed insufficient, retraining the classifier with new training data (or employing a different architecture) does not affect the system’s performance on other classes. Adding new categories is straightforward and does not require rebuilding an entire multi-class or multi-label classifier.

Another advantage is more flexibility for the end-user. When used in the context of an alert broker, the ACAI scores can be mixed and matched with other criteria such as the information on cross-matches with external catalogs to create alert stream filters tailored to a particular science case.

Finally, our approach implicitly allows for anomaly detection. For example, when an object is classified as real but does not belong to any of the non-bogus categories with high confidence, or if it simultaneously receives high scores in multiple categories, it is potentially of interest.

³<https://github.com/dmitryduev/acai>

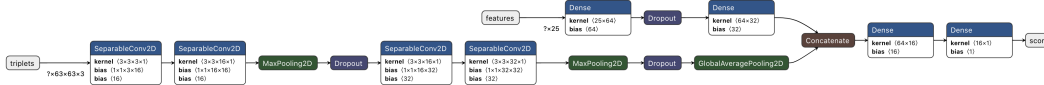


Figure 1: The architecture of the ACAI models.

2.1 Data set

In the case of alert classification, the training set must reflect the possible variations across different seeing conditions, filters, sky location, sensors, and include data artifacts caused by, for example, cross-talk or telescope reflections.

We have assembled a large training data set consisting of over 200,000 individual alerts – about 70k alerts from known variable stars of different types, 35k from nuclear transients originating from known AGN and TDEs, 30k “hosted” transients from known supernovae and novae, 50k bogus alerts, and 30k orphan events, mostly from known Solar system objects.

We employed an active-learning-like approach of alternating between data labeling and classifier training with subsequent sampling of their predictions, both of confident ones and, more importantly, of those near the decision boundary and with the highest loss.

2.2 Deep neural network architecture and training

When building the classifiers, we had to explore a vast hyperparameter space.

As input, our classifiers use 25 features from the “candidate” section of the alert packets⁴ that include both derived characteristics/parameters of events as well as contextual information, such as cross-matches with external catalogs and full-sized (63x63 pixels) triplet (epochal science, reference, and difference) image cutouts.

The classifiers were implemented using the TensorFlow library and its high-level Keras API [16, 9]. We employed the standard binary cross-entropy loss function, the Adam optimizer [15], a batch size of 32, and a 81%/9%/10% training/validation/test data split. Given the amount and diversity of gathered training data, we did not perform data augmentation.

The input features were normalized; the same norms were used for all classifiers. We applied class balancing; the classifier performance was subsequently checked on the originally dropped “negative” examples and the small number of misclassifications (typically on the order of 1 – 2%) were added to the training set. The training data were weighted per class. The class weights were further adjusted to balance precision (purity) and recall (completeness). We used standard techniques to achieve high performance, such as learning rate reduction on a plateau and early stopping based on validation loss.

Initially we used a simple architecture that demonstrated satisfying performance, with a minimal, arbitrarily-chosen number of fully-connected and convolutional layers. As we expanded the input data sets, we ran several rounds of hyperparameter tuning using the Weights & Biases’ sweeps⁵ library [4]. We tuned hyperparameters such as the number of fully-connected layers and neurons therein, number of convolutional filters and their sizes and types (regular or separable convolution), flattening versus global average pooling, dropout rates, activation functions, and the initial learning rate for the Adam optimizer. We also experimented with the inclusion/exclusion of particular alert features in/from the input.

We chose the simplest of the best-performing architectures (shown in Fig. 1) and adopted it for all of the individual classifiers.

Training (including grid hyper-parameter search) was performed on a single Nvidia Tesla V100 16GB GPU on the Google Cloud Platform, and took just under 1 day.

⁴<https://zwicktransientfacility.github.io/ztf-avro-alert>

⁵<https://github.com/wandb/sweeps>

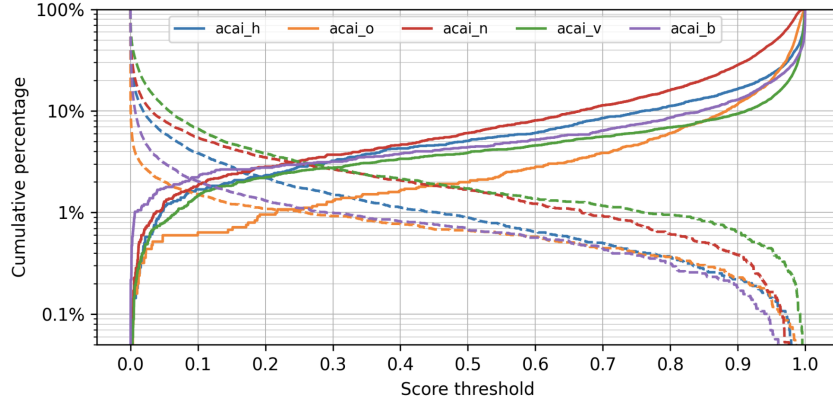


Figure 2: False negative (FNR; solid lines) and false positive rates (FPR; dashed lines) of the ACAI models as functions of the score threshold. The classifiers were evaluated on the test set consisting of $\sim 20,000$ alerts.

3 ACAI performance

3.1 Test set performance

When evaluated on the test sets, all of the ACAI classifiers demonstrate 1-3% balanced false positive / false negative rates for score thresholds around 0.2 (see Fig. 2).

3.2 Performance in production

The ACAI classifiers have been deployed on Fritz⁶, an open-source data platform and alert broker used by the ZTF collaboration (consisting of hundreds of astronomers worldwide), in a production setting since early December 2020. When used in combination with auxiliary/contextual event information available from the Fritz’s broker, Kowalski⁷, ACAI has proven to be a powerful tool (with configurable purity/completeness based on the score thresholds) for alert stream filtering tailored to different science cases, yielding numerous discoveries. These include early detection of SN-like events (often after only a single alert), identification of tidal disruption events, identification of unknown Solar system objects, and more. Finally, we have been able to identify multiple astrophysically-interesting anomalous objects, detailed investigation of which is currently underway.

3.2.1 Early detection of “hosted” SNe

As an demonstration, we have been running a simple alert stream filter that selects events with high *acai_h* scores and low scores from other ACAI classifiers, detected fewer than 5 times as a positive source in the difference image, and not originating from a know Solar system object. This can be expressed in pseudo code as follows:

```
acai_h > 0.8 AND MAX(acai_b, acai_n, acai_v, acai_o) < 0.1
AND n_detections < 5 AND is_positive_subtraction AND NOT is_known_ss_object
```

From January through September 2021, this filter has yielded 12,035 objects, of which only 52 have been marked as likely bogus (0.4% FPR). The majority of the selected objects exhibit SN-like light curves, at peak fainter than $\sim 19.5 - 20$ mag.

4 Limitations

We note that the training sets for the different classifiers overlap, so while the classifiers are executed independently, their output is expected to be correlated. Furthermore, ACAI is designed to aid in

⁶<https://github.com/fritz-marshal/fritz>

⁷<https://github.com/dmitryduev/kowalski>

(early) discovery of interesting events and is intended to be used alongside other information. More work is required to characterize completeness of samples selected with the use of ACAI.

5 Conclusions

We have demonstrated that, by putting together a large, representative, and uncontaminated data set with a relatively simple deep model, we can achieve excellent classification performance. To improve it even further, we will retrain and deploy new classifiers as more labeled data are collected. Our setup will be useful for other surveys including the Vera Rubin Observatory’s Legacy Survey of Space and Time (LSST, [14]) in the near future.

Acknowledgments and Disclosure of Funding

D. A. Duev would like to thank Ivan Duev for assistance with data labeling.

D. A. Duev acknowledges support from Google Cloud and from the Heising-Simons Foundation under Grant No. 12540303.

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 Grant No. AST-1440341 and a collaboration including Caltech, IPAC, the Weizmann Institute for Science, the Oskar Klein Center at Stockholm University, the University of Maryland, the University of Washington, Deutsches Elektronen-Synchrotron and Humboldt University, Los Alamos National Laboratories, the TANGO Consortium of Taiwan, the University of Wisconsin at Milwaukee, and Lawrence Berkeley National Laboratories. Operations are conducted by COO, IPAC, and UW.

References

- [1] G. Aldering et al. “Overview of the Nearby Supernova Factory”. In: *Survey and Other Telescope Technologies and Discoveries*. Ed. by J. A. Tyson and S. Wolff. Vol. 4836. SPIE. Dec. 2002, pp. 61–72. DOI: 10.1117/12.458107.
- [2] S. Bailey et al. “How to Find More Supernovae with Less Work: Object Classification Techniques for Difference Imaging”. In: *ApJ* 665 (Aug. 2007), pp. 1246–1253. DOI: 10.1086/519832. arXiv: 0705.0493.
- [3] E. C. Bellm et al. “The Zwicky Transient Facility: System Overview, Performance, and First Results”. In: *PASP* 131.995 (Jan. 2019), p. 018002. DOI: 10.1088/1538-3873/aaecbe. arXiv: 1902.01932 [astro-ph. IM].
- [4] L. Biewald. *Experiment Tracking with Weights and Biases*. Software available from wandb.com. 2020. URL: <https://www.wandb.com/>.
- [5] J. Bloom et al. “Towards a real-time transient classification engine”. In: *Astronomische Nachrichten* 329.3 (Mar. 2008), pp. 284–287. ISSN: 00046337. DOI: 10.1002/asna.200710957. URL: <http://onlinelibrary.wiley.com/doi/10.1002/asna.200710957/pdf> (visited on 03/03/2011).
- [6] H. Brink et al. “Using Machine Learning for Discovery in Synoptic Survey Imaging”. In: *arXiv:1209.3775* (Sept. 2012). URL: <http://arxiv.org/abs/1209.3775> (visited on 10/22/2012).
- [7] G. Cabrera-Vives et al. “Deep-HiTS: Rotation Invariant Convolutional Neural Network for Transient Detection”. In: *ApJ* 836.1, 97 (Feb. 2017), p. 97. DOI: 10.3847/1538-4357/836/1/97. arXiv: 1701.00458 [astro-ph. IM].
- [8] R. Carrasco-Davis et al. “Alert Classification for the ALeRCE Broker System: The Real-time Stamp Classifier”. In: *arXiv e-prints*, arXiv:2008.03309 (Aug. 2020), arXiv:2008.03309. arXiv: 2008.03309 [astro-ph. IM].
- [9] F. Chollet et al. *Keras*. <https://keras.io>. 2015.
- [10] R. Dekany et al. “The Zwicky Transient Facility: Observing System”. In: *PASP* 132.1009, 038001 (Mar. 2020), p. 038001. DOI: 10.1088/1538-3873/ab4ca2. arXiv: 2008.04923 [astro-ph. IM].

- [11] D. A. Duev et al. “Real-bogus classification for the Zwicky Transient Facility using deep learning”. In: *MNRAS* 489.3 (Nov. 2019), pp. 3582–3590. DOI: 10.1093/mnras/stz2357. arXiv: 1907.11259 [astro-ph.IM].
- [12] D. A. Goldstein et al. “Automated Transient Identification in the Dark Energy Survey”. In: *AJ* 150, 82 (Sept. 2015), p. 82. DOI: 10.1088/0004-6256/150/3/82. arXiv: 1504.02936 [astro-ph.IM].
- [13] M. J. Graham et al. “The Zwicky Transient Facility: Science Objectives”. In: *PASP* 131.1001 (July 2019), p. 078001. DOI: 10.1088/1538-3873/ab006c. arXiv: 1902.01945 [astro-ph.IM].
- [14] Z. Ivezić et al. “LSST: from science drivers to reference design and anticipated data products”. In: (2008). arXiv: 0805.2366v4.
- [15] D. P. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. In: *arXiv e-prints*, arXiv:1412.6980 (Dec. 2014), arXiv:1412.6980. arXiv: 1412.6980 [cs.LG].
- [16] Martin Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org, 2015. URL: <https://www.tensorflow.org/>.
- [17] F. J. Masci et al. “The Zwicky Transient Facility: Data Processing, Products, and Archive”. In: *PASP* 131.995 (Jan. 2019), p. 018003. DOI: 10.1088/1538-3873/aae8ac. arXiv: 1902.01872 [astro-ph.IM].
- [18] U. Rebbapragada, B. Bue, and P. R. Wozniak. In: *Time-domain Surveys and Data Shift: Case Study at the intermediate Palomar Transient Factory*. Vol. 225. Jan. 2015, p. 434.02. URL: <http://adsabs.harvard.edu/abs/2015AAS...22543402R> (visited on 04/20/2016).
- [19] E. Reyes et al. “Enhanced Rotational Invariant Convolutional Neural Network for Supernovae Detection”. In: *arXiv e-prints*, arXiv:1808.03626 (Aug. 2018), arXiv:1808.03626. arXiv: 1808.03626 [astro-ph.IM].
- [20] P. R. Wozniak et al. In: *Automated Variability Selection in Time-domain Imaging Surveys Using Sparse Representations with Learned Dictionaries*. Vol. 221. Jan. 2013, p. 431.05. URL: <http://adsabs.harvard.edu/abs/2013AAS...22143105W> (visited on 04/20/2016).
- [21] D. E. Wright et al. “Machine learning for transient discovery in Pan-STARRS1 difference imaging”. In: *MNRAS* 449 (May 2015), pp. 451–466. DOI: 10.1093/mnras/stv292. arXiv: 1501.05470 [astro-ph.IM].
- [22] B. Zackay, E. O. Ofek, and A. Gal-Yam. “Proper Image Subtraction—Optimal Transient Detection, Photometry, and Hypothesis Testing”. In: *ApJ* 830, 27 (Oct. 2016), p. 27. DOI: 10.3847/0004-637X/830/1/27. arXiv: 1601.02655 [astro-ph.IM].

Checklist

1. For all authors...
 - (a) Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope? [Yes]
 - (b) Did you describe the limitations of your work? [Yes]
 - (c) Did you discuss any potential negative societal impacts of your work? [N/A] Could not think of any.
 - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
 - (a) Did you state the full set of assumptions of all theoretical results? [N/A]
 - (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] A significant chunk (about 20%) of the data used for training (and necessary to reproduce the graphs) from the paper is private. It is expected that these data will be released in ZTF DR9 in early 2022, at which time the complete data sets will be open sourced. The code and supplemental materials are open-source.
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]

- (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
- 4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
 - (a) If your work uses existing assets, did you cite the creators? [N/A]
 - (b) Did you mention the license of the assets? [N/A]
 - (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
 - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
- 5. If you used crowdsourcing or conducted research with human subjects...
 - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
 - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]