



Exploring the Universe with SNAD: Anomaly Detection in Astronomy

Alina A. Volnova^{1,6}(✉) , Patrick D. Aleo^{2,3} , Anastasia Lavrukhina⁴,
Etienne Russeil⁵ , Timofey Semenikhin^{4,6}, Emmanuel Gangler⁵,
Emille E. O. Ishida⁵ , Matwey V. Kornilov^{6,7} ,
Vladimir Korolev^{1,2,3,4,5,6,7,8} , Konstantin Malanchev^{2,6} ,
Maria V. Pruzhinskaya⁵ , and Sreevarsha Sreejith⁸ 

- ¹ Space Research Institute of the Russian Academy of Sciences, Profsoyuznaya
84/32, Moscow 117997, Russia
alinusss@gmail.com
- ² Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W.
Green St., Champaign, IL 61801, USA
- ³ Center for AstroPhysical Surveys, National Center for Supercomputing
Applications, Urbana, IL 61801, USA
- ⁴ Faculty of Space Research, Lomonosov Moscow State University, Leninsky Gori 1
bld. 52, Moscow 119234, Russia
- ⁵ Université Clermont Auvergne, CNRS/IN2P3, LPCA,
63000 Clermont-Ferrand, France
- ⁶ Sternberg Astronomical Institute, Lomonosov Moscow State University,
Universitetsky 13, Moscow 119234, Russia
- ⁷ National Research University Higher School of Economics, 21/4 Staraya
Basmannaya Ulitsa, Moscow 105066, Russia
- ⁸ Physics Department, University of Surrey, Stag Hill, Guildford GU2 7XH, UK

Abstract. SNAD is an international project with a primary focus on detecting astronomical anomalies within large-scale surveys, using active learning and other machine learning algorithms. The work carried out by SNAD not only contributes to the discovery and classification of various astronomical phenomena but also enhances our understanding and implementation of machine learning techniques within the field of astrophysics. This paper provides a review of the SNAD project and summarizes the advancements and achievements made by the team over several years.

Keywords: Methods · data analysis · Supernovae · general · Transients · Astronomical data bases

1 Introduction

In modern astronomy, discoveries of new objects are based on a huge flow of data coming from all-sky surveys (e.g., the Sloan Digital Sky Survey, SDSS [7], the Zwicky Transient Facility, ZTF [5], the Vera Rubin Observatory Legacy

Survey of Space and Time, LSST [19]). The terabytes of data generated every night contain information, allowing the discovery of several hundred transients per year. Undoubtedly, the human resource is limited, and each scientist should choose from a variety of discovered objects those that are of the greatest interest to him. However, this is not a trivial task, given the number of transients to choose from. In addition, the total mass of discovered objects may contain rare or not yet known phenomena. Thus, we are faced with the problem of classifying a large number of objects, and specifically in this case: the search for anomalies or outliers.

It is natural to expect that such large volumes of data that are generated in astronomy today require machine learning (ML) methods for processing. Despite the fact that ML has become an integral part of data analysis in almost all areas of science in recent years, astronomy has benefited from it only recently, starting from solving problems in classification and regression (see, e.g., [21]). In the light of the search for the most interesting objects among all detected ones anomaly detection (AD) algorithms have a wide field of application: would it be search for galaxies with anomalous spectra [4], transient with extraordinary light curves [32], or unusual variable stars [27].

However, most of those algorithms are based on statistical models, and the anomaly is defined as an object that does not fit the model. Observational defects lead to multiple identifications of non-physical events as anomalous, giving the researcher hundreds of candidates for further investigations. To decrease the amount of “not interesting” objects and to detect the astrophysical anomalies AD algorithms need some advisory from human in which object should be considered as anomalous, i.e., active learning (AL) is required. AL is a subclass of ML algorithms where the user may adjust the model by interactively setting scores to the objects which are suspicious for the machine. In this learning paradigm, the user can select which type of object should be considered as anomalous, extracting some rare classes of objects from the bulk of huge sky-survey data (see, e.g., [21, 40]).

Given to its important impact in the future of astronomical discoveries with modern data sets, the problem of anomaly detection have already been explored in the literature. This includes recent studies using data streams [30], deep [12], generative [38] and active [15, 22] methods. Nevertheless, the incidence of false positives continues to be an important bottle neck to be overcome if we intend to take full advantage of modern astronomical data. We describe here one of such attempts, which has been consistently focusing on anomaly detection for astronomy in the last 5 years.

This paper is an overview of the SNAD project – an international group of researchers (including astronomers, physicists, data scientists and mathematicians) which focus on developing new AL algorithms for the search for anomalous objects in large data-sets of all-sky astronomical surveys. In Sect. 2 we provide the description and main goals of the project and the team, and briefly summarise data sources and used techniques. In Sect. 3 the most prominent results of the project are presented. Section 4 describes some auxiliary products devel-

oped by the team, which are now publicly available. Section 5 gives conclusive remarks for future prospects of the project.

2 What is the SNAD Project?

2.1 Goals and Objectives

The goal of the SNAD¹ project is to develop a pipeline where human expertise and modern machine learning techniques can complement each other in the task of identifying unusual astronomical objects mostly by their photometrical features. The team concentrates on the search for unusual, rare or yet unknown objects in the sets of photometric light curves (LCs) by combining different AL AD algorithms with the additional information provided by the expert in order to label a significant fraction of the most obvious outliers and choose those which are true astronomical anomalies. Enabling reliable anomaly/outlier detection based solely on photometric observations is one of the fundamental puzzles to be solved before we can convert the full potential of large-scale surveys into scientific results. This project represents an effective strategy to guarantee we shall not overlook exciting new science hidden in the data we fought so hard to acquire.

2.2 Team Members and Expertise

The SNAD team is composed of young researchers, each offering a unique set of skills and experiences that are utilized within the project. All team members are involved in various stages of the anomaly detection process, from feature engineering to expert analysis.

In addition to their collective expertise in anomaly detection, each team member also contributes their unique knowledge from different fields: gamma-ray bursts (Alina Volnova), supernovae (Maria Pruzhinskaya), fast transients (Anastasia Lavrukhina), accretion flows in astrophysics (Konstantin Malanchev), adapting machine learning algorithms (Emille E. O. Ishida), astronomical site characterization (Matwey Kornilov), time-domain astronomy and photometric classification (Partick Aleo), astronomical image analysis and dwarf galaxy detection (Sreevarsha Sreejith), real-bogus classification (Timofey Semenikhin), aeronautical science (Vladimir Korolev), symbolic regression for light curve analysis (Etienne Russeil), observational cosmology (Emmanuel Gangler).

Over the years, several other researchers, including Florian Mondon, Anastasia Malancheva, Alexandra Novinskaya, and Anastasiya Voloshina have also contributed their skills and knowledge to the project.

SNAD activities are designed around annual meetings whose format was inspired by other initiatives which focused on boosting innovation and creativity (e.g. the silicon valley model [2] and the Cosmostatistics Initiative²). We also

¹ <https://snad.space>.

² <https://cosmostatistics-initiative.org/>.

invite³ other researchers and students of different levels for collaboration in the field of astronomical anomalies detection. The resulting environment is one of the key components which play an important role in the development of products described in this work.

2.3 Data Sources and Techniques

The first data-set used for our research was extracted from the Open Supernova Catalog⁴ (OSC; [18]). It contains all publicly available data on all SNe from a dozen of catalogues, including, in different cases, multi-colour LCs, spectra, redshift estimations and classification. We extracted 1999 LCs with the number of data points enough to fit the LC within some particular interval of time relative to the maximum, and presented in 3 different photometric pass-bands (gri , $g'r'i'$ or BRI). These 1999 LCs allowed us to test our main ideas by finding several peculiar and super-luminous SNe, along with a few dozens of miss-classified objects [32].

The results of the first try encouraged us to use more numerous data-set of photometric LCs, and the Zwicky Transient Facility (ZTF) survey data were chosen. The ZTF survey started on March 2018 and during its initial phase has observed around a billion objects [5]. Each object is represented with a bunch of LCs in filters g , r , i with a cadence – on average ~ 1 day for the Galactic plane and ~ 3 days for the Northern-equatorial sky. To minimize the affection of different cadences on the results we selected 3 different fields: 1 in the M31 galaxy, 1 in the Galactic plane, and 1 far above the Galactic disk and analysed data from the first 9.4 months of the ZTF survey, between 17 March and 31 December 2018. The total amount of objects with at least 100 data-point in the LC is 2.25 millions. The results of the search for anomalies in this set are described in Sect. 3.

Investigating all data described above we experimented with a series of different anomaly detection algorithms. Our goal was to quantify their effectiveness in identifying scientifically interesting anomalous objects within large data sets. Among others, we used Isolation Forest (IF) [26], Local Outlier Factor [9], Gaussian Mixture Models [29] and one-class Support Vector Machines [35]. These were applied to small [32] as well as large [27] data sets with encouraging results.

However, we noticed that frequently, the majority of objects with high anomaly score represent non-astrophysical artificial effects. These were most times border effects, glitches in the detector, bad CCDs, cosmic rays, unexpected telescope movements, satellites, etc. This discovery was important, and generated an additional line of investigation, described in Sect. 3.4. Nevertheless, the need of an adaptive algorithm, which could allow the user to define which type of anomaly was interesting became increasingly more evident.

We started experimenting with adaptive learning techniques by using the Active Anomaly Discovery algorithm (AAD) [13], which uses a human-in-the-loop strategy to apply a series of sequential modifications to a traditional IF. By

³ <https://snad.space/#contact>.

⁴ <https://github.com/astrocatalogs>.

downgrading decision paths which disagree with the expert’s definition of what is an interesting anomaly the algorithm allows the construction of a personalized AD model. We showed that this strategy is effective in small [22] and large [31] data sets. The successful experiences with AAD motivated many of the results described below.

3 Key Contributions and Discoveries

3.1 Supernova Catalog

The SNAD team discovered potential, unreported supernovae—powerful and bright explosion of a star, indicating the final stage of its evolution [8]—within the ZTF DRs during a non-targeted anomaly detection search [27]. This led to the inception of a new experiment aimed at developing specialized machine learning models. The AAD algorithm (Sect. 2.3) and the SNAD Transient Miner (Sect. 4.3) have been trained and refined based on our long-term experience with supernovae. Applying these algorithms, the SNAD team identified 144 new supernova candidates hidden within the vast photometric data of the ZTF survey.

Each of these candidates has been thoroughly inspected and validated by our domain experts. Detailed information about the candidates has been made publicly available in the SNAD supernova catalog⁵.

Furthermore, these supernova candidates have also been reported to the Transient Name Server⁶ (TNS). The TNS is the official resource for announcing new astronomical transients, and our reports ensure that these candidates are available for further study by the international astronomical community. With this dual approach of cataloguing and reporting, we aim to foster collaboration and further our understanding of supernovae.

3.2 Superluminous Supernova Candidates

Among the supernova candidates discovered using the AAD algorithm (Sect. 2.3) we report four objects: SNAD120, SNAD121, SNAD160, SNAD187, which could belong to the superluminous supernovae class. They display an unusually broad LC when compared to the Nugent’s models⁷ and multiple redshift estimations indicate their high absolute brightness [31]. Among them, SNAD160 displays a particularly broad LC and further analysis indicates that it could belong to the pair instability supernovae class [33]. It is a theoretical category of thermonuclear explosions of extremely massive stars, as described by [3] [16]. There have been no reports of observational confirmation of such event yet and any good candidate contributes to enrich our understanding of the phenomena.

Additional analysis has been performed by inputting SNAD160 to a symbolic regression algorithm. It is a machine learning method based on genetic programming [25] that computes a mathematical expression with independent variables

⁵ <https://snad.space/catalog/>.

⁶ <https://www.wis-tns.org/>.

⁷ https://c3.lbl.gov/nugent/nugent_templates.html.

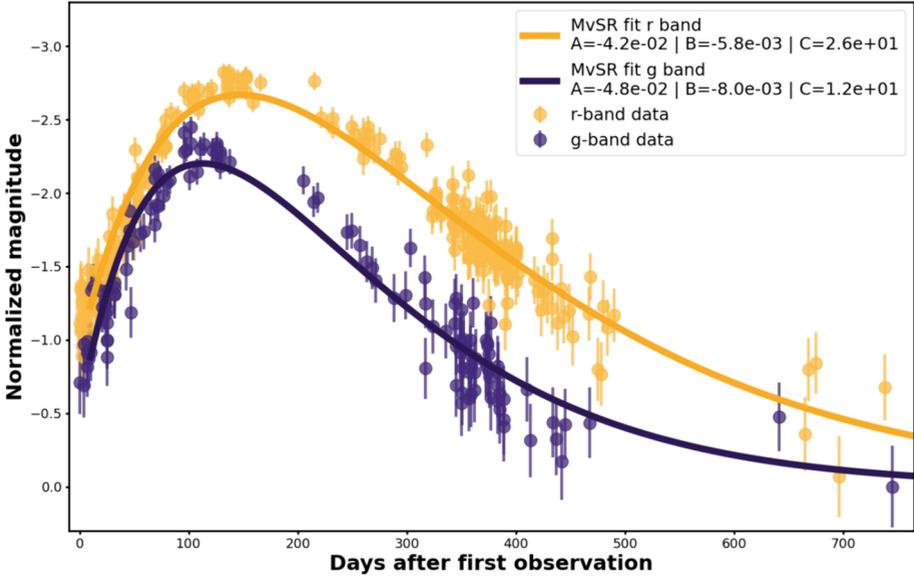


Fig. 1. Fit of SNAD160 using Eq. 1 resulting from MvSR procedure. The magnitude is normalized by subtraction of the peak magnitude.

to optimally fits some input data. We developed a more complex approach of the algorithm called Multiview symbolic regression (MvSR) [34]. It allows for the input of multiple datasets supposedly generated by the same parametric function, and directly returns the common parametric function that generated the examples. MvSR was applied to SNAD160 with LCs in passbands g and r being used as two separate datasets. The resulting parametric function describes transient like behaviors with a linear rising and an exponential decay.

$$f(t) = A(t + C) \times e^{B(t+C)} \quad (1)$$

It is particularly well fitting SNAD160 in both pass-bands as shown in Fig. 1. It has been applied to other SLSN candidates and provides an accurate representation of the LCs. With three free parameters, this equation offers a noticeably dense description of supernova events. Using this representation for further feature extraction analysis could lead to more interesting SLSN discoveries.

3.3 Red Dwarf Flares

Within the SNAD group, we are currently working on detecting red dwarf flares in the high cadence data of the ZTF survey releases. Stellar flares are very energetic phenomena, during which the optical luminosity of a star increases by several fold over tens of seconds. The currently accepted version of stellar flares' nature is a magnetic field, generated in the convective stellar interior [17].

The detection of rapid stellar flares, such as those from red dwarfs, is of significant importance in the field of exoplanet science, particularly when studying exoplanet habitability. Additionally, creating a statistically significant sample of red dwarf flares can greatly aid in our understanding of the physical processes involved in these events.

By this time, approximately a hundred new candidates for red dwarf flares have been found [39]. Some of them were discovered using the AAD algorithm (Sect. 2.3). Others were detected by fitting the LCs with a model that accurately describes the properties and shape of red dwarf flares' LCs, followed by further analysis of the goodness of fit. The candidates found using presented methods are then subject to further analysis and approval by experts within the SNAD team, with the assistance of the SNAD ZTF Viewer (Sect. 4.1).

Examples of found red dwarf flare candidates LCs presented in Fig. 2. Found candidates vary both by a LC shape (multiple outburst or one outburst flare) and brightness, which allows exploring the complexity of flare events and forming a more comprehensive sample.

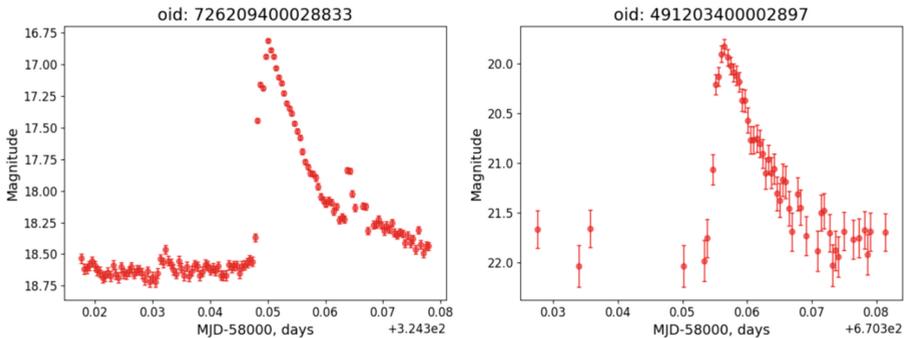


Fig. 2. Two examples of red dwarf flares candidates found by SNAD team. (Color figure online)

3.4 Catalog of Artefacts

Another interesting data product that resulted from the construction of the SNAD database is the SNAD catalog of artefacts. The term artefact is generally used as an umbrella term to denote observational, phenomenological or instrumental oddities in signal that manifest as diffraction spikes, a step-effect in brightness, colour saturation etc. to name a few. We are collating a catalog of such occurrences in the SNAD database that is aimed for both outreach and science goals. The catalog consists of the OIDs (identifying number that's referenced in ZTF), the FITS images of the respective fields and the SNAD link to the main object page denoted by the OID. Some examples of the artefacts thus identified are given in Fig. 3. While spurious in the strictest sense, several of these artefacts present as visually breathtaking making them perfect for outreach purposes.

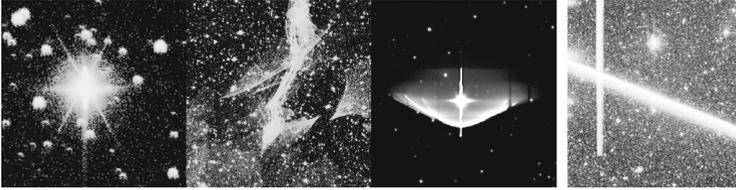


Fig. 3. Some examples from the SNAD catalog of artefacts.

The catalog of artefacts also provides a data set for training machine learning algorithms that could potentially be used for identifying and/or removing these objects from fields of interest. Using the labels of 2230 objects from SNAD knowledge database (Sect. 4.2), roughly half of which are artefacts, we developed an algorithm that predicts whether an object is an artefact or not based on the sequence of object frames from the ZTF survey. The algorithm consists of two parts: (1) a variational auto-encoder that returns the compressed representation of each frame after being trained on all the images from the sample and, (2) a recurrent neural network that takes the sequence of compressed frame representations and an object label as inputs, and returns the probability that the object is an artefact. Currently the best classification result is $\text{ROC} - \text{AUC} = 0.856 \pm 0.010$ and $\text{Accuracy} = 0.802 \pm 0.023$ [36].

4 Scientific Tools and Resources

4.1 SNAD ZTF Viewer

The SNAD Viewer⁸ [28] is a web portal specifically designed for astronomers, providing a centralized view of individual objects from various data releases. It integrates data from multiple publicly available astronomical archives and sources, thereby offering a comprehensive platform for time-domain data analysis and interpretation.

In the era of big data in astronomy, the SNAD Viewer emerges as a significant tool designed to centralize and streamline the management of astronomical data. Initially conceived to facilitate expert feedback in active machine learning applications (see Sects. 4.3, 4.4), the SNAD Viewer has evolved into a valuable

⁸ <https://ztf.snad.space>.

community asset. It centralizes public information and provides a multi-dimensional view of individual objects from the ZTF data releases.

The SNAD Viewer’s infrastructure is characterized by its scalability and flexibility. It is designed to accommodate a wide range of data types and user needs, demonstrating its adaptability to the changing landscape of astronomical research. This adaptability extends to its ability to be personalized and used by other surveys and for various scientific goals, underscoring its broad applicability within the field.

Importantly, the SNAD Viewer is not an isolated entity but is part of a larger network of astronomical data portals. It is linked to by other significant portals such as Antares, Fink, YSE PZ, and Astro-COLIBRI, which further enhances its accessibility and utility within the astronomical community. This interconnectedness highlights the collaborative nature of modern astronomical research and the crucial role of the SNAD Viewer within this ecosystem.

The SNAD Viewer is publicly available online, emphasizing its commitment to open science and the democratization of astronomical data. It serves as a testament to the crucial roles that domain experts continue to play in the era of big data in astronomy.

4.2 SNAD Knowledge Database

In addition to the discovery of candidates in anomalies, our research also led to the creation of a knowledge database. One of the core features of the SNAD ZTF Viewer (Sect. 4.1) is the ability to assign specific tags to each ZTF object, a function our experts extensively used when reviewing potential anomalies. This tagging system incorporates a wide range of general astronomical classes, including variable stars, transients, active galactic nuclei, along with their various specific types and subtypes. We also devised custom tags for internal purposes, as well as non-astrophysical tags like artefacts and their subtypes. Multiple tags can be assigned to a single object, with the history of tag changes stored in the database.

During the expert analysis, a total of 7238⁹ objects were labelled. Despite the ZTF data processing pipeline’s procedure to distinguish astrophysical events from bogus ones, almost a half of them are artefacts. These assigned labels serve as a reliable source of curated data for training supervised machine learning models. Furthermore, they can help to refine the ZTF’s pre-processing pipeline, ensuring more efficient identification and classification of astronomical events in the future.

4.3 SNAD Transient Miner

The SNAD Miner is an exhaustive similarity search for LC features for transient discovery, using simulated LCs as a guide to identify transients in a large dataset

⁹ on Sep. 18, 2024

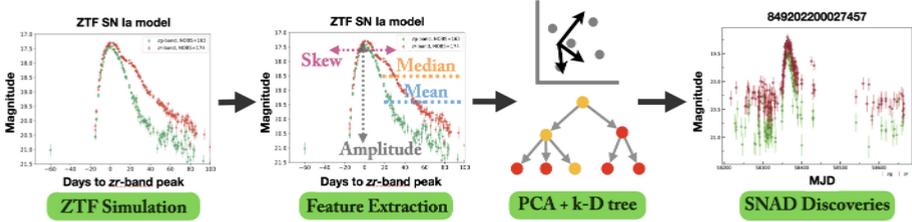


Fig. 4. SNAD Miner schematic. We use bright ZTF SNe simulations (left) and extract their LC features (left center). Then, we apply a (PCA+) k-D tree on these features, to search for real ZTF DR events nearest neighbors (right center). Some of these matched nearest neighbors were previously missed SNe (right).

mostly comprised of variable stars [11]. Specifically, it leverages extracted statistical features from simulated SNe LCs, from which a k-D tree is applied to search for similar feature values in the ZTF DR4. As a result, we discovered 11 previously missed transients, 7 SNe candidates and 4 AGN candidates.

Realistic ZTF simulations were generated using SNANA [24], with ZTF data release 3 cadence and magnitude error distribution (see [10] for details). Starting from the template models originally developed for PLAsTiCC, [20, 23], we selected the seven brightest, well-sampled LCs to use as a reference (3 SLSN-I, 1 SN Ia, 1 SN II and 2 TDE with peak magnitude $\sim 17^m$).

Using these objects, we applied a k-D tree [6] to their extracted 82 non-normalized features. Subsequently, we identified the 15 nearest neighbors for each simulation (105 matches in total, resulting in 89 unique ZTF DR4 sources), and manually inspected the results. Ultimately, we discovered 11 previously unreported supernova and active galactic nucleus candidates. The remaining 94 matches (81 unique ZTF DR4 sources) were either known/already reported transients or variable stars. Given the large number of variable stars previously estimated in ZTF data releases [11], it is reasonable to expect many well sampled, high-amplitude variable stars whose coverage in parameter space significantly overlaps with the regions populated by transients (see, e.g. Figure 4 in [1]). Thus, a ratio of 18 transients (11 newly-discovered) to 44 variable stars out of 89 unique sources selected from 990,220 considered ZTF sources is a very successful result. Considering the extreme case where ≈ 3000 SNe discovered by ZTF [14] were part of the data set, the expected incidence of SNe when choosing 100 sources at random would be of < 1 (≈ 0.3) event.

This “SNAD Miner” process is flexible, and can use simulated or real objects as the input, any number of features, and any number of nearest neighbors to “mine” in the existing dataset. A schematic is shown in Fig. 4.

4.4 Coniferest Python Library

The `coniferest` library [37] is aiming to add adaptive capabilities to isolation forest algorithm which is inherently static.

The library has implementations of two adaptive learning algorithms. One of them is an already mentioned earlier AAD algorithm [13]. And the second one is the implementation of our own adaptive learning isolation forest algorithm named PineForest. It is based of tree filtering approach. After every new observation with some label given by an expert the PineForest filters out trees that do not push forward right observations.

This approach have a few remarkable advantages. First, the algorithm has not much hyperparameters, so it is easy to tune. Second, it maybe used for both – as an adaptive learning with an expert in loop and as an accumulator of prior knowledge about data. Finally, it has a very good performance characteristics making it suitable for data intensive applications.

Also, as a bonus the library includes our own implementations of classical isolation forest with much better performance in scoring compared to `scikit-learn`'s one. Refer to `coniferest` package documentation for more details.

5 Future Directions and Challenges

Large scale sky surveys have fundamentally changed the process of astronomical investigation and discovery. In the era of LSST, when millions of new transients will be detected every night, serendipitous discoveries will not happen. On the other hand, targeted searches are bound to identify objects which fall within our domain knowledge.

The SNAD team has been consistently working on the development of adaptable algorithms and tools which allow the user to probe the boundaries of their domain knowledge – thus enabling scientific discovery in large data sets. These have been proven to be effective in current state of the art catalog data.

In the near future we intend to concentrate our efforts in two important bottlenecks: ensuring that our tools are scalable to meet LSST requirements and make additional connections with expert communities to guarantee scientific impact of SNAD products. The latter will allow the SNAD team to develop increasingly more personalized tools which will fulfill the requirements of individual experts and ensure that we can fully exploit the scientific potential of modern astronomical surveys.

Acknowledgements. Authors thank the Ministry of Science and Higher Education of Russian Federation for financial support, grant 075-15-2022-1221 (2022-BRICS-8847-2335). We used the equipment funded by the Lomonosov Moscow State University Program of Development. This work made use of the Illinois Campus Cluster, a computing resource that is operated by the Illinois Campus Cluster Program (ICCP) in conjunction with the National Center for Supercomputing Applications (NCSA) and which is supported by funds from the University of Illinois at Urbana-Champaign.

References

1. Aleo, P.D., Malanchev, K.L., Pruzhinskaya, M.V., et al.: SNAD transient miner: finding missed transient events in ZTF DR4 using k-D trees. *New Astron.* **96**, 101846 (2022). <https://doi.org/10.1016/j.newast.2022.101846>
2. Annika Steiber, S.A.: *The Silicon Valley Model*. Springer, Cham (2016)
3. Barkat, Z., Rakavy, G., Sack, N.: Dynamics of supernova explosion resulting from pair formation. *Phys. Rev. Lett.* **18**, 379–381 (1967). <https://doi.org/10.1103/PhysRevLett.18.379>
4. Baron, D., Poznanski, D.: The weirdest SDSS galaxies: results from an outlier detection algorithm. *Mon. Not. R. Astron. Soc.* **465**(4), 4530–4555 (2017). <https://doi.org/10.1093/mnras/stw3021>
5. Bellm, E.C., Kulkarni, S.R., Graham, M.J., et al.: The Zwicky transient facility: system overview, performance, and first results. *Publ. Astron. Soc. Pac.* **131**(995), 018002 (2019). <https://doi.org/10.1088/1538-3873/aaecbe>
6. Bentley, J.L.: Multidimensional binary search trees used for associative searching. *Commun. ACM* **18**(9), 509–517 (1975). <https://doi.org/10.1145/361002.361007>
7. Blanton, M.R., Bershady, M.A., Abolfathi, B., et al.: Sloan digital sky survey IV: mapping the milky way, nearby galaxies, and the distant universe. *Astron. J.* **154**(1), 28 (2017). <https://doi.org/10.3847/1538-3881/aa7567>
8. Branch, D., Wheeler, J.C.: *Supernova explosions* (2017). <https://doi.org/10.1007/978-3-662-55054-0>
9. Breunig, M.M., Kriegel, H.P., Ng, R.T., et al.: LOF: identifying density-based local outliers. In: *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pp. 93–104 (2000)
10. Chatterjee, D., Narayan, G., Aleo, P.D., et al.: El-CID: a filter for gravitational-wave electromagnetic counterpart identification. *arXiv e-prints* [arXiv:2108.04166](https://arxiv.org/abs/2108.04166) (2021)
11. Chen, X., Wang, S., Deng, L., et al.: The Zwicky transient facility catalog of periodic variable stars. *Astrophys. J. Suppl. Ser.* **249**(1), 18 (2020). <https://doi.org/10.3847/1538-4365/ab9cae>
12. Ćiprijanović, A., Lewis, A., Pedro, K., et al.: DeepAstroUDA: semi-supervised universal domain adaptation for cross-survey galaxy morphology classification and anomaly detection. *Mach. Learn.: Sci. Technol.* **4**(2), 025013 (2023). <https://doi.org/10.1088/2632-2153/acca5f>
13. Das, S., Wong, W.K., Fern, A., et al.: Incorporating feedback into tree-based anomaly detection. In: *Workshop on Interactive Data Exploration and Analytics (IDEA 2017)*, p. [arXiv:1708.09441](https://arxiv.org/abs/1708.09441). *KDD Workshop* (2017)
14. Dhawan, S., Goobar, A., Smith, M., et al.: The Zwicky transient facility type Ia supernova survey: first data release and results. *Mon. Not. Roy. Astron. Soc.* **510**(2), 2228–2241 (2022). <https://doi.org/10.1093/mnras/stab3093>
15. Etsebeth, V., Lochner, M., Walmsley, M., et al.: Astronomy at scale: searching for anomalies amongst 4 million galaxies. *arXiv e-prints* [arXiv:2309.08660](https://arxiv.org/abs/2309.08660) (2023). <https://doi.org/10.48550/arXiv.2309.08660>
16. Gal-Yam, A.: The most luminous supernovae. *Annu. Rev. Astron. Astrophys.* **57**(1), 305–333 (2019). <https://doi.org/10.1146/annurev-astro-081817-051819>
17. Gershberg, R.E.: *Solar-Type Activity in Main-Sequence Stars*. Springer, Cham (2005). <https://doi.org/10.1007/3-540-28243-2>
18. Guillochon, J., Parrent, J., Kelley, L.Z., et al.: An open catalog for supernova data. *Astrophys. J.* **835**(1), 64 (2017). <https://doi.org/10.3847/1538-4357/835/1/64>

19. Hambleton, K.M., Bianco, F.B., Street, R., et al.: Rubin observatory LSST transients and variable stars roadmap. arXiv e-prints [arXiv:2208.04499](https://arxiv.org/abs/2208.04499) (2022). <https://doi.org/10.48550/arXiv.2208.04499>
20. Hložek, R., Ponder, K.A., Malz, A.I., et al.: Results of the photometric LSST astronomical time-series classification challenge (PLAsTiCC). arXiv e-prints [arXiv:2012.12392](https://arxiv.org/abs/2012.12392) (2020)
21. Ishida, E.E.O., Beck, R., González-Gaitán, S., et al.: Optimizing spectroscopic follow-up strategies for supernova photometric classification with active learning. *Mon. Not. R. Astron. Soc.* **483**(1), 2–18 (2019). <https://doi.org/10.1093/mnras/sty3015>
22. Ishida, E.E.O., Kornilov, M.V., Malanchev, K.L., et al.: Active anomaly detection for time-domain discoveries. *Astron. Astrophys.* **650**, A195 (2021). <https://doi.org/10.1051/0004-6361/202037709>
23. Kessler, R., Narayan, G., Avelino, A., et al.: Models and simulations for the photometric LSST astronomical time series classification challenge (PLAsTiCC). *Publ. Astron. Soc. Pac.* **131**(1003), 094501 (2019). <https://doi.org/10.1088/1538-3873/ab26f1>
24. Kessler, R., Bernstein, J.P., Cinabro, D., et al.: SNANA: a public software package for supernova analysis. *Publ. Astron. Soc. Pac.* **121**(883), 1028 (2009). <https://doi.org/10.1086/605984>
25. Koza, J., Koza, J.: Genetic Programming: On the Programming of Computers by Means of Natural Selection. A Bradford Book, Bradford (1992). <https://books.google.com.br/books?id=Bhtxo60BV0EC>
26. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422. IEEE (2008)
27. Malanchev, K.L., Pruzhinskaya, M.V., Korolev, V.S., et al.: Anomaly detection in the Zwicky transient facility DR3. *Mon. Not. R. Astron. Soc.* **502**(4), 5147–5175 (2021). <https://doi.org/10.1093/mnras/stab316>
28. Malanchev, K., Kornilov, M.V., Pruzhinskaya, M.V., et al.: The SNAD viewer: everything you want to know about your favorite ZTF object. *Publ. Astron. Soc. Pac.* **135**(1044), 024503 (2023). <https://doi.org/10.1088/1538-3873/acb292>
29. McLachlan, G.J., Peel, D.: Finite mixture models. Wiley Series in Probability and Statistics. New York (2000)
30. Perez-Carrasco, M., Cabrera-Vives, G., Hernandez-García, L., et al.: Alert classification for the ALeRCE broker system: the anomaly detector. *Astron. J.* **166**(4), 151 (2023). <https://doi.org/10.3847/1538-3881/ace0c1>
31. Pruzhinskaya, M.V., Ishida, E.E.O., Novinskaya, A.K., et al.: Supernova search with active learning in ZTF DR3. *Astron. Astrophys.* **672**, A111 (2023). <https://doi.org/10.1051/0004-6361/202245172>
32. Pruzhinskaya, M.V., Malanchev, K.L., Kornilov, M.V., et al.: Anomaly detection in the open supernova catalog. *Mon. Not. R. Astron. Soc.* **489**(3), 3591–3608 (2019). <https://doi.org/10.1093/mnras/stz2362>
33. Pruzhinskaya, M., Volnova, A., Kornilov, M., et al.: Could SNAD160 be a pair-instability supernova? *Res. Notes Am. Astron. Soc.* **6**(6), 122 (2022). <https://doi.org/10.3847/2515-5172/ac76cf>
34. Russeil, E., et al.: Multi-view symbolic regression. arXiv e-prints [arXiv:2402.04298](https://arxiv.org/abs/2402.04298) (2024). <https://doi.org/10.48550/arXiv.2402.04298>
35. Schölkopf, B., Williamson, R., Smola, A., et al.: Support vector method for novelty detection, vol. 12, pp. 582–588 (1999)

36. Semenikhin, T.A.: Neural network architecture for artifacts detection in ZTF survey. *Syst. Means Inf.* **34**(1), 70–79 (2024). <https://doi.org/10.14357/08696527240106>
37. SNAD: Coniferest python package. <https://github.com/snad-space/coniferest>
38. Storey-Fisher, K., Huertas-Company, M., Ramachandra, N., et al.: Anomaly detection in Hyper Suprime-Cam galaxy images with generative adversarial networks. *Mon. Notes R. Astron. Soc.* **508**(2), 2946–2963 (2021). <https://doi.org/10.1093/mnras/stab2589>
39. Voloshina, A.S., et al.: SNAD catalogue of M-dwarf flares from the Zwicky Transient Facility. *Mon. Not. R. Astron. Soc.* **533**(4), 4309–4323 (2024). <https://doi.org/10.1093/mnras/stae2031>
40. Walmsley, M., Smith, L., Lintott, C., et al.: Galaxy zoo: probabilistic morphology through Bayesian CNNs and active learning. *Mon. Not. R. Astron. Soc.* **491**(2), 1554–1574 (2020). <https://doi.org/10.1093/mnras/stz2816>