Astronomy and Computing 55 (2026) 101049

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

# Hunting the outliers: Machine learning for anomalous time series detection ☆

N. De Bonis [a,b,1,*], S. Vaccaro [a,b,1], Y. Maruccia [a,b], G. Riccio [a], R. Crupi [c], S. Rubini [c], D. De Cicco [d,a,e], M. Brescia [d,a], S. Cavuoti [a,f]

[a] INAF - Astronomical Observatory of Capodimonte, Via Moiariello 16, Napoli, I-80131, Italy
[b] ICSC—Centro Nazionale di Ricerca in HPC, Big Data e Quantum Computing, Via Stalingrado 84/3, Bologna, I-40121, Italy
[c] Intesa Sanpaolo S.p.A., Corso Inghilterra 3, Turin, I-10138, Italy
[d] Department of Physics "E. Pancini", University Federico II of Napoli, Via Cinthia 21, Napoli, I-80126, Italy
[e] Millennium Institute of Astrophysics (MAS), Nuncio Monsenor Sotero Sanz 100, Santiago, CL-7500000, Chile
[f] INFN section of Naples, via Cinthia 6, Napoli, I-80126, Italy

ARTICLE INFO

*Keywords:*
Anomaly detection
Machine learning
Autoencoder
Isolation forest
Feature engineering

ABSTRACT

The increasing availability of large-scale time series datasets from modern astronomical surveys, such as those provided by Gaia and the forthcoming Legacy Survey of Space and Time (LSST) to be conducted with the Simonyi Survey Telescope at the Vera C. Rubin Observatory, is transforming time-domain astrophysics, enabling the systematic study of variable and transient phenomena across billions of sources. However, the sheer volume and heterogeneity of these data present significant challenges for traditional analysis techniques. Feature-based representations have emerged as a powerful solution, allowing the application of machine learning methods for efficient characterization and classification of astrophysical sources, including the reliable identification of Active Galactic Nuclei (AGNs). In this work, we introduce a general-purpose methodology for anomaly detection in time series data that transfers this feature engineering framework to the financial domain, where time series display complexities, such as noise and irregular patterns, closely resembling those in astrophysics. By combining domain-informed features with unsupervised algorithms (specifically Isolation Forests and Autoencoders), our approach effectively detects anomalous time series relative to the sample studied, demonstrating strong performance and highlighting its cross-domain transferability. Moreover, we propose an extension based on adaptive temporal windows to localize anomalies at a finer temporal resolution, further enhancing detection capabilities. Finally, we discuss the potential for reapplying this adaptive strategy to astrophysical time series, aiming to improve the identification of rare or unexpected behaviors in future studies.

## 1. Introduction

The current landscape of time-domain astrophysics is shaped by the increasing availability of large-scale time series datasets produced by modern surveys, such as those provided by Gaia (Prusti et al., 2016), the Zwicky Transient Facility (ZTF, Bellm et al., 2018) or the forthcoming Legacy Survey of Space and Time (LSST) to be conducted with the Simonyi Survey Telescope at the Vera C. Rubin Observatory (abe; Ivezić et al., 2019), all in the optical bands. Other surveys, such as the High Time Resolution Universe (HTRU; Keith et al. 2010) and the MPIfR-MeerKAT Galactic Plane Survey (MMGPS; Padmanabh et al. 2023), explore the radio band. These surveys systematically monitor the sky over extended periods, producing billions of light curves that encode the temporal evolution of a wide variety of astrophysical sources. The analysis of such large and complex datasets is essential to study transient phenomena, variable stars, and rare events. However, the volume and heterogeneity of the data pose significant challenges, especially when relying on traditional analysis techniques that often require manual intervention, domain-specific heuristics, or substantial computational effort.

In this context, machine learning (ML) methods have gained increasing relevance, offering scalable and flexible tools to extract knowledge from high-throughput astronomical datasets. ML techniques can be tailored to a wide range of tasks, including classification, clustering, regression, and anomaly detection, depending on the scientific goals and the nature of the available data. For instance, supervised learning methods have been extensively applied for the classification of

☆ This article is part of a Special issue entitled: 'HPC in Cosmology and Astrophysics' published in Astronomy and Computing.

* Corresponding author at: INAF - Astronomical Observatory of Capodimonte, Via Moiariello 16, Napoli, I-80131, Italy.
*E-mail address:* natale.debonis@inaf.it (N. De Bonis).

[1] The first two authors share the first authorship and are in alphabetical order.

https://doi.org/10.1016/j.ascom.2025.101049
Received 30 September 2025; Received in revised form 7 December 2025; Accepted 15 December 2025
Available online 23 December 2025

AGN (De Cicco et al., 2021, 2025), variable stars (Armstrong et al., 2016; Wang et al., 2025), star formation (Maruccia et al., 2025a) or exoplanets (Rao et al., 2021) where labeled datasets are available for training. Unsupervised clustering techniques, on the contrary, do not need labels, causing a huge increase in their popularity, as shown by their extensive use in astronomy (e.g., Álvarez et al., 2025; Pinciroli Vago et al., 2025; Zhu et al., 2025; Fotopoulou, 2024). Anomaly detection, broadly defined as the identification of patterns that deviate from expected behavior, has a long-standing tradition in a variety of domains, including cybersecurity (e.g., Salman et al., 2025; Rajendran et al., 2024), manufacturing (e.g., Engbers and Freitag, 2024; Xie et al., 2024), healthcare (e.g., Sriraam et al., 2025; Nawaz et al., 2024), and finance (e.g., Maruccia et al., 2025b; Chandola et al., 2009) providing a comprehensive overview of the applicability of this technique across many domains. In astrophysics, early applications focused on image-based outlier detection (e.g., via convolutional neural networks and vision transformers, Parimucha et al., 2025), and more recently on time series (Gupta et al., 2025), where the task is particularly challenging due to irregular sampling, observational noise, and intrinsic variability. Several studies have applied machine learning methods such as clustering, autoencoders, and deep generative models to characterize variability and identify atypical sources (e.g., Campagne, 2025; Gebran and Bentley, 2025; Tang et al., 2025; Webb et al., 2020).

A critical component in the success of methods for time series analysis is the design of an appropriate representation of the data. Given the complexity and heterogeneity of astrophysical light curves, often characterized by irregular sampling, missing data, and varying lengths due to non-contemporaneous observations, it is challenging to directly apply ML methods that require uniform input structures. Feature extraction techniques have been widely adopted to summarize time-domain information into a structured feature space. In particular, they address the above cited limitations by mapping each light curve into a fixed-dimensional representation, regardless of its length or sampling pattern. These representations provide a consistent and compact description of the time-domain information, enabling the application of both supervised and unsupervised learning methods in a computationally efficient manner, even at a large scale. In this work, we make use of the Feature Analysis for Time Series library (FATS; Nun et al., 2015), which provides a widely used implementation of such methods.

Several studies in recent years have demonstrated the effectiveness of feature-based approaches for identifying Active Galactic Nuclei (AGNs) among different types of objects, including galaxies and stars (De Cicco et al., 2019; Sánchez-Sáez et al., 2021; De Cicco et al., 2021, 2025; Maruccia et al., 2025c). As shown in De Cicco et al. (2019), these features capture key aspects of the sources variability, as well as their statistical properties and morphological characteristics, thus transforming the problem from a time-domain analysis to one based on features. This framework supports a variety of analyses, ranging from classical variability studies (De Cicco et al., 2015, 2019) to modern ML approaches, including both supervised (see De Cicco et al., 2021, 2025) and unsupervised methods (Maruccia et al., 2025c). Notably, these strategies have consistently achieved strong results in the reliable identification of AGNs.

In this work, we present a general-purpose methodology for anomaly detection in time series data, in which we transfer the previously described feature engineering approach to the domain of financial time series. In particular, the proposed framework combines domain-informed feature engineering with unsupervised machine learning algorithms, specifically Isolation Forests (Liu et al., 2008) and Autoencoders (Hinton and Zemel, 1993). The method is developed and validated on financial transaction time series, demonstrating both its transferability to a new application domain and its potential industrial applicability.

Financial and astrophysical time series substantially differ in their nature: the former are typically non-stationary and abrupt regime changes. The latter are often characterized by periodic signals, long-term trends, and observational noise. Nevertheless, despite these fundamental differences, both domains face similar challenges related to the identification of outliers in complex temporal data, especially in the absence of labeled ground truth. A first example of such an application was presented in Maruccia et al. (2025b), where features were extracted from financial time series using the FATS library and subsequently analyzed with Isolation Forests and Self-Organizing Maps (Kohonen, 2001) for anomaly detection. In this work, we extend the methodology by introducing adaptive temporal windows during feature extraction, which enables the localization of anomalies not only at the global object level but also within specific segments of the time series. This validation underscores the flexibility and robustness of the proposed approach, originally developed to address astrophysical variability, and demonstrates its potential for broader application in heterogeneous, high-throughput environments. Finally, as a perspective for future work, we plan to transfer the adaptive temporal windowing strategy back to the astrophysical domain, with the goal of enhancing the results previously obtained.

The remainder of this paper is structured as follows. Section 2 briefly describes the dataset used for our analysis. In Section 3 we outline the method adopted in this work, while Section 4 describes our experiments. Finally, we discuss the findings and limitations in Section 5, and in Section 6 we conclude the paper and outline directions for future research.

## 2. Dataset

The initial dataset has been extracted from Intesa Sanpaolo (ISP) database and arranged in time series of daily balances, each associated with a specific user, as well described in Maruccia et al. (2025b). The full dataset comprises 2,636,027 different time series, spanning a two-year period. For the work described in this article, we extracted 300,000 time series in order to limit computational costs. These time series were randomly selected and are representative of the overall sample, as it will be shown in Section 3.1. These time series were further analyzed and cleaned by removing all cases where the length of the series (and thus the available data for a given user) was shorter than 30 days. Such cases typically correspond either to accounts that were closed shortly after the beginning of the analysis period, or to accounts that were opened close to its end. Finally, we removed those time series exhibiting a flat behavior, since this specific pattern is not typically associated with fraud- or anomaly-related profiles. Therefore, the final dataset used in our work consists of 266,388 time series. Some examples of the time series used in this work can be seen in Fig. 1.

## 3. Method

To investigate anomalous behavior in time series, we designed a two-branches methodological framework, both starting from the raw time series, illustrated in Fig. 2.

The first branch starts with the systematic extraction of descriptive features that capture the statistical and temporal characteristics of the time series. In the second stage, we adopt a dual-model approach to anomaly detection, applying both AutoEncoders (AE; Hinton and Zemel, 1993) and Isolation Forests (IF; Liu et al., 2008), to detect anomalous time series in the global feature space (both models will be described in Section 3.2). Their outputs are subsequently compared to assess their performance independently and in combination. This design allows us to explore the robustness of anomaly identification by evaluating areas of agreement and disagreement between the two models. In parallel, a second branch focuses on localizing the source of anomalous behavior within individual series. To this end, we implement a sliding window masking strategy, progressively removing contiguous temporal segments and re-extracting features for each configuration. The masked variants are then analyzed with the same models (AE and IF), allowing us to isolate the specific time segments that contribute to anomalous behavior. Together, these steps provide a comprehensive framework to capture, interpret, and localize anomalies within complex time series.

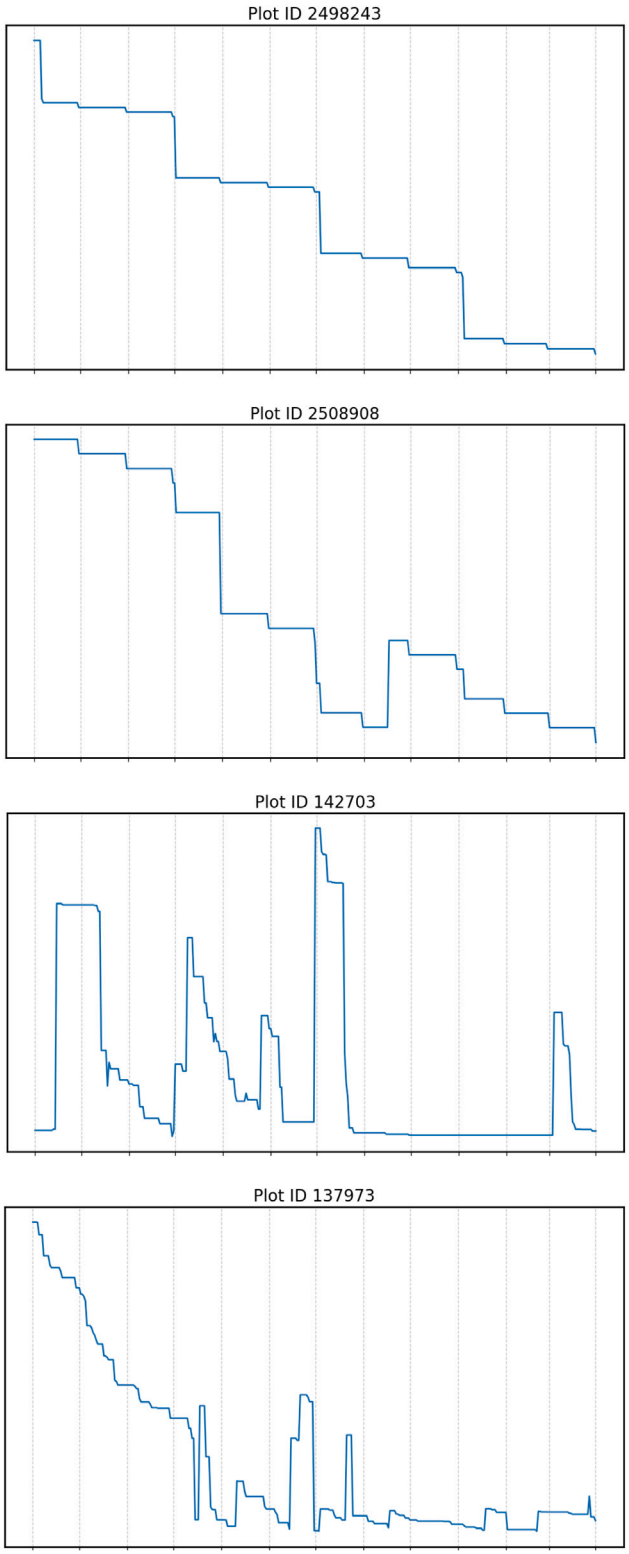


**Fig. 1.** Examples of the time series present in the data set.

### 3.1. Feature engineering

As described in Section 2, our dataset is made up of ∼ 260,000 time series of daily balances. Each original two-year time series was first split into year 1 and year 2. The resulting set of approximately 260,000 series was then divided into two non-overlapping datasets, **A** and **B**, yielding four subsets: **A** Year 1, **A** Year 2, **B** Year 1, and **B** Year 2. The **A** subsets were used for training, while the **B** subsets were reserved for testing.

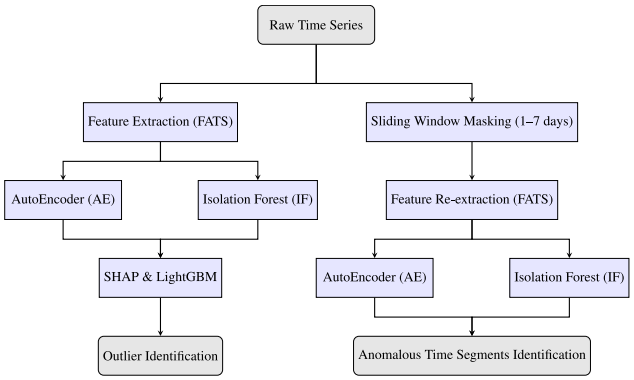


**Fig. 2.** Framework adopted in this study for outlier detection and for identifying anomalies among the outliers.

Each time series was first analyzed using the Python library FATS (Nun et al., 2015), which was used to extract from the time series the set of features, described in Table 1, used for the later analysis.

In Fig. 3 it is shown the distribution of the FATS features computed for the whole dataset and for the representative sample adopted in this work. As it can be seen, the histograms for the whole dataset (in blue) and the subset used in this paper (in red) reveal that the distributions are compatible, ensuring that the sample is representative of the entire dataset.

### 3.2. Outlier identification

To detect global anomalies in the dataset, we first relied on the features extracted from the complete time series using the FATS library. These features provide a compact representation of the statistical and temporal properties of the light curves, enabling the identification of objects that deviate from the bulk distribution in the feature space. For this task, we adopted two complementary unsupervised approaches, namely an AE and an IF. The rationale behind employing both models lies in their distinct anomaly detection mechanisms: while the AE leverages reconstruction error to capture deviations from learned patterns, the IF relies on random partitioning to isolate rare instances. By applying them independently to the same feature set, we can assess the robustness of the results, exploit potential complementarities, and better characterize ambiguous cases that may only be detected by one of the two methods.

An AE (Hinton and Zemel, 1993) is a neural network architecture designed to learn efficient and compressed representations of input data by attempting to reconstruct the original input as accurately as possible. It consists of two main components: an encoder, which progressively reduces the dimensionality of the input data and maps it into a lower dimensional latent representation, and a decoder, which aims to reconstruct the input from this latent space. The encoder and decoder are trained jointly in an unsupervised way, typically by minimizing a reconstruction error (RE), such that the latent representation captures the most relevant information in the data.

The logic behind using the autoencoder for anomaly detection lies in its ability to reconstruct the input data accurately when the data resembles those seen during training. Consequently, time series exhibiting large RE are flagged as anomalous. The RE is computed as the mean squared error (MSE) between the original input sequence and its reconstruction, defined as:

$$\mathbf{RE}_i = \frac{1}{d}\sum_{j=1}^{d}(x_{ij} - \hat{x}_{ij})^2, \tag{1}$$

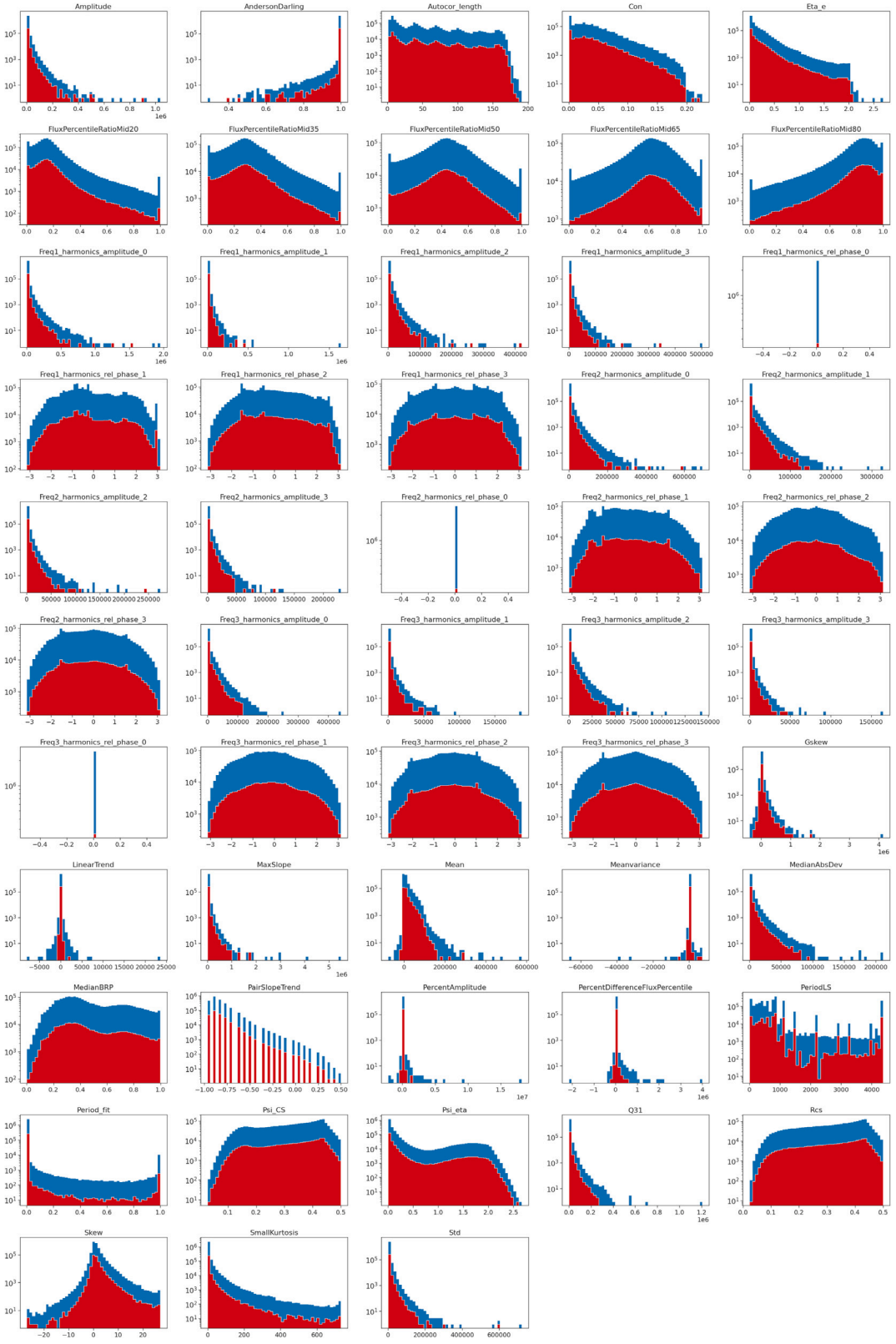


**Fig. 3.** FATS features extracted from the dataset: in blue the distributions of the whole dataset, in red the subset used in this paper.

**Table 1**
List of time series features extracted for this analysis, the indices i and j variates between 1 and 3.

| Feature | Description | Reference |
|---|---|---|
| Amplitude | Half of the difference between the median of the maximum 5% and of the minimum 5% values | Richards et al. (2011) |
| AndersonDarling | Test of whether a sample of data comes from a population with a specific distribution | Nun et al. (2015) |
| Autocor_length | Lag value where the autocorrelation function becomes smaller than $\eta^e$ | Kim et al. (2011) |
| Con | It is defined by the count of the number of three consecutive data points that are above or below $2\sigma$ from the mean, and normalize this number by $N-2$. | Kim et al. (2011) |
| $\eta^e$ | Ratio of the mean of the squares of successive value differences to the variance | Kim et al. (2014) |
| FluxPercentileRatioMid20 | Ratio between the difference between 60% and 40% values, and 95% and 5% values | Richards et al. (2011) |
| FluxPercentileRatioMid35 | Ratio between the difference between 32.5% and 67.5% values, and 95% and 5% values | Richards et al. (2011) |
| FluxPercentileRatioMid50 | Ratio between the difference between 25% and 75% values, and 95% and 5% values | Richards et al. (2011) |
| FluxPercentileRatioMid65 | Ratio between the difference between 17.5% and 82.5% values, and 95% and 5% values | Richards et al. (2011) |
| FluxPercentileRatioMid80 | Ratio between the difference between 10% and 90% values, and 95% and 5% values | Richards et al. (2011) |
| Freq1_harmonics_amplitude_0 | Amplitude of one of the harmonics extracted from light-curves using Lomb–Scargle periodogram | Richards et al. (2011) |
| Freqi_harmonics_amplitude_j | Amplitudes of the harmonics extracted from light-curves using Lomb–Scargle periodogram | Richards et al. (2011) |
| Freqi_harmonics_rel_phase_j | Phases of the harmonics extracted from light-curves using Lomb–Scargle periodogram | Richards et al. (2011) |
| Gskew | Median-based measure of the skew | — |
| LinearTrend | Slope of a linear fit to the light curve | Richards et al. (2011) |
| MaxSlope | Maximum absolute value slope between two consecutive observations | Nun et al. (2015) |
| Mean | Mean value | Nun et al. (2015) |
| Meanvariance | Ratio of the standard deviation to the mean value | Nun et al. (2015) |
| MedianAbsDev | Median discrepancy of the data from the median data | Nun et al. (2015) |
| MedianBRP | Fraction of photometric points within amplitude/10 of the median mag | Nun et al. (2015) |
| PairSlopeTrend | Fraction of increasing first differences minus the fraction of decreasing first differences over the last 30 time-sorted mag measures | Richards et al. (2011) |
| PercentAmplitude | Largest percentage difference between either max or min mag and median mag | Richards et al. (2011) |
| PercentDifferenceFluxPercentile | Ratio of the difference between the 95% and the 5% values over the median value | Nun et al. (2015) |
| PeriodLS | Period derived from the Lomb–Scargle method | Nun et al. (2015) |
| Period_fit | False-alarm probability of the largest periodogram value obtained with LS | Kim et al. (2011) |
| $\Psi_{CS}$ | Range of a cumulative sum applied to the phase-folded light curve | Kim et al. (2011) |
| $\Psi_{\eta^e}$ | $\eta^e$ index calculated from the folded light curve | Kim et al. (2011) |
| Q31 | Difference between the third and the first quartile of the light curve | Kim et al. (2014) |
| Rcs | Range of a cumulative sum | Kim et al. (2011) |
| Skew | Skewness of the time series | Richards et al. (2011) |
| SmallKurtosis | Kurtosis of the light curve. | Richards et al. (2011) |
| Std | Standard deviation of the light curve | Nun et al. (2015) |

where $d$ denotes the dimensionality of the feature space, while $x_{ij}$ and $\hat{x}_{ij}$ represent the reconstructed and target feature value for the $i$th time series. A high MSE indicates that the input pattern significantly deviates from those learned by the model during training, thereby suggesting anomalous behavior. In practice, a threshold must be defined: a time series is classified as anomalous if its corresponding reconstruction error exceeds this threshold, which is commonly set according to the empirical distribution of the errors (e.g., the 95th percentile).

The second algorithm used for detecting outliers within the dataset has been the IF. The IF (Liu et al., 2008) is a machine learning algorithm that identifies anomalies or outliers in data by isolating them through a process of random partitioning within a collection of decision trees, each of them is called Isolation Tree. An Isolation Tree is built by recursively partitioning the data using random splits. Starting from the root node containing a set of instances, the algorithm randomly selects a feature and a split value within the range of that feature for the current node. The data is then divided into two child nodes based on this split. This process continues recursively until each leaf contains a single instance or a predefined maximum depth is reached. The number of splits needed to isolate an instance defines its path length in a tree. The anomalies, which are easier to isolate, typically have shorter paths. By averaging and normalizing these path lengths across all trees in the forest, the algorithm assigns higher anomaly scores to instances that are consistently isolated with fewer splits. Finally, the algorithm assigns an anomaly score between 0 and 1 to each sample, with higher scores indicating more anomalous behavior. Based on a predefined threshold (such as, the parameter *contamination* of the model), the samples are classified as inliers or outliers.

By applying AE and IF independently, we can:

- isolate model-specific behavior: identify which types of anomalies are detected by each model;

- quantify the agreement between the two approaches;
- analyze the disagreement, which may highlight ambiguous cases or limitations in one of the methods.

To this end we implemented the use of the Python library called SHapley Additive exPlanations (SHAP, Lundberg and Lee 2017), a game-theoretic approach designed to explain the output of any machine learning model. In practice, SHAP computes the marginal contribution of each feature by considering all possible subsets of features and measuring how the prediction changes when a given feature is added. This results in a set of additive attributions that are consistent (features that always increase the prediction receive positive contributions and vice versa) and locally accurate (the sum of the contributions equals the model output). However, SHAP requires a predictive model, as the method quantifies how each feature affects the model's numerical output.[2] To meet this requirement, we trained Light Gradient Boosting Machine Regressor (LGBMRegressor) models from the library LightGBM (Ke et al., 2017). LightGBM is a gradient boosting algorithm based on decision trees that adopts a leaf-wise growth strategy, which grants high scalability, reduced computational cost, and strong predictive performance in regression tasks.

### 3.3. Identification of the anomalies within the anomaly

While global anomaly detection can flag entire time series as outliers, it does not indicate which specific segments are responsible for the anomalous behavior. Determining the origin of anomalies within a series is therefore not straightforward and requires a more fine-grained approach. To this end, we employed a multi-day sliding window approach. Rather than extracting features from the entire time series at once, we applied a chronological masking strategy: starting from the first day, we progressively masked contiguous segments of the time series using windows of varying lengths (from 1 to 7 days). After each masking step, features were re-extracted using the FATS library. This process was repeated 2534 times, i.e., the total number of combinations, for each time series (once for every window configuration) resulting in a considerable computational load. The final step of this procedure produced a new dataset containing all combinations of FATS features for each window configuration and time series ID. This additional dataset allowed us to identify which specific segments of the original time series contributed to its anomalous behavior, as the trained models could be directly applied to these masked variants to isolate the anomalous windows. In this way, the approach not only localizes anomalies within the series, but also provides deeper insights into the temporal patterns underlying the observed deviations.

## 4. Experiments

To evaluate the effectiveness of our anomaly detection and localization approach, we conducted two experiments. The first one is focused on applying IF and AE to detect outliers at the global time series level. This allowed to assess the capability of each model to isolate anomalous series without considering their internal temporal structure. The second experiment leverages the multi-day sliding window methodology introduced above. By systematically masking segments of each time series and re-extracting features, we tested the ability of the trained models to pinpoint specific temporal regions responsible for anomalous behavior.

[2] SHAP values are defined with respect to a function f(x) representing the model output. Without a predictive model providing such a function, the marginal contribution of each feature cannot be computed.

**Table 2**
Compact representation of the autoencoder architecture.

| Encoder | Decoder |
|---|---|
| Input (50) | Latent (10) |
| Dense (1024, ReLU) | Dense (128, LeakyReLU) |
| Dropout (0.3) + BatchNorm | Dropout (0.5) + BatchNorm |
| Dense (768, ReLU) | Dense (256, LeakyReLU) |
| Dropout (0.3) + BatchNorm | Dropout (0.5) + BatchNorm |
| Dense (512, ReLU) | Dense (512, LeakyReLU) |
| Dropout (0.3) + BatchNorm | BatchNorm |
| Dense (256, ReLU) | Output (50, Linear) |
| Dropout (0.3) + BatchNorm | |
| Dense (128, ReLU) | |
| Dense (10, Tanh) | |

### 4.1. Experiment with independent IF and AE

The starting point is the dataset described in Section 2. As aforementioned, we split the data into two main datasets (for both years), namely the dataset **A** and dataset **B**: the former has been used for training the machine learning (ML) algorithms, while we made predictions on the latter, for both the years studied. Furthermore, each feature was scaled using a standard scaler, which standardizes features by removing the mean and scaling to unit variance.

We then applied both AE and IF, independently. The Encoder and Decoder are both described in Table 2, which summarizes the architecture of the autoencoder used in this study. The encoder progressively reduces the input dimensionality starting from 50 features, mapping it into a 10-dimensional latent space. Each dense layer applies a nonlinear activation (ReLU) to capture complex patterns in the data. We also used Dropout layers with a probability of firing a neuron of 30% to avoid overfitting, and Batch normalization is applied after most layers to stabilize and accelerate training by normalizing the layer inputs. The decoder reconstructs the original input from the latent representation, using similar techniques but in reverse, with LeakyReLU activations to improve gradient flow and a linear output layer to match the input scale.

We set the anomaly detection threshold as the 95th percentile of the RE distribution over the training set. A time series is then classified as anomalous if its corresponding reconstruction error exceeds this threshold.

The percentage of anomalous time series identified in the test set is 5.42% for the first year. For the second year, applying the same detection logic, the percentage rises slightly to 6.16%. The overlap between the two sets of anomalies is 26.61%, indicating that the majority of time series flagged as anomalous in the first year for a given account are not considered anomalous in the second by the AE.

As done with the autoencoder, we trained the IF on dataset A and evaluated it on dataset B. We set the *contamination* parameter to 5%, which defines the expected proportion of anomalies in the data, and used 100 decision trees in the ensemble.

The 24.44% of the anomalies detected by the IF remain anomalies in the second year too, meaning that most of the points detected as anomalous in the first year, are not labeled as such in the second year

The anomalies detected from both models in the two years are listed in Table 3. Some examples of anomalous time series can be seen in Fig. 4.

To see both the isolated behavior and the agreement between the two models, we used a t-distributed stochastic neighbor embedding algorithm (t-SNE), to highlight the anomalies detected by the two models. t-SNE is a non-linear dimensionality reduction technique that maps high-dimensional data into a two- or three-dimensional space, preserving local neighborhood structures, basically points that were similar in the original space will be mapped closely, while different points will be mapped further away.

As we can see in Figs. 5 and 6 AE detected anomalies tend to be more concentrated at the border of the distributions, while IF detected

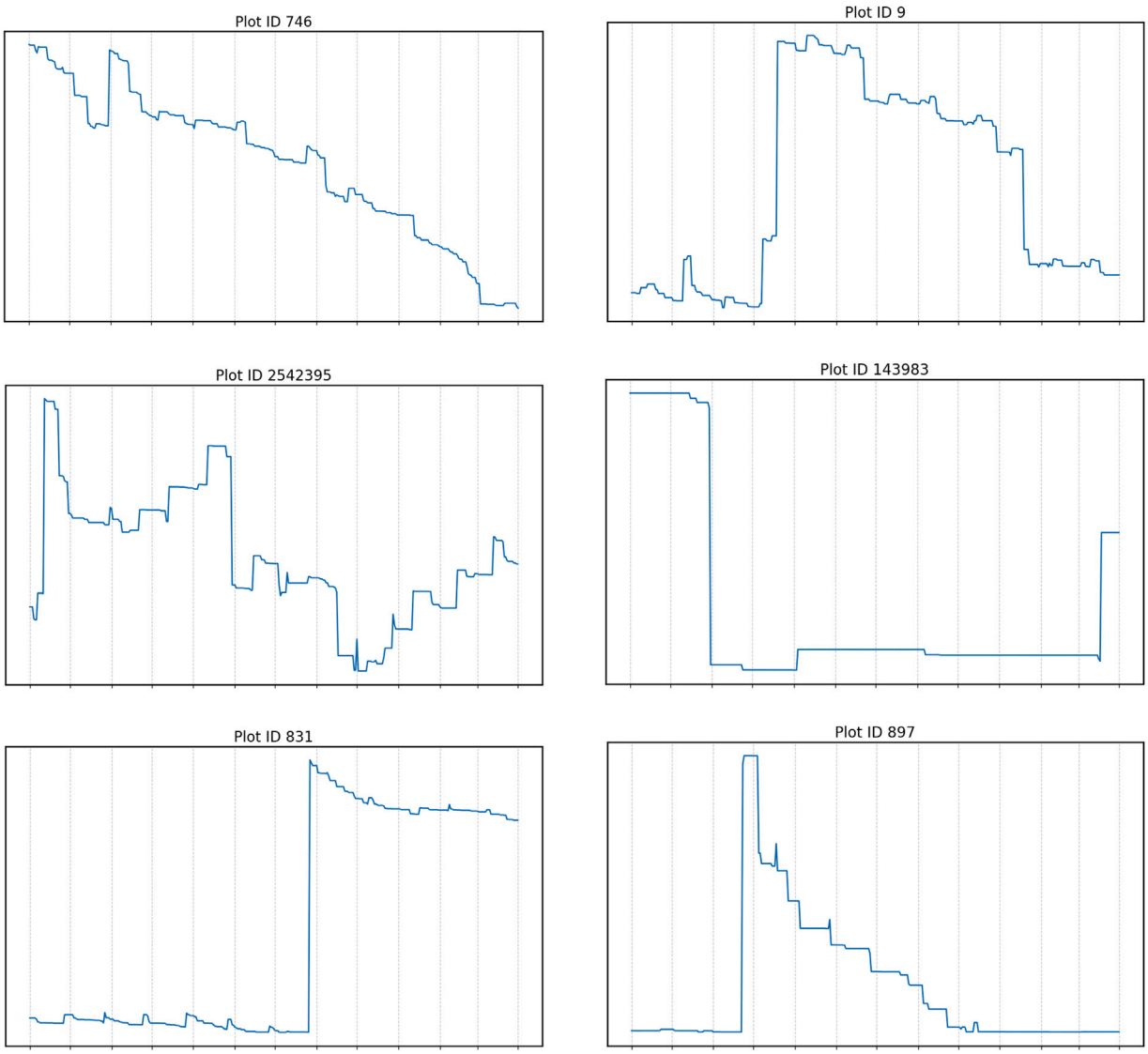


**Fig. 4.** Examples of anomalous time series. The first row shows anomalies detected by both algorithms, the second row contains anomalies identified only by the AE, and the third row displays anomalies detected exclusively by the IF.

**Table 3**
Numerical visualization of the anomalies detected by the two models year per year. This breakdown highlights the degree of overlap and complementarity between the two methods across different observation periods.

| | First year | Second year |
|---|---|---|
| **AE** exclusive | 3003 | 3411 |
| **IF** exclusive | 2503 | 1936 |
| Detected by both | 4219 | 4790 |

anomalies are more sparse. Commonly detected anomalies also tend to cluster near the distribution boundaries. It is important to note that the use of t-SNE for visualization introduces non-linear distortions and should be interpreted as indicative. In Table 3, we can numerically observe the degree of agreement between the two models. And, as we can see, there is a good agreement between the two models for the same year, 64.2% of common anomalies for the first year, and 64.5% for the second year. This suggests that, in the majority of cases, both models detect the same anomalous behavior, despite detecting anomalies in different ways. The remaining $\sim 35\%$ of anomalies not shared between the models may also be informative. These cases could correspond to borderline instances near the detection threshold, the choice of which was arbitrary.

To investigate this possibility, we analyzed the distribution of RE produced by the AE for the entire sample, highlighting both the common and exclusive anomalies. As shown in Fig. 7, we can see that the AE exclusive anomalies are distributed differently with respect to the common ones. Also, if we look at the reconstruction error of the IF exclusive anomalies, we can see that they lie very close to the threshold. We repeated this analysis for the IF model by plotting the IF score distributions for the whole sample, along with the common and exclusive anomalies. Similarly, the scores distribution for the IF exclusive anomalies is different from the common ones, and also, while in general the number of points grows with the scores, the same thing cannot be said for the score relative to the AE exclusive anomalies.

Another explanation may be that they may reflect distinct types of anomalies that are more readily captured by one model than the other. Specifically, if this were the case, using multiple algorithms, such as in this work, could lead to a much broader and complete detection. To investigate the second possibility, we adopted the following strategy: since we are working with unsupervised algorithms, we do not have a clear or immediate explanation for why a certain data point is

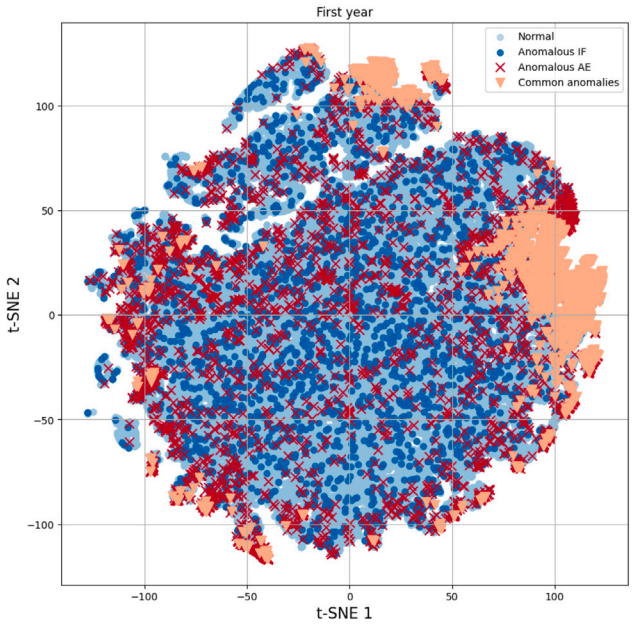


**Fig. 5.** t-SNE visualization of first-year data.

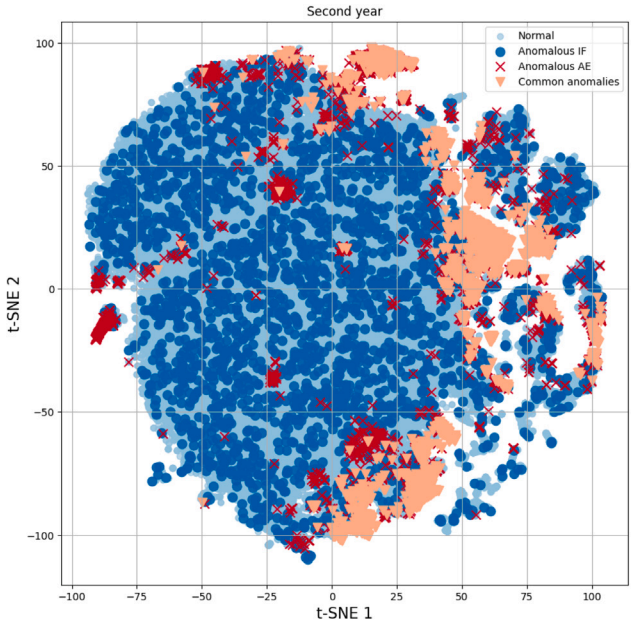


**Fig. 6.** t-SNE visualization of second-year data.

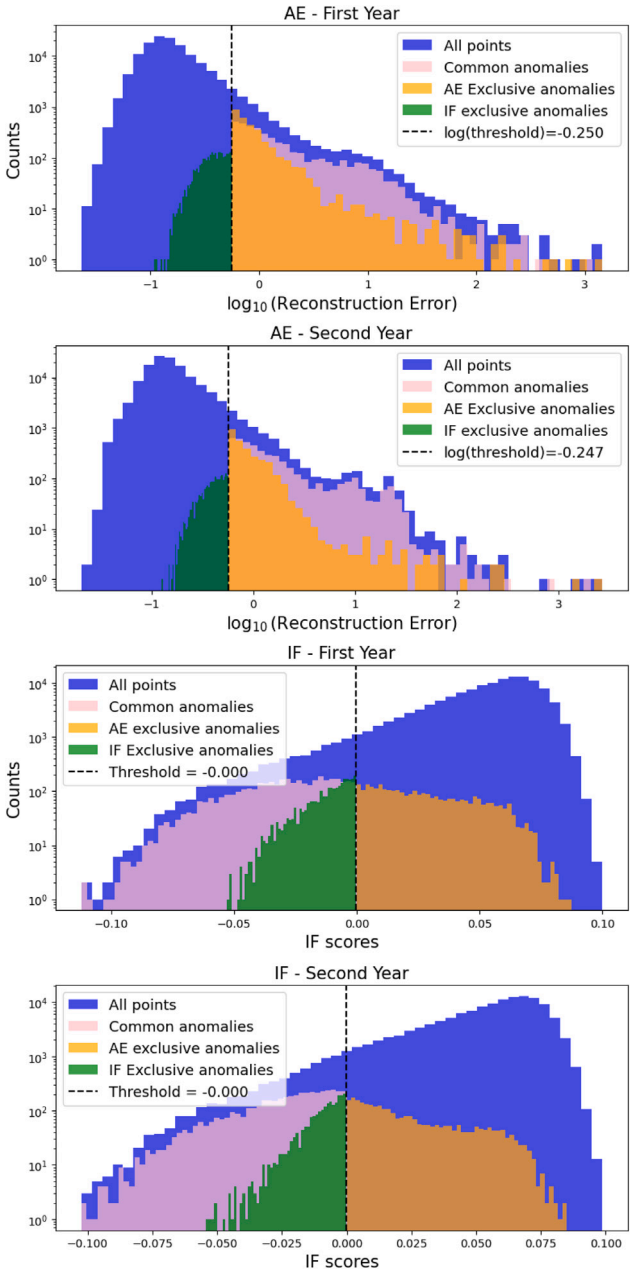


**Fig. 7.** From top to bottom: the first two panels show the reconstruction error distributions for the two years studied; points to the right of the black dashed lines are considered anomalous by the Autoencoder (AE). The bottom two panels show the Isolation Forest (IF) score distributions for the same two years; points to the left of the black dashed lines are considered anomalous according to the IF.

classified as anomalous, nor do we know which features contribute the most to this decision. To overcome this limitation we used SHAP in combination with a LightGBM regressor. We used the RE for the autoencoder and the anomaly score for the IF as target variables for the LightGBM, and we interpreted the results via SHAP.

In total, we trained four LGBMRegressor models, two for each year, and applied SHAP to interpret their predictions on the exclusive anomalies. As shown in Figs. 8 and 9, which display the top 10 most influential features, we observe that different sets of features characterize the anomalies detected exclusively by the two models.

Furthermore, a clear pattern seems to emerge: the AE consistently highlights features related to periodicity and harmonic content, such as Freq1_harmonics_rel_phase_1, Freq3_harmonics_amplitude_1, and Fre q2_harmonics_amplitude_2. These features suggest that the AE is particularly sensitive to deviations from the periodic or quasi-periodic structures in the light curves. In contrast, the IF highlights features more indicative of abrupt changes or statistical outliers, such as MaxSlope, Amplitude, and LinearTrend. These features reflect sudden variations or global deviations, consistent with the density-based nature of the IF algorithm.

Taken together, this evidence strongly suggests that the two sets of exclusive anomalies are qualitatively different. Therefore, using both models in combination enhances the breadth and completeness of anomaly detection. These differences could plausibly be connected to broader financial conditions affecting all accounts, although we cannot

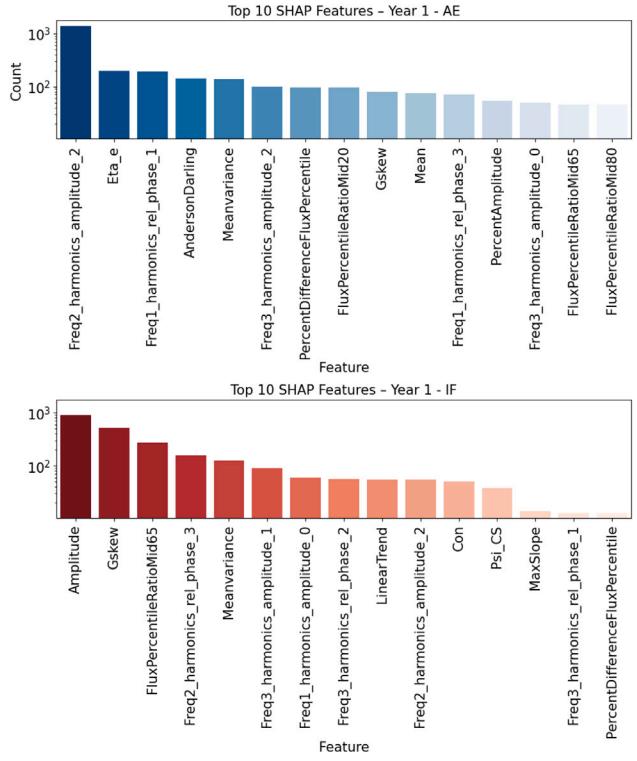


**Fig. 8.** Comparison of the most influential features for the exclusive features of the two models in first year.

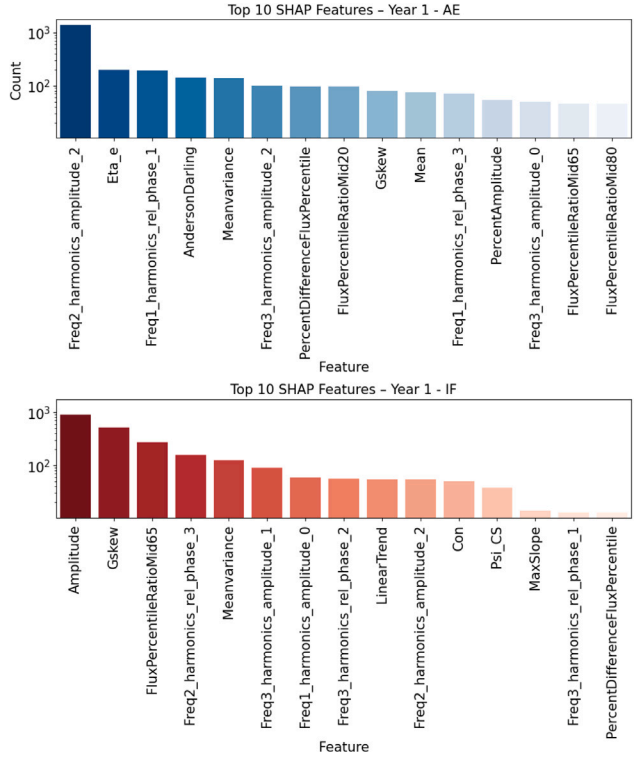


**Fig. 9.** Comparison of the most influential features for the exclusive features of the two models in the second year.

**Table 4**
Summary of time series with classification changes according to IF and AE models. Percentages are relative to the total sample ($N = 4652$).

| Type | Inliers (%) | Outliers (%) | Total (%) |
|---|---|---|---|
| Any change | 224 (4.82%) | 759 (16.32%) | 983 (21.13%) |
| Only IF | 109 (2.34%) | 300 (6.45%) | 409 (8.79%) |
| Only AE | 83 (1.78%) | 100 (2.15%) | 183 (3.93%) |
| Both models | 32 (0.69%) | 277 (7.72%) | 391 (8.40%) |

assert this with certainty. While these differences could plausibly reflect broader financial conditions affecting all accounts, this connection cannot be confirmed with certainty.

### 4.2. Experiments with multi-day sliding window

As described in Section 3, to fully characterize the model outputs, it is essential to identify which specific segments of the time series lead the models to classify them as anomalous. To achieve this, we performed a multi-day sliding window analysis: we applied sliding windows of varying sizes (ranging from one to seven days) to each time series. Beginning on day one, we progressively masked segments of the series using the sliding window, removing the corresponding points in chronological order. For each modified version of the time series, we recomputed the full set of features listed in Table 1 using the FATS library. This procedure resulted in 2534 different feature sets per time series, one for each masked version of the time series. For each representation of these sources, we performed a prediction with the models being trained on the dataset **A**, and at each iteration, we confronted the results of the prediction on the current representation with the previous one. Whenever the classification changed (i.e., from inlier to outlier or vice versa), we flagged the corresponding time window that had been masked. This allowed us to identify which specific periods of the time series were the most influential in altering the models' predictions.

From the initial sample, we extracted a subset of 2317 outliers and 2335 inliers, which showed consistent predictions for both models used above (AE and IF). Out of the 4652 time series analyzed, 983 of them showed at least a state change during the process. Of these 983, 759 ($\sim 77\%$) were originally labeled as outliers by the two models (16.32% of the original sample studied), and 224 were originally labeled as inliers by the two models (4.82% of the original sample). Now focusing on how the two models tackle the problem, we have that: 409 time series show changes only for the IF (8.79% of the total; thus divided: 109 inliers and 300 outliers), 183 time series show changes only for the AE (3.93% of the total, thus divided: 83 inliers and 100 outliers), and 391 time series show changes for both models (8.40% of the total; thus divided: 32 inliers and 359 outliers). All these results are summarized in Table 4.

To test whether the algorithm shows any preferences in detecting changes in any of the two classes, we characterized the probability distributions associated with the number of changes, we used both the Kolmogorov–Smirnov (KS; Hodges (1958)) test and the t-test (Student, 1908) from the scipy Python library. Since the number of outliers is overall 4 times larger than the inliers, we also used the same test by comparing the distribution of the original inliers and the distribution of outliers sampled every four. The PDFs are shown in Fig. 10, which visually highlights the similarity between them. The results obtained by both tests in both cases are presented in Table 5. As can be seen in both cases, the two distributions are statistically indistinguishable. These results suggest that the algorithm does not show any preferences or bias towards either of the two classes.

Among all the classes defined in Table 4, we have a particularly interesting class, the originally inliers, which exhibit changes in both models used. This class is particularly interesting not only because we are talking about inliers, which is interesting by itself, since we do

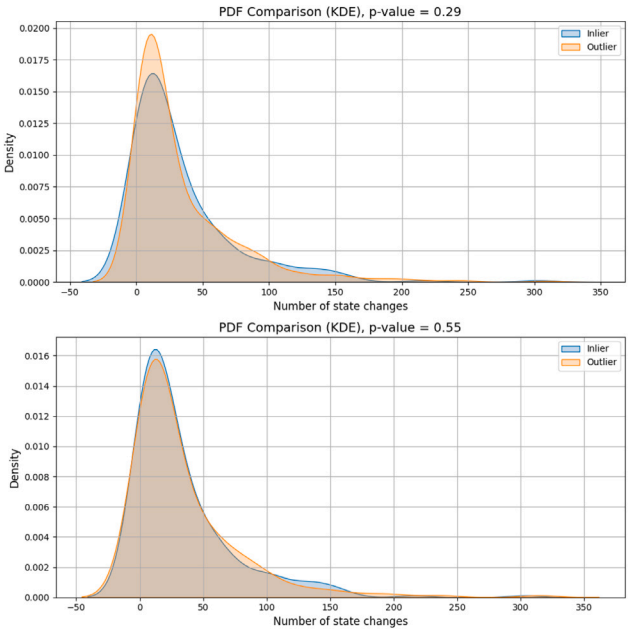


**Fig. 10.** PDFs for number of changes. Top: the PDFs for the whole populations; bottom: the PDFs for the whole population of inliers plus the population of outliers sampled every four.

**Table 5**
Results of statistical tests (KS and t-test) comparing the PDFs of the number of changes for inliers and outliers. The top two rows refer to the full distributions of inliers and outliers; the bottom two rows compare the inlier distribution with that of outliers sampled every four.

| Comparison | Test statistic | p-value |
|---|---|---|
| Full distributions (KS test) | 0.0738 | $2.86 \times 10^{-1}$ |
| Full distributions (t-test) | −0.1337 | $8.94 \times 10^{-1}$ |
| Inliers vs sampled outliers (KS test) | 0.0766 | $5.48 \times 10^{-1}$ |
| Inliers vs sampled outliers (t-test) | −0.2435 | $8.08 \times 10^{-1}$ |

not expect them to show any status changes, but also because they can highlight the use of this algorithm as a complementary means to anomaly detection. Specifically, this approach can highlight anomalous behavior that otherwise might not have been detected by conventional means. Some examples of the algorithm results can be seen in Fig. 11.

In general, different types of local anomalies can be observed. In some cases, the detected windows align with sharp structural changes, where the series shifts abruptly to a new level. In other cases, the anomalies are concentrated in periods of irregular variability, contrasting with the otherwise smooth trend of the series. More complex situations are also observed, where multiple bursts of anomalous behavior are distributed across the series, reflecting recurrent irregular dynamics rather than a single deviation. Overall, these examples highlight how the sliding window approach can disentangle the specific temporal segments responsible for the anomalous classification, thereby refining the interpretability of anomaly detection beyond a purely global assessment.

## 5. Discussion

Time series are a common type of data found in many real-world scenarios, such as traffic flow, network KPIs, and financial data, all of which can be represented as time series (Lei et al., 2023). Anomaly detection in time series is an interesting research topic in the data mining field, with a wide range of real world applications, such as manufacturing, healthcare, finance, and so on (Salman et al., 2025;

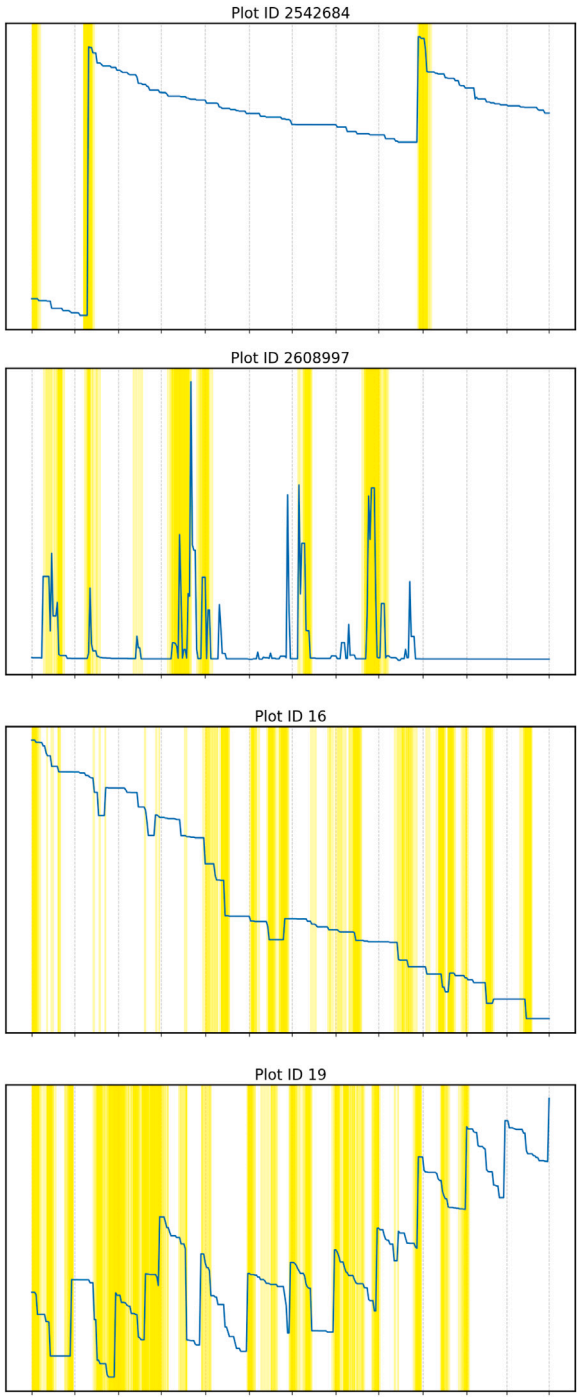


**Fig. 11.** Examples of regions where the sliding window algorithm detects changes for two different time series. The yellow shaded areas highlight the time intervals where the sliding window analysis led to a classification change.

Maruccia et al., 2025b; Engbers and Freitag, 2024; Lei et al., 2023). In this study, we presented a general-purpose methodology for anomaly detection in time series data, adapting a feature engineering framework originally developed for astrophysical applications (De Cicco et al., 2019; Sánchez-Sáez et al., 2021; De Cicco et al., 2021, 2025; Maruccia et al., 2025c) to the financial domain Maruccia et al. (as also shown in 2025b). By combining feature engineering with the use of unsupervised algorithms (specifically, IF and AE), we were able to successfully identify atypical patterns, like those ones shown in Fig. 4. In addition, we introduced an extension based on adaptive temporal

windows, which allowed us to localize the anomalies within the time series, moving beyond a purely global characterization of anomalous objects. Some examples have been shown in Fig. 11, where the highlighted segments correspond to specific intervals contributing to the anomalous behavior. These results illustrate the added value of the sliding window strategy where, rather than identifying only whether a time series is anomalous or not, the method highlights the specific temporal segments that drive the anomalous behavior. Identifying the causes behind anomalies is a difficult task, nevertheless it is crucial for actionable insights. As seen before, time series data contain many different patterns (Li et al., 2022), each of them often requiring various detection methods (Lei et al., 2023). Therefore, understanding why a time series exhibits anomalous behavior allows practitioners not only to detect irregularities but also to interpret their significance, anticipate potential risks, and design appropriate interventions. This aligns with recent research emphasizing that explainability in multivariate time series is essential, as small changes in one variable can propagate and trigger larger anomalies elsewhere, and interpreting these relationships is a key step in early detection and risk mitigation (e.g., Li and Jung, 2023).

The results presented in this work highlight the adaptability of our approach in detecting anomalies beyond its original astrophysical context, bridging the gap between astrophysical studies and finance, and underscoring its cross-domain transferability. A key aspect of our methodology is, in fact, its domain-agnostic design: by relying on feature engineering, time series of diverse origin can be transformed into a unified representation, enabling the application of unsupervised algorithms without requiring prior domain-specific labels. This aspect is particularly relevant for large, heterogeneous datasets where label annotation is infeasible, and it strengthens the potential of the framework as a general-purpose strategy.

The comparative performance of IF and AE within this framework further illustrates the flexibility of the approach. IF provides a computationally efficient, interpretable baseline, while AE captures more subtle, non-linear relationships in the feature space (Berahmand et al., 2024). Together, these methods demonstrate complementary strengths, suggesting that hybrid strategies may further enhance anomaly detection capabilities.

Another notable outcome concerns the transferability of knowledge between domains. The successful adaptation of a framework originally designed for astrophysical light curves to financial time series not only validates the underlying principles but also illustrates how methodological advances in one scientific field can benefit another. This cross-disciplinary applicability highlights the growing importance of methodological convergence in data-intensive sciences, where shared challenges such as irregular sampling, incomplete data, and the scarcity of labeled examples recur (Lei et al., 2023).

This study offers significant implications for both practitioners and researchers. For practitioners, the proposed framework provides a practical and scalable tool for anomaly detection in time series where labeled ground truth is unavailable or incomplete. Its reliance on feature engineering ensures interpretability, a crucial aspect in domains such as finance, where regulatory compliance and explainability are essential, as well as in astrophysics, where physical insight must guide the interpretation of detected anomalies. The introduction of adaptive temporal windows further enables practitioners to pinpoint anomalous behavior at finer temporal scales, supporting actionable decisions in operational settings. In fact, the ability to localize anomalies in specific intervals is crucial in real-world applications, where events of interest are often transient and context-dependent. For researchers, this work underscores the value of cross-domain methodological transfer, demonstrating that techniques originally developed for astrophysical variability can be successfully adapted to financial datasets. It also opens avenues for further investigations into hybrid anomaly detection strategies that combine feature-based and deep representation learning approaches. Moreover, the adaptive windowing concept provides a fertile ground for methodological innovation, inviting exploration of automated optimization schemes and domain-informed strategies to enhance temporal resolution.

## 6. Conclusions

The methods and techniques used in this paper offer a new combined approach in the field of anomaly detection for time series. The combined use of IF and AE allows for a more comprehensive and diversified detection of anomalous behavior across the dataset. Our comparative analysis shows that there is a good overlap between the two models: $\sim 64\%$ of common anomalies, and using both algorithms can help us recover also the anomalies which may be detected by just one of the two.

Furthermore, the SHAP analysis of the feature can help in understanding which features are the most influential in the final classification, possibly helping in detecting systematic behaviors or particularly interesting cases.

Nonetheless, several limitations deserve attention. It is worth mentioning that this kind of feature analysis is computationally expensive, especially when applied to complex models or large datasets. One possible mitigation for the problem would be to apply this analysis only to a representative subset of the data, which will surely reduce the computational time. The sliding-window procedure is another computationally demanding step, as it requires re-extracting features for each window configuration. In our experiments, the processing time for a single user averaged approximately 35 min, which is manageable at moderate scales but becomes prohibitive when extended to very large datasets or real-time applications. This highlights the need for more efficient implementations, such as parallelized pipelines, dimensionality reduction prior to feature extraction, or the adoption of approximate methods that preserve localization accuracy while significantly reducing runtime. Future work should also explore automated strategies for window optimization, possibly informed by domain knowledge or data-driven heuristics, to balance precision and computational tractability.

Finally, our study points towards several promising research directions. In astrophysics, the integration of adaptive windowing into large-scale surveys could refine the detection of transient events and rare variable sources, improving the precision of anomaly localization. In finance, extending the framework to real-time transaction monitoring systems could enhance fraud detection capabilities, particularly when combined with ensemble methods or semi-supervised approaches that leverage partially available labels. More broadly, the growing availability of large and heterogeneous time series datasets across domains suggests that feature-based anomaly detection, complemented by adaptive representations, may become a cornerstone of scalable, interpretable, and transferable methodologies in data-driven science.

## CRediT authorship contribution statement

**N. De Bonis:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Data curation. **S. Vaccaro:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Data curation. **Y. Maruccia:** Writing – review & editing, Writing – original draft, Software, Project administration, Data curation, Conceptualization. **G. Riccio:** Writing – review & editing, Supervision, Conceptualization. **R. Crupi:** Writing – review & editing, Project administration, Data curation, Conceptualization. **S. Rubini:** Writing – review & editing, Project administration, Data curation. **D. De Cicco:** Writing – review & editing, Supervision, Project administration, Conceptualization. **M. Brescia:** Writing – review & editing, Supervision, Conceptualization. **S. Cavuoti:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This paper is supported by Italian Research Center on High Performance Computing Big Data and Quantum Computing (ICSC), project funded by European Union - NextGenerationEU - and National Recovery and Resilience Plan (NRRP) - Mission 4 Component 2 within the activities of Spoke 3 (Astrophysics and Cosmos Observations). The authors acknowledge Intesa Sanpaolo S.p.A. The views and opinions expressed are those of the authors and do not necessarily reflect the views of Intesa Sanpaolo, its affiliates or its employees. We acknowledge the use of the ADHOC (Astrophysical Data HPC Operating Center) resources, within the project "Strengthening the Italian Leadership in ELT and SKA (STILES)", proposal nr. IR0000034, admitted and eligible for funding from the funds referred to in the D.D. prot. no. 245 of August 10, 2022 and D.D. 326 of August 30, 2022, funded under the program "Next Generation EU" of the European Union, "Piano Nazionale di Ripresa e Resilienza" (PNRR) of the Italian Ministry of University and Research (MUR), "Fund for the creation of an integrated system of research and innovation infrastructures", Action 3.1.1 "Creation of new IR or strengthening of existing IR involved in the Horizon Europe Scientific Excellence objectives and the establishment of networks".

## References

Álvarez, M., Dafonte, C., Manteiga, M., Garabato, D., Santoveña, R., Pallas, L., 2025. Unsupervised clustering visualisation tool for gaia DR3. In: McIver, J., Mahabal, A., Fluke, C. (Eds.), In: IAU Symposium, vol. 19, pp. 98–100. doi:10.1017/S1743921323000716.

Armstrong, D.J., Kirk, J., Lam, K.W.F., McCormac, J., Osborn, H.P., Spake, J., Walker, S., Brown, D.J.A., Kristiansen, M.H., Pollacco, D., West, R., Wheatley, P.J., 2016. K2 variable catalogue - II. Machine learning classification of variable stars and eclipsing binaries in K2 fields 0-4. Mon. Not. R. Astron. Soc. 456 (2), 2260–2272. doi:10.1093/mnras/stv2836, arXiv:1512.01246.

Bellm, E.C., Kulkarni, S.R., Graham, M.J., Dekany, R., Smith, R.M., Riddle, R., Masci, F.J., Helou, G., Prince, T.A., Adams, S.M., et al., 2018. The zwicky transient facility: system overview, performance, and first results. Publ. Astron. Soc. Pac. 131 (995), 018002.

Berahmand, K., Daneshfar, F., Salehi, E.S., Li, Y., Xu, Y., 2024. Autoencoders and their applications in machine learning: a survey. Artif. Intell. Rev. 57 (2), 28.

Campagne, J.-E., 2025. Galaxy imaging with generative models: insights from a two-models framework. Mon. Not. R. Astron. Soc. 539 (4), 3445–3458.

Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection: A survey. ACM Comput. Surv. 41 (3), 1–58.

De Cicco, D., Bauer, F., Paolillo, M., Cavuoti, S., Sánchez-Sáez, P., Brandt, W., Pignata, G., Vaccari, M., Radovich, M., 2021. A random forest-based selection of optically variable AGN in the VST-COSMOS field. Astron. Astrophys. 645, A103.

De Cicco, D., Paolillo, M., Covone, G., Falocco, S., Longo, G., Grado, A., Limatola, L., Botticella, M.T., Pignata, G., Cappellaro, E., et al., 2015. Variability-selected active galactic nuclei in the VST-SUDARE/VOICE survey of the COSMOS field. Astron. Astrophys. 574, A112.

De Cicco, D., Paolillo, M., Falocco, S., Poulain, M., Brandt, W., Bauer, F., Vagnetti, F., Longo, G., Grado, A., Ragosta, F., et al., 2019. Optically variable AGN in the three-year VST survey of the cosmos field. Astron. Astrophys. 627, A33.

De Cicco, D., Zazzaro, G., Cavuoti, S., Paolillo, M., Longo, G., Petrecca, V., Saccheo, I., Sánchez-Sáez, P., 2025. Selection of optically variable active galactic nuclei via a random forest algorithm. Astron. Astrophys. 697, A204.

Engbers, H., Freitag, M., 2024. Automated model selection for multivariate anomaly detection in manufacturing systems. J. Intell. Manuf. 1–19.

Fotopoulou, S., 2024. A review of unsupervised learning in astronomy. Astron. Comput. 48, 100851. doi:10.1016/j.ascom.2024.100851.

Gebran, M., Bentley, I., 2025. The use of conditional variational autoencoders in generating stellar spectra. Astronomy 4 (3), 13.

Gupta, R., Muthukrishna, D., Lochner, M., 2025. A classifier-based approach to multiclass anomaly detection for astronomical transients. RAS Tech. Instruments 4, rzae054.

Hinton, G.E., Zemel, R., 1993. Autoencoders, minimum description length and Helmholtz free energy. Adv. Neural Inf. Process. Syst. 6.

Hodges, J.L., 1958. The significance probability of the smirnov two-sample test. Ark. Mat. 3 (5), 469–486. doi:10.1007/BF02589501.

Ivezić, Ž., Kahn, S.M., Tyson, J.A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S.F., Andrew, J., et al., 2019. LSST: from science drivers to reference design and anticipated data products. Astrophys. J. 873 (2), 111.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30.

Keith, M.J., Jameson, A., Van Straten, W., Bailes, M., Johnston, S., Kramer, M., Possenti, A., Bates, S.D., Bhat, N.D.R., Burgay, M., Burke-Spolaor, S., D'Amico, N., Levin, L., McMahon, P.L., Milia, S., Stappers, B.W., 2010. The high time resolution universe pulsar survey - I. System configuration and initial discoveries: HTRU - I. System configuration. Mon. Not. R. Astron. Soc. 409 (2), 619–627. doi:10.1111/j.1365-2966.2010.17325.x.

Kim, D.W., Protopapas, P., Bailer-Jones, C.A., Byun, Y.I., Chang, S.W., Marquette, J.B., Shin, M.S., 2014. The EPOCH project-I. Periodic variable stars in the EROS-2 LMC database. Astron. Astrophys. 566, A43.

Kim, D.W., Protopapas, P., Byun, Y.I., Alcock, C., Khardon, R., Trichas, M., 2011. Quasi-stellar object selection algorithm using time variability and machine learning: Selection of 1620 quasi-stellar object candidates from MACHO large magellanic cloud database. Astrophys. J. 735 (2), 68. doi:10.1088/0004-637X/735/2/68, arXiv:1101.3316.

Kohonen, T., 2001. Self-organizing maps, third ed. In: Springer Series in Information Sciences, vol. 30, Springer, Berlin, URL http://www.worldcat.org/isbn/3540679219.

Lei, T., Gong, C., Chen, G., Ou, M., Yang, K., Li, J., 2023. A novel unsupervised framework for time series data anomaly detection via spectrum decomposition. Knowl.-Based Syst. 280, 111002.

Li, G., Jung, J.J., 2023. Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. Inf. Fusion 91, 93–102.

Li, Y., Shen, D., Nie, T., Kou, Y., 2022. A new shape-based clustering algorithm for time series. Inform. Sci. 609, 411–428.

Liu, F.T., Ting, K.M., Zhou, Z.H., 2008. Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining. pp. 413–422. doi:10.1109/ICDM.2008.17.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), In: Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., pp. 4765–4774, URL http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

Maruccia, Y., Cavuoti, S., Brescia, M., Riccio, G., Molinari, S., Elia, D., Schisano, E., 2025a. The evolutionary path of star-forming clumps in Hi-GAL. Astron. Comput. 100985.

Maruccia, Y., Cavuoti, S., Crupi, R., Riccio, G., Cicco, D.D., Barone, P., Rubini, S., Koleci, A., Pavanelli, P., Sabatino, A.D., Desiante, R., 2025b. Anomaly detection in the banking sector using feature engineering. Financ. Innov. Submitted.

Maruccia, Y., De Cicco, D., Cavuoti, S., Riccio, G., Sánchez-Sáez, P., Paolillo, M., Borrelli, N.L., Crupi, R., Brescia, M., 2025c. Navigating AGN variability with self-organizing maps. Astron. Astrophys. 699, A303. doi:10.1051/0004-6361/202553866.

Nawaz, A., Khan, S.S., Ahmad, A., 2024. Ensemble of autoencoders for anomaly detection in biomedical data: A narrative review. IEEE Access 12, 17273–17289. doi:10.1109/ACCESS.2024.3360691.

Nun, I., Protopapas, P., Sim, B., Zhu, M., Dave, R., Castro, N., Pichara, K., 2015. FATS: Feature analysis for time series. arXiv:1506.00010.

Padmanabh, P.V., Barr, E.D., Sridhar, S.S., Rugel, M.R., Damas-Segovia, A., Jacob, A.M., Balakrishnan, V., Berezina, M., Bernadich, M.C., Brunthaler, A., Champion, D.J., Freire, P.C.C., Khan, S., Klöckner, H.-R., Kramer, M., Ma, Y.K., Mao, S.A., Men, Y.P., Menten, K.M., Sengupta, S., Venkatraman Krishnan, V., Wucknitz, O., Wyrowski, F., Bezuidenhout, M.C., Buchner, S., Burgay, M., Chen, W., Clark, C.J., Künkel, L., Nieder, L., Stappers, B., Legodi, L.S., Nyamai, M.M., 2023. The MPIfR–MeerKAT galactic plane survey – I. System set-up and early results. Mon. Not. R. Astron. Soc. 524 (1), 1291–1315. doi:10.1093/mnras/stad1900.

Parimucha, Š., Gabdeev, M., Markus, Y., Vaňko, M., Gajdoš, P., 2025. Morphological classification of eclipsing binary stars using computer vision methods. Astron. Comput. 53, 100998. doi:10.1016/j.ascom.2025.100998, arXiv:2508.12802.

Pinciroli Vago, N.O., Amato, R., Imbrogno, M., Israel, G.L., Belfiore, A., Kovlakas, K., Fraternali, P., Pasquato, M., 2025. The hunt for new pulsating ultraluminous X-ray sources: A clustering approach. Astron. Astrophys. 701, A96. doi:10.1051/0004-6361/202553739, arXiv:2507.15032.

Prusti, T., De Bruijne, J., Brown, A.G., Vallenari, A., Babusiaux, C., Bailer-Jones, C., Bastian, U., Biermann, M., Evans, D.W., Eyer, L., et al., 2016. The gaia mission. Astron. Astrophys. 595, A1.

Rajendran, T., Mohamed Imtiaz, N., Jagadeesh, K., Sampathkumar, B., 2024. Cybersecurity threat detection using deep learning and anomaly detection techniques. ICKECS, In: 2024 International Conference on Knowledge Engineering and Communication Systems, vol. 1, pp. 1–7. doi:10.1109/ICKECS61492.2024.10617347.

Rao, S., Mahabal, A., Rao, N., Raghavendra, C., 2021. Nigraha: Machine-learning-based pipeline to identify and evaluate planet candidates from TESS. Mon. Not. R. Astron. Soc. 502 (2), 2845–2858. doi:10.1093/mnras/stab203, arXiv:2101.09227.

Richards, J.W., Starr, D.L., Butler, N.R., Bloom, J.S., Brewer, J.M., Crellin-Quick, A., Higgins, J., Kennedy, R., Rischard, M., 2011. On machine-learned classification of variable stars with sparse and noisy time-series data. Astrophys. J. 733 (1), 10.

Salman, A.M., Al-Nuaimi, B.T., Subhi, A.A., Alkattan, H., Alfilh, R.H., 2025. Enhancing cybersecurity with machine learning: A hybrid approach for anomaly detection and threat prediction. Mesopotamian J. CyberSecurity 5 (1), 202–215.

Sánchez-Sáez, P., Reyes, I., Valenzuela, C., Förster, F., Eyheramendy, S., Elorrieta, F., Bauer, F., Cabrera-Vives, G., Estévez, P., Catelan, M., et al., 2021. Alert classification for the ALeRCE broker system: The light curve classifier. Astron. J. 161 (3), 141.

Sriraam, N., Chinta, B., Seshadri, S., Suresh, S., 2025. A comprehensive review of artificial intelligence-based algorithm towards fetal facial anomalies detection (2013–2024). Artif. Intell. Rev. 58 (5), 1–49.

Student, 1908. The probable error of a mean. Biometrika 6 (1), 1–25. doi:10.1093/biomet/6.1.1.

Tang, Y., Fan, L., Wan, Z., Liu, Y., Lu, Y., 2025. Real-time light curve classification framework for the wide field survey telescope using modified semisupervised variational autoencoder. Astron. J. 169 (6), 304.

Wang, L.H., Li, K., Gao, X., Guo, Y.N., Sun, G.Y., 2025. Using machine learning method for variable star classification using the TESS sectors 1–57 data. Astrophys. J. 986 (1), 19. doi:10.3847/1538-4357/add159, arXiv:2504.00347.

Webb, S., Lochner, M., Muthukrishna, D., Cooke, J., Flynn, C., Mahabal, A., Goode, S., Andreoni, I., Pritchard, T., Abbott, T.M., 2020. Unsupervised machine learning for transient discovery in deeper, wider, faster light curves. Mon. Not. R. Astron. Soc. 498 (3), 3077–3094.

Xie, G., Wang, J., Liu, J., Lyu, J., Liu, Y., Wang, C., Zheng, F., Jin, Y., 2024. Im-iad: Industrial image anomaly detection benchmark in manufacturing. IEEE Trans. Cybern. 54 (5), 2720–2733.

Zhu, S.Y., Shao, L., Tam, P.H.T., Zhang, F.W., 2025. Unsupervised machine learning classification of gamma-ray bursts based on the rest-frame prompt emission parameters. arXiv e-prints, arXiv:2509.08224.