

# Unsupervised Anomaly Detection in TESS Light Curves Using an Uncertainty-aware RNN-VAE and Consensus Ensemble of Latent-space Outlier Detectors

William Zhang<sup>1</sup> & Daniel Muthukrishna<sup>2</sup>

Received August 2, 2025

Accepted February 28, 2026

Electronic access March 31, 2026

Astronomical light curves provide a high-cadence record of stellar brightness variability, but the scale of modern surveys makes systematic discovery of rare behaviors difficult without automated methods. We present an unsupervised anomaly-detection pipeline for processed Transiting Exoplanet Survey Satellite (TESS) light curves that combines representation learning with complementary outlier detectors. Starting from 3,659 processed light curves produced by an MIT lab preprocessing pipeline, we apply non-imputing missing-value handling and a minimum-information filter ( $< 20$  valid paired samples) to obtain 3,593 light curves for analysis. A recurrent variational autoencoder (RNN-VAE) is trained to learn a compact latent representation; anomalies are scored using (i) VAE reconstruction error (a proxy for low reconstruction likelihood), (ii) density-based outlier structure in standardized latent space using HDBSCAN, and (iii) isolation scores in the standardized latent space using Isolation Forest. We select operating points via sensitivity analyses over reconstruction thresholds (across training epochs), HDBSCAN hyperparameters, and Isolation Forest contamination, prioritizing shortlist stability under small perturbations. Under the selected settings, reconstruction scoring flags 398 objects (11.1%), HDBSCAN flags 256 (7.1%), and Isolation Forest flags 230 (6.4%). An agreement-based ensemble reduces method-specific false positives and yields 226 consensus candidates (6.29%) flagged by at least two detectors, including a strict subset of 116 objects (3.23%) flagged by all three. We provide a reproducible workflow, robustness criteria, and a manual vetting checklist to support scientific follow-up of ensemble-selected candidates.

**Keywords:** anomaly detection, light curves, RNN-VAE, HDBSCAN, isolation forest

## Introduction

Astronomical time-domain surveys now deliver light curves at a scale where rare and scientifically valuable behaviors can be missed by manual inspection alone. In the Transiting Exoplanet Survey Satellite (TESS) mission, instrumental systematics (e.g., sector-dependent trends, discontinuities, and data gaps) are interleaved with astrophysical variability, complicating rule-based screening and making labeled anomaly sets scarce<sup>1–5</sup>. Unsupervised anomaly detection provides a natural alternative: instead of relying on predefined templates, models can learn the dominant structure of the data and flag observations that deviate from it. In this study, we analyze 3,659 processed TESS light curves provided by a Massachusetts Institute of Technology (MIT) lab processing pipeline and apply filtering designed to avoid imputation-driven artifacts and low-information series, yielding a final working set of 3,593 light curves for scoring and sensitivity analysis. We

develop an end-to-end pipeline that (i) applies per-object robust clipping and normalization to preserve intrinsic morphology while controlling scale variation and outliers, (ii) learns a compact representation with an uncertainty-aware recurrent variational autoencoder (RNN-VAE), and (iii) scores anomalies using complementary detectors on both reconstruction space and latent space<sup>6,7</sup>. Specifically, we compute three anomaly signals: (1) VAE reconstruction-based scores that highlight light curves the model cannot accurately reproduce (a reconstruction-likelihood proxy for atypical structure), (2) density-based outlier structure in standardized latent space using HDBSCAN, and (3) tree-ensemble isolation scores in the same standardized latent space using Isolation Forest<sup>8–10</sup>. Because each detector can fail in different ways—reconstruction metrics can be sensitive to noise and calibration, while latent-space detectors depend on the learned embedding geometry—we ensemble their outputs and prioritize consensus candidates for follow-up. Our results show partial overlap between detectors, consistent with their complementary inductive biases. The consensus ensemble yields 226 candidates (6.29% of the dataset) flagged by at least two meth-

<sup>1</sup> (MIT) International Academy High School of Michigan

<sup>2</sup> Massachusetts Institute of Technology (MIT) Kavli Institute for Astrophysics and Space Research.

---

ods, and a strict shortlist of 116 candidates (3.23%) flagged by all three. Beyond producing these shortlists, we contribute a sensitivity-driven selection procedure based on threshold stability and hyperparameter robustness, and we provide a manual vetting checklist intended to support downstream inspection and scientific follow-up of ensemble-selected anomalies<sup>11,12</sup>.

## Data

We analyze a curated dataset of processed TESS light curves provided by an MIT laboratory preprocessing pipeline. Because the files were delivered as preprocessed time-series products, mission acquisition metadata (e.g., sector identifiers and cadence) were not consistently available and are therefore not used as selection variables in this study. We treat the provided columns (*relative\_time*, *tess\_flux*, *tess\_uncert*) as the primary observables and describe all additional transformations that we apply in Section Preprocessing: Per-object Robust Clipping, Normalization, and Uncertainty Weights. The dataset consists of 3,659 unlabeled CSV light-curve files, each corresponding to a single object. Each file contains time-series data with three columns: *relative\_time*, *tess\_flux*, and *tess\_uncert*. Figure 1 is a randomly selected light curve file, representing the original light curve of object 2019soh, including the relative time, flux, and flux uncertainty values. To construct a unified dataset for this analysis, we combined all individual light curve files into a single dataset. To associate each data point with its corresponding object, we extracted an identifier from the file name and added it as a new column, *object\_name*. For example, for a file named *lc\_2019soh\_processed*, the *object\_name* was derived as 2019soh, where "lc" and "processed" are common prefixes and suffixes shared across all file names. The final dataset comprises four columns: *relative\_time*, *tess\_flux*, *tess\_uncert*, and *object\_name*.

To ensure that anomaly signals are not driven by imputation artifacts, we removed missing measurements without filling (i.e., no interpolation or forward/backward filling). After removing missing values, we excluded any light curve with fewer than 20 valid time–flux pairs since sequences below this length provide insufficient information for stable representation learning and reconstruction-error estimation. The final cleaned dataset contains 3,593 light curves (98.2% of the initial population), which we randomly partitioned at the object level into a training set and validation set using a 0.9/0.1 split with a fixed random seed (*randomstate* = 42). This object-level split prevents leakage of information between training and validation and provides a consistent basis for all sensitivity studies and downstream anomaly scoring.

Since light curves represent time-series data, we organized the dataset as a nested 2D array. To standardize the time values

across all objects, we shifted the start time of each light curve to zero using (1):

$$t' = t - \min(t) \quad (1)$$

## Methods

This section describes the data preprocessing, representation learning model, anomaly scoring methods, and ensemble decision rule used to construct a shortlist of unusual light curves<sup>4,5</sup>.

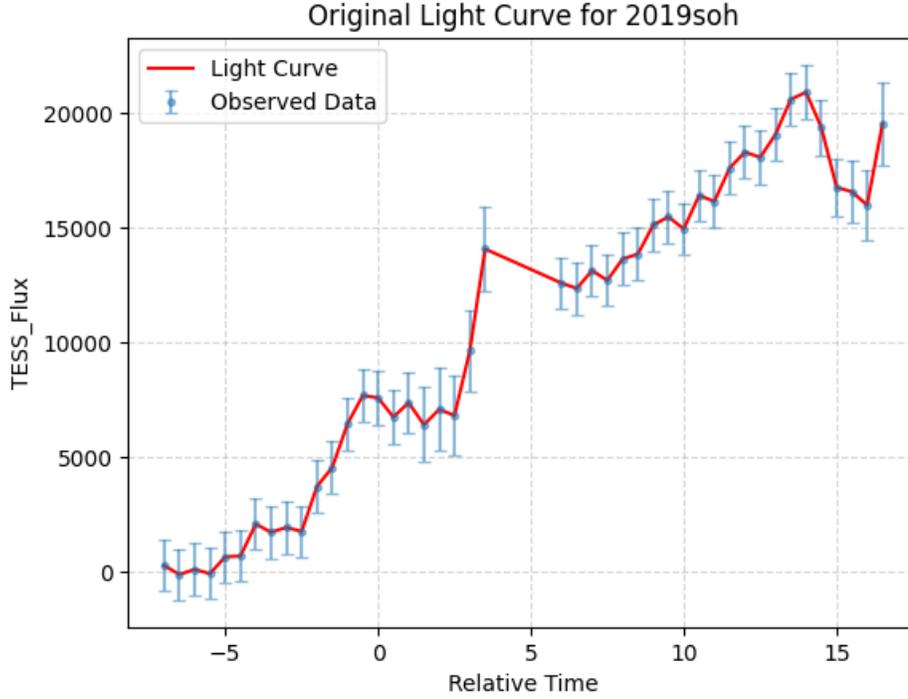
### Pipeline Overview and Design Goals

Our goal is to detect rare or unusual light curve morphologies with minimal assumptions and without a labeled training set. Figure 2 summarizes the end-to-end workflow. The pipeline consists of<sup>11,13</sup>:

- I. **Preprocessing** of each light curve with robust per-object clipping and normalization, incorporating uncertainty weights when available (Section Preprocessing: Per-object Robust Clipping, Normalization, and Uncertainty Weights).
- II. **Representation learning** with an uncertainty-aware RNN-VAE that generates both a reconstruction  $\hat{x}$  and latent statistics  $\mu, \log \sigma^2$  (Section Representation Learning with an Uncertainty-aware RNN-VAE).
- III. **Three anomaly signals**:
  - (a) Reconstruction error  $e(x)$  (Section VAE Reconstruction-error Anomaly Detection),
  - (b) HDBSCAN outlier score in latent space (Section HDBSCAN Anomaly Detection),
  - (c) Isolation Forest outliers in latent space (Section Isolation Forest latent-space anomalies).
- IV. **Ensemble consensus** to produce robust shortlists (Section Ensemble Intersection and Manual Vetting).

### Preprocessing: Per-object Robust Clipping, Normalization, and Uncertainty Weights

TESS light curves can contain outliers, discontinuities, and sector-dependent systematics. To reduce sensitivity to rare, extreme artifacts while preserving object-specific morphology, we apply per-object robust clipping using the median and a scaled median absolute deviation (MAD). We then normalize each object to reduce scale differences across stars



**Fig. 1** Original light curve for object 2019soh.

and to allow the model to focus on shape rather than amplitude. When uncertainty estimates are available, we incorporate them as weights so that high-uncertainty points contribute less to the training objective<sup>14</sup>. When performing the raw data distribution for all data points, the minimum  $tess\_flux$  is  $-3.1e8$ , the maximum  $tess\_flux$  is  $1.2e7$ , approximately 0.1% of the  $tess\_flux$  exceeded  $-1.1e6$  or fell below  $-1.1e6$ , and approximately 7.3% of the  $tess\_flux$  exceeded  $1e4$  or fell below  $-1e4$ . The distribution of  $tess\_flux$  is illustrated in Figure 3, which shows that most of the data lies within the range of  $-5,000$  to  $5,000$ . The extreme data is from peak of transient light curves typically, a transient light curve example shown in Figure 5 after normalization. Per-object processing is central because TESS transients exhibit large between-object differences in baseline and scale. If a global normalization is applied across all measurements, embeddings may reflect brightness differences more than temporal shape. We therefore normalize each object independently using robust statistics.

**Robust clipping.** For each object, we compute the median flux  $m$  and the median absolute deviation (MAD). We then cap only extreme flux values outside  $m \pm k \cdot MAD$  at the boundary. In this study, we use  $k = 25$ , chosen via a sensitivity sweep to balance artifact suppression against preservation of astrophysical morphology.

$$f_{clipped}(t) = \text{clip}(f(t), m - k \cdot MAD, m + k \cdot MAD) \quad (2)$$

**Per-object normalization.** After clipping, we compute  $robust\_median$  and  $robust\_mad$  per object and normalize:

$$f_{norm}(t) = \frac{f_{clipped}(t) - robust\_median}{robust\_mad} \quad (3)$$

**Uncertainty scaling and weights.** Uncertainties are scaled by the same  $robust\_mad$ . We compute inverse-variance weights  $\omega(t) = \frac{1}{(\sigma_{norm}(t)^2 + \epsilon)}$  and (optionally) normalize weights within each object so  $\max(w) = 1$  to stabilize training.

$$uncert_{scaled}(t) = \frac{tess\_uncert(t)}{robust\_mad} \quad (4)$$

$$w_{raw}(t) = \frac{1}{uncert_{scaled}(t)^2 + \epsilon}, \quad \epsilon = 10^{-10} \quad (5)$$

$$weights_{normalized}(t) = \frac{w_{raw}(t)}{\max(w_{raw}(t))} \quad (6)$$

### Sensitivity Studies for Preprocessing Choices

We justify preprocessing parameters using two complementary sensitivity analyses<sup>14</sup>:

(A) Robust clipping factor  $k$ : evaluate  $k \in \{3, 5, 10, 15, 25\}$  and track

- i) fraction of points clipped,

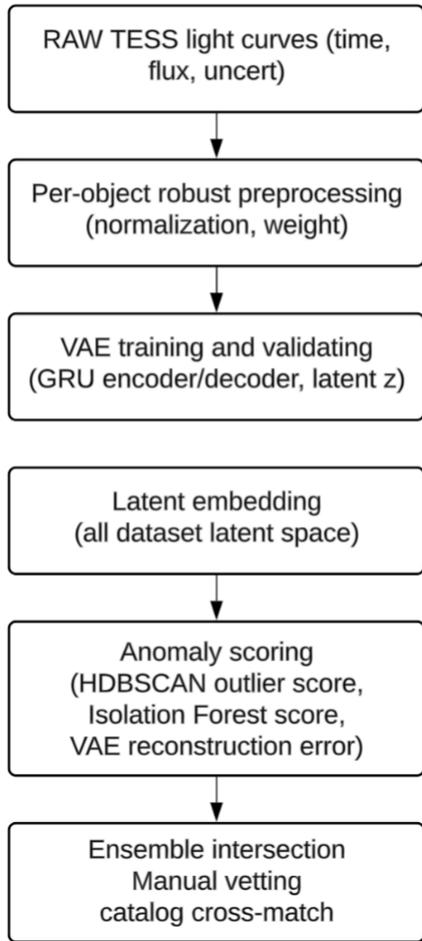


Fig. 2 End-to-end workflow.

- ii) morphology preservation (peak amplitude and turning points)

To determine the optimal  $k$  value for the robust normalization procedure, we conducted a sensitivity analysis by evaluating multiple  $k$  values, and the clipped point percentage result is shown in Figure 4. The choice of  $k$  represents a critical balance between removing spurious outliers and preserving genuine astrophysical signals in transient light curves. Figure 5 shows the normalized flux profiles with unclipped and clipped by various  $k$ , whose object name is 2019kli with the maximum  $tes\text{.flux}$  of  $1.2e7$ . It can be observed that the maximum  $tes\text{.flux}$  is over 40 times as MAD for this object. Figure 6 shows the percentage of objects clipped with  $k$ . When  $k = 25$ , only 2% of the total 3,593 objects were impacted by the capping process. Based on this analysis, we selected  $k = 25$  for all subsequent normalization procedures. This represents a con-

servative compromise that:

1. **Preserves astrophysical signals:** Only 0.17% of data points are affected, minimizing the risk of clipping genuine transient events for the majority of light curves.
2. **Mitigates extreme artifacts:** The most extreme outliers (beyond  $m \pm 16.9\sigma$  equivalents) are capped, reducing their influence on subsequent analysis.
3. **Accommodates distribution characteristics:** The chosen  $k$  respects the heavy-tailed nature of the data while providing consistent normalization across all objects

(B) Per-object vs global normalization: we did a comparison study by normalizing the flux globally, which means all flux is divided by a common scale factor, which is the minimum of absolute 5 or 95 percentiles. The normalized result for the same object 2019kli was shown in Figure 7. It can be observed that the light curve pattern was completely altered, as compared to the same object shown in Figure 5 using robust MAD normalization and clipping.

### Evaluation Metrics and Robustness Criteria

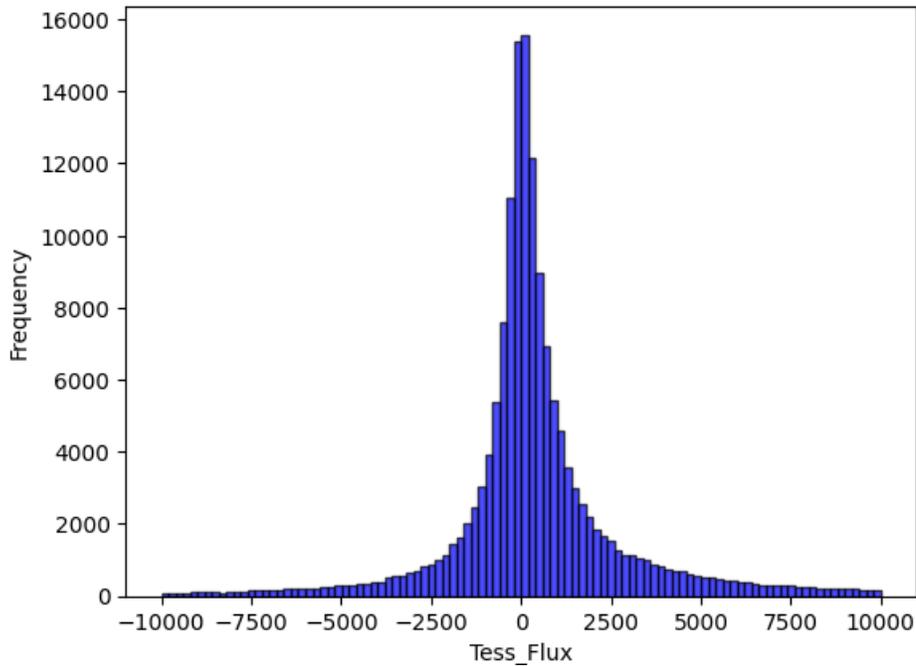
Because ground-truth anomaly labels are not available, we evaluate robustness using internal consistency metrics and stability under parameter perturbation<sup>4</sup>:

- Anomaly fraction: proportion of objects flagged as anomalous.
- Shortlist stability: Jaccard similarity between anomaly sets produced under adjacent parameter choices (e.g., neighboring hyperparameter values).
- Noise fraction (HDBSCAN): fraction of points labeled as noise.
- Cluster persistence and membership probability (HDBSCAN): diagnostic measures of cluster stability and assignment confidence.
- Sensitivity regimes (Isolation Forest): analysis of how the anomaly signal changes as contamination varies, separating stable from overly sensitive settings.

These metrics guide parameter selection toward stable, reproducible behavior rather than arbitrary anomaly rates.

### Representation Learning with an Uncertainty-aware RNN-VAE

Variational Autoencoders (VAEs) are generative models that encode complex data into a lower-dimensional latent space while preserving a probabilistic interpretation. Introduced



**Fig. 3** Distribution of raw *tess\_flux* values across all cadences before robust clipping.

by Kingma and Welling in 2013<sup>11</sup>, VAEs combine neural networks with variational inference to learn both a reconstruction model and a structured latent distribution, making them well-suited for representation learning in unlabeled time-series data<sup>7,15</sup>. VAEs are built on the concept of an autoencoder, which consists of two neural networks: an encoder and a decoder. The encoder compresses input data into a latent representation, and the decoder reconstructs the original data from this compressed latent space. Unlike traditional autoencoders, VAEs incorporate probabilistic modeling, which provides two main advantages: they can generate new samples from the learned distribution, and they provide a meaningful latent space that captures the underlying structure of the data. The VAE in this study uses a Gated Recurrent Unit (GRU) encoder–decoder. GRUs are recurrent neural networks designed to mitigate vanishing-gradient issues and model temporal dependencies efficiently, which is appropriate for variable-length TESS light-curve sequences<sup>16</sup>. Related recurrent architectures such as LSTMs have also been applied to time-series anomaly detection<sup>17</sup>. Each light curve is a variable-length sequence  $x(t) = [t, f_{norm}(t)]$ . The encoder outputs  $\mu(x)$  and  $\log \sigma^2(x)$ ; a latent sample  $z$  is obtained via the reparameterization trick and decoded to reconstruct  $\hat{x}(t)$ . The loss function is a weighted Evidence Lower Bound (ELBO), which consists of two components: -Reconstruction Loss: Computed as the mean squared error (MSE) between the original input and the reconstructed output, this ensures that the reconstruction of the

data is accurate enough for analysis. Here, we use weighted MSE to incorporate flux uncertainty. -Kullback-Leibler Divergence (KLD): This term regularizes the model by approximating a Gaussian distribution in the latent space. It provides a continuous latent representation and prevents overfitting. The total ELBO loss is represented as:

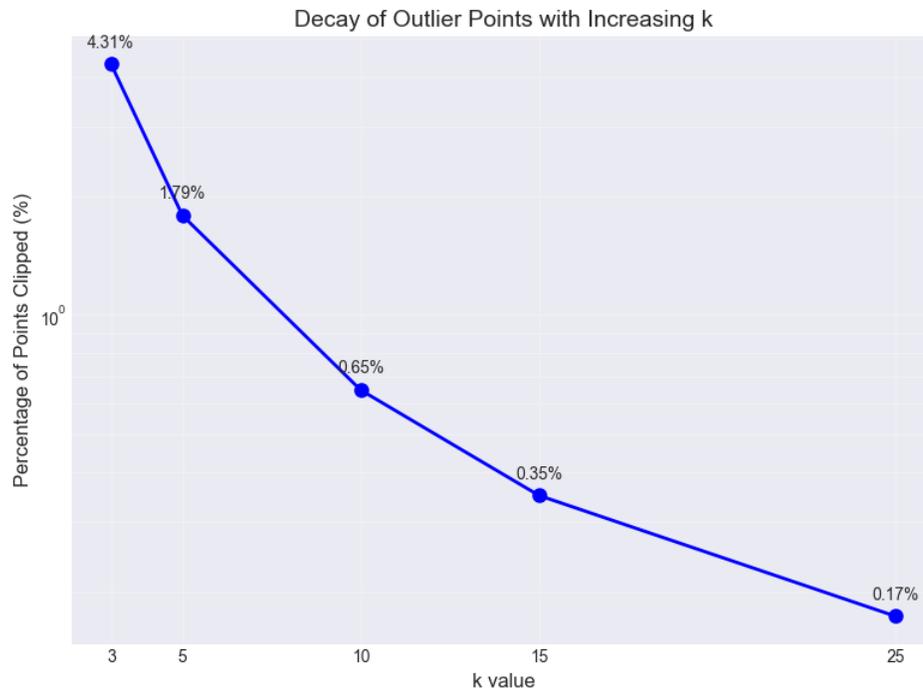
$$ELBO = \omega(t) \cdot MSE + KLD \quad (7)$$

The architecture of VAE and subsequent anomaly detection is shown in Figure 8.

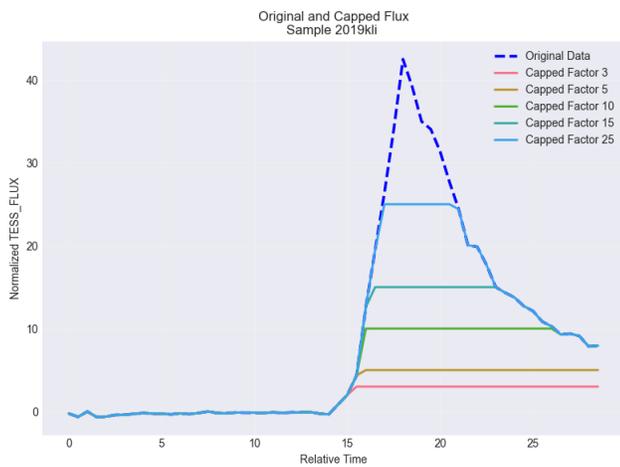
### VAE Setup and Convergence Monitoring

We use a recurrent encoder–decoder architecture with hidden size of 128, a latent size of 16, and 4 recurrent layers. We train with Adam (learning rate 10<sup>-4</sup>) for up to 1400 epochs, saving checkpoints periodically. Convergence is monitored using validation reconstruction error and ELBO components (reconstruction and KLD)<sup>18</sup>. The basic parameters used in the VAE are listed as follows:

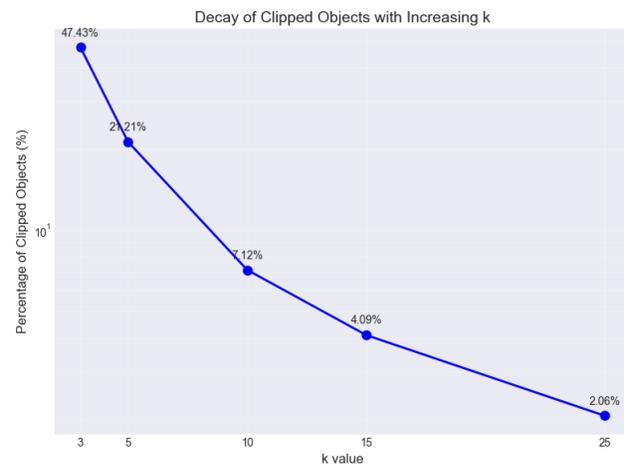
- Input size = 2
- Hidden size = 128
- Latent size = 16
- Output size = 1
- Number of GRU layers (encoder and decoder) = 4
- Optimizer = *Adam*
- Learning rate = 1.e - 4



**Fig. 4** Fraction of flux points clipped as a function of MAD factor  $k$  in the per-object robust clipping step.



**Fig. 5** Example normalized light curve (object 2019kli) under different MAD factors  $k$ .



**Fig. 6** Fraction of objects affected by clipping as a function of MAD factor  $k$ .

Drop out = 0.2  
Epochs = 1400

As stated in the previous section, we use a 90/10 train-validation split (*random\_state* = 42) and monitor ELBO components (reconstruction loss and KLD) to verify stable convergence before downstream scoring. Early stopping was used for a case study, with patient parameter as of 2,

which means checking validation loss increasing for the last 2 epochs, and the training process was not triggered by early stopping until it is completed at 800 epochs. The training and validation loss behavior had the same pattern as in Figure 11, which did not use early stopping.

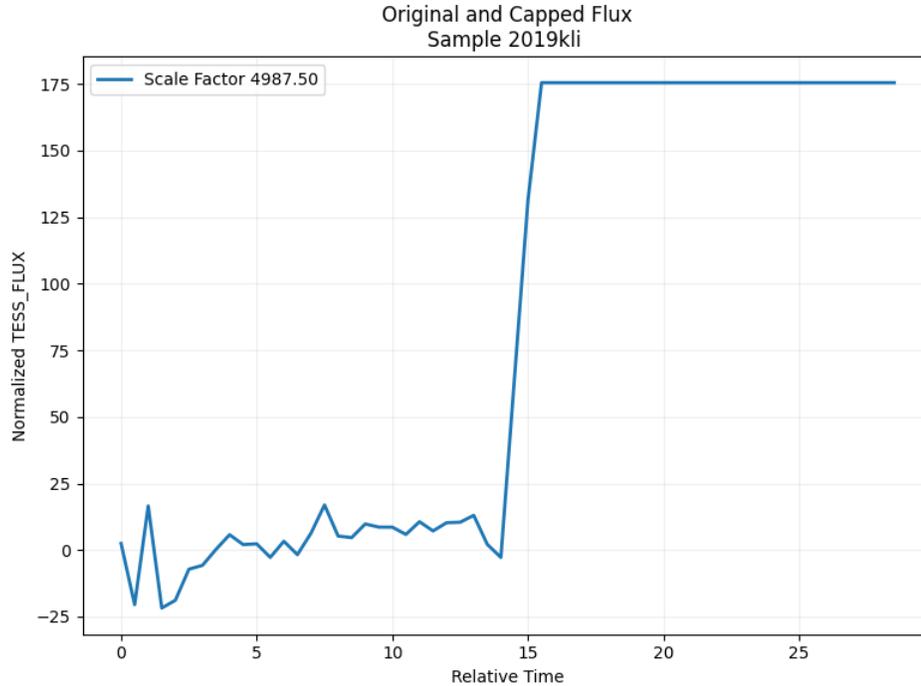


Fig. 7 Example normalization using a global scale factor (shown for comparison).

### VAE Reconstruction-error Anomaly Detection

Reconstruction error is the (weighted) mean squared error over normalized flux values across all timesteps in an object. Reconstruction error is used as an anomaly score: objects in the high-error tail are candidates<sup>6</sup>. **Sensitivity study (epoch × threshold)** To select a defensible checkpoint and threshold, we perform a two-dimensional sweep: epochs from 200 to 1400 (step 100) and reconstruction-error thresholds from 0.2 to 0.9 (step 0.1). For each epoch,  $t$ , we compute the anomaly fraction:

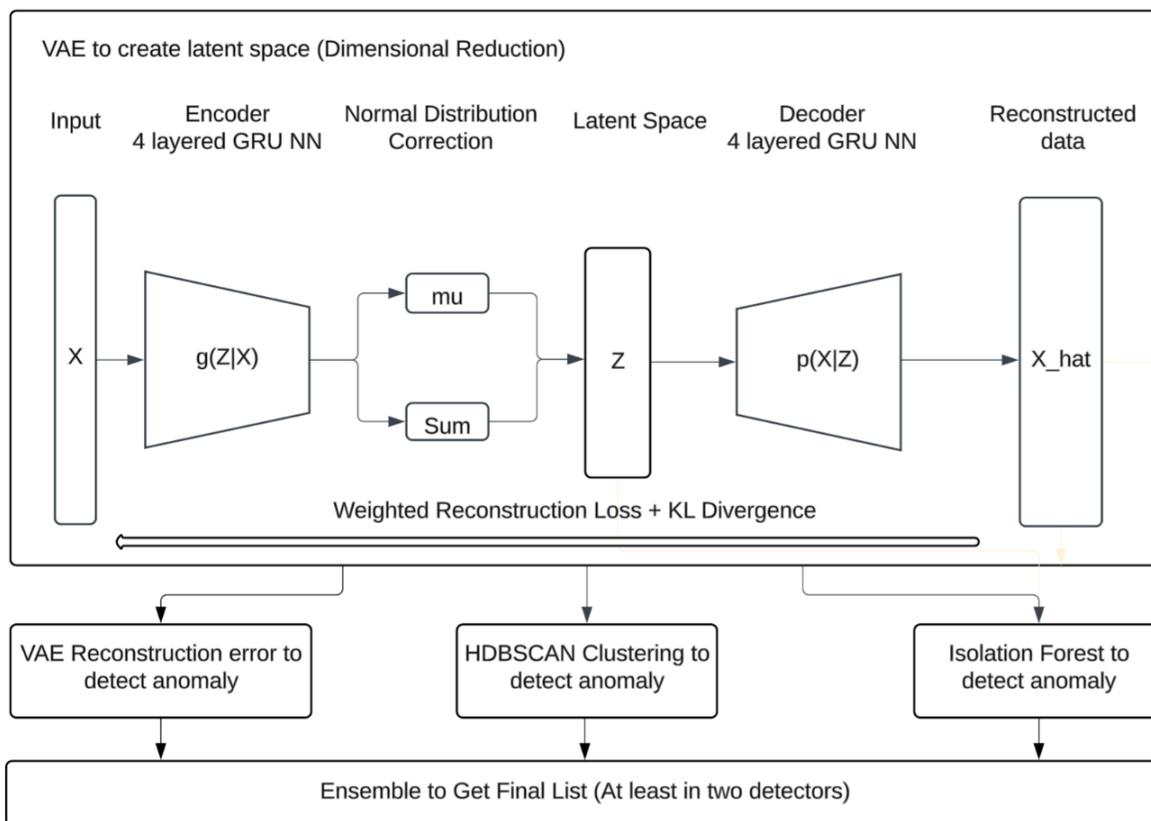
$$p(\tau; epoch) = \frac{1}{N} \sum_{i=1}^N I[e_i(epoch) \geq \tau] \quad (8)$$

where  $I[\cdot]$  is the indicator function, when the condition is met, then 1, else 0. Results are summarized in Figure 9 and Table 1. The validation reconstruction curve flattens around 700 epochs, indicating convergence, and threshold–anomaly curves stabilize thereafter. We therefore select the converged checkpoint (epoch  $\approx$  700) and choose  $\tau = 0.5$ , which corresponds to 398 anomalies (11.0%).

### HDBSCAN Anomaly Detection

Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) is an advanced clustering algorithm

that identifies clusters based on density variations in the data. Unlike traditional density-based clustering methods such as DBSCAN, HDBSCAN builds a hierarchical tree of clusters and condenses it to select the most stable structures. This capability makes HDBSCAN particularly effective for datasets with variable densities, such as light curves derived from astronomical observations<sup>8,19</sup>. The latent space features generated by the VAE were input into the HDBSCAN algorithm for clustering analysis. HDBSCAN identifies dense clusters representing typical behaviors in the dataset and labels sparse regions as noise or anomalies. Points classified as noise or outliers by HDBSCAN are flagged as potential anomalies, which may correspond to astrophysical phenomena or instrumental irregularities. HDBSCAN and related density-based methods have seen growing use in astronomy for exploratory clustering and anomaly discovery in high-dimensional representations (e.g., within the Astronomy framework and related work combining clustering with anomaly scoring). Related unsupervised approaches have been applied to transient discovery in survey light curves<sup>20</sup>. We apply HDBSCAN to the latent mean vectors  $\mu_i$  (standardized per dimension) to identify low-density outliers. HDBSCAN yields (i) cluster labels, (ii) membership probabilities, and (iii) outlier scores (higher indicates more outlier-like). We define anomalies by thresholding the outlier score. For qualitative visualization only, we include a 2-D t-SNE embedding of the standardized latent vec-



**Fig. 8** Pipeline overview: preprocessing → uncertainty-aware RNN-VAE → anomaly signals (reconstruction error, HDBSCAN, Isolation Forest) → consensus ensemble.

tors in the Supplement; all clustering and anomaly scoring are performed in the full latent space.

**Sensitivity study** ( $mcs \times ms \times threshold$ ) We sweep  $min\_cluster\_size$  and  $min\_samples$  over a grid and evaluate multiple outlier-score thresholds, prioritizing shortlist stability and diagnostic quality. We compute anomaly shortlist stability using Jaccard similarity of anomaly sets across adjacent settings, and we track noise fraction, mean membership probability, and cluster persistence. Based on stability and diagnostic criteria, we select  $min\_cluster\_size = 15$  and  $min\_samples = 8$  and outlier-score threshold 0.6. This yields 256 anomalies (7.125%) and strong shortlist stability as shown in Table 2. ( $mcs$ :  $min\_cluster\_size$ ,  $ms$ :  $min\_samples$ ,  $jac$ : Jaccard similarity score) Additional HDBSCAN parameters (defaults unless stated):  $metric = euclidean$ ,  $cluster\_selection\_method = eom$ .

### Isolation Forest Anomaly Detection

Isolation Forest (IF) is an unsupervised algorithm for anomaly detection that isolates rare points using random partitioning:

anomalies tend to be separated with fewer splits than typical points<sup>9</sup>. We apply Isolation Forest to the standardized latent vectors  $\mu$ . IF isolates anomalies via random partitions, assigning lower decision-function scores to more anomalous points. IF requires specifying contamination  $c$ , the expected outlier fraction.

### Score distribution and contamination sensitivity

We first inspect the distribution of IF decision scores (Figure 10) to confirm a concentrated bulk with a tail of low scores. We then perform a sensitivity sweep over contamination  $c$ , tracking (i) a signal-stability metric based on the mean score of a fixed extreme subset (top K%) and (ii) the expected linear increase in predicted outlier count with  $c$ . The sensitivity curve exhibits a transition from a highly sensitive regime at small  $c$  to a flatter regime where increases in  $c$  primarily add borderline points. We select  $c = 0.064$  near the onset of the stable regime, yielding approximately 230 outliers (6.4%). Isolation Forest parameters (defaults unless stated):  $n\_estimators = 1000$ ,  $random\_state = 42$ .

## Anomaly Detection Analysis Across Training Epochs

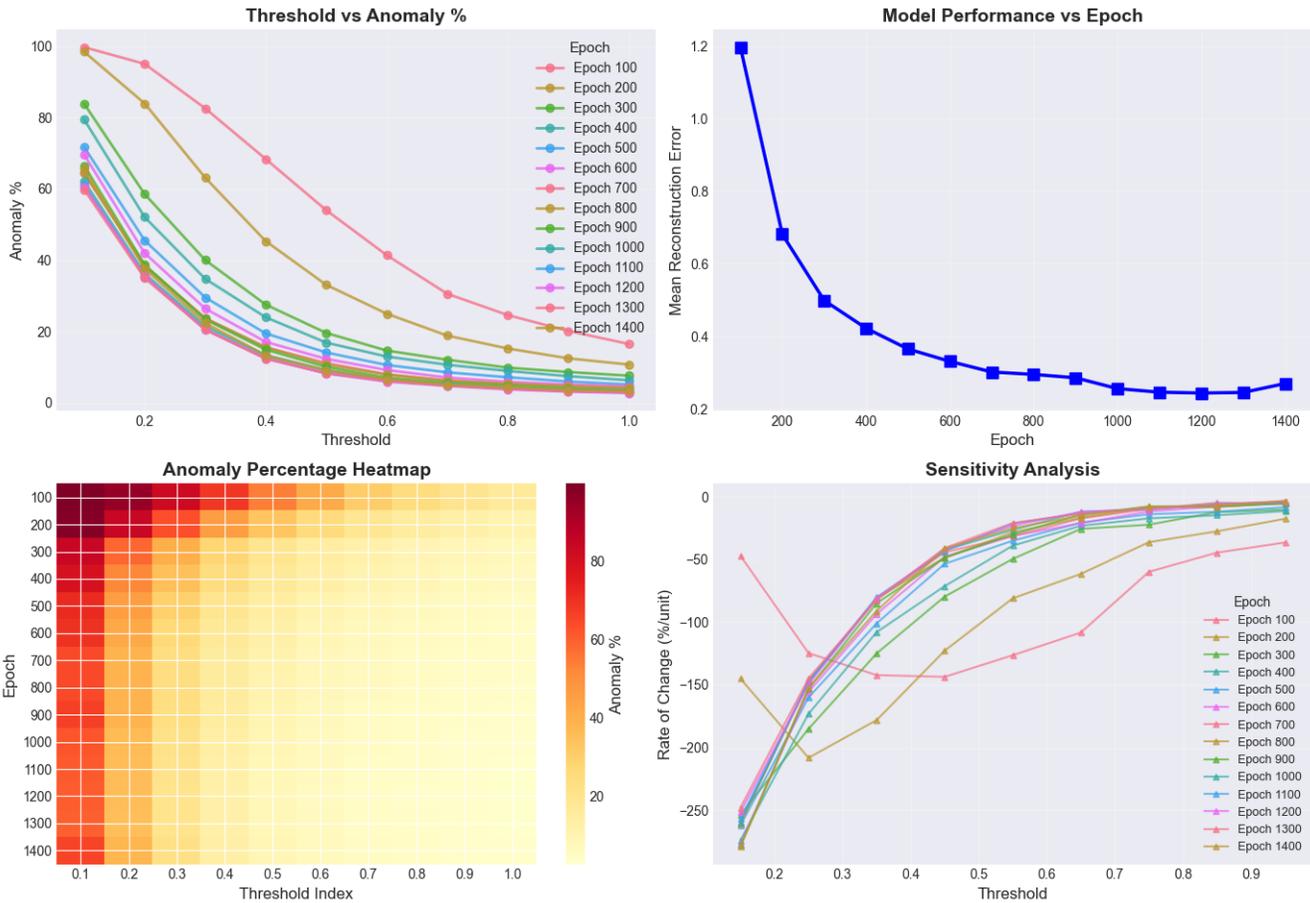


Fig. 9 VAE Reconstruction Error Sensitivity Study.

### Ensemble Intersection and Manual Vetting

Each detector flags anomalies under a different definition. To improve robustness and reduce detector-specific false positives, we combine detector outputs using an agreement-based ensemble. We assign each object a consensus score equal to the number of detectors that flag it (0–3). We report two shortlists<sup>11,13,17</sup>:

- Moderate consensus ( $\geq 2$ ): increases recall while suppressing single-detector artifacts.
- Strict consensus (3-of-3): a high-confidence shortlist robust across all detectors.

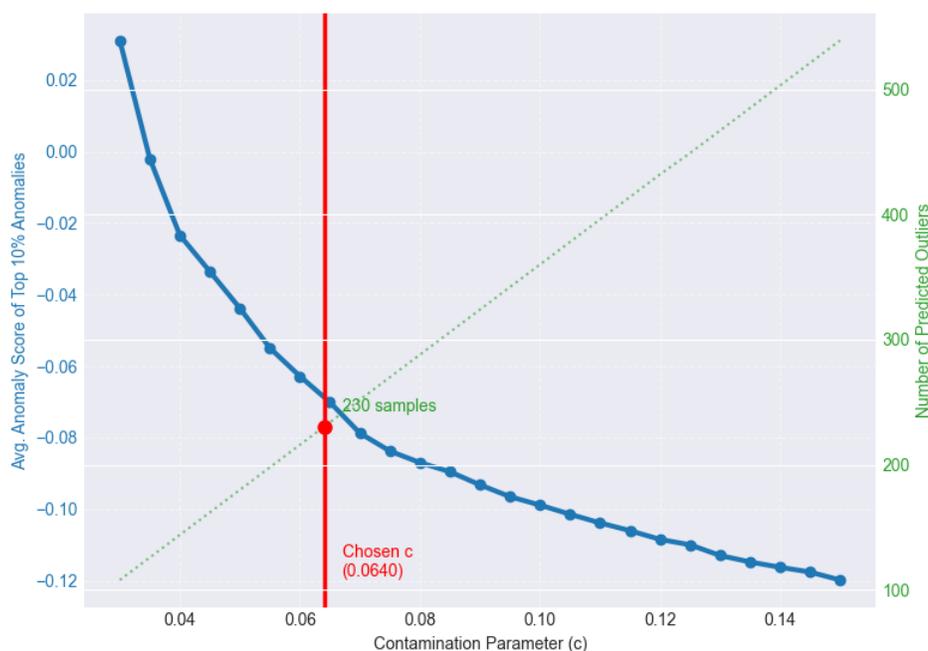
Manual vetting is applied as a final qualitative check to confirm that shortlisted objects exhibit non-trivial astrophysical structure rather than trivial preprocessing failures. For each candidate we (i) verify that the time–flux series contains  $\geq 20$

valid cadences after cleaning, (ii) check for step-like discontinuities consistent with scattered-light or stitching artifacts, (iii) confirm that anomalies persist under small changes to preprocessing hyperparameters within the stable regime, and (iv) visually inspect the normalized light curve for coherent variability (e.g., transit-like dips, flare-like spikes, or structured trends) before prioritizing follow-up.

### Implementation Details and Reproducibility

All experiments were implemented in Python using PyTorch for the RNN-VAE<sup>8,21,22</sup>. We fix random seeds where applicable (including `random_state = 42` for the object-level 0.9/0.1 train–validation split in Section Data to ensure repeatability of sensitivity sweeps and shortlist comparisons<sup>21–23</sup>). Key model hyperparameters are listed in Section VAE Setup and Convergence Monitoring (latent dimension 16; GRU hidden size

### Isolation Forest Sensitivity Analysis in VAE Latent Space



**Fig. 10** Isolation Forest sensitivity analysis in VAE latent space; selected contamination  $c = 0.064$  (red).

128; 4-layer encoder/decoder; dropout 0.2; Adam optimizer with learning rate  $1 \times 10^{-4}$ ; 1400 epochs with checkpointing). All latent-space detectors operate on standardized latent mean vectors (zero mean, unit variance per dimension) to place distances and axis-wise densities on a comparable scale before clustering or isolation-based scoring.

## Results

We report convergence diagnostics for the representation learner and summarize the anomaly candidates identified by each detector, followed by the intersection-based ensemble shortlist<sup>11,13</sup>.

### Model convergence and reconstruction behavior

Training of the RNN-VAE showed rapid improvement in reconstruction quality during early epochs, followed by a clear plateau, indicating convergence of the learned representation. The average validation loss (weighted  $MSE + KLD$ ) decreased steeply up to mid-training and then flattened, with diminishing improvements beyond approximately 700 epochs, as shown in Figure 11. This convergence pattern is consistent with the threshold–anomaly curves across epochs (Figure 9): at early epochs, small changes in the reconstruction-error threshold produce large swings in anomaly percentage, whereas

after convergence, the anomaly-rate curves become substantially more stable. Based on this stabilization, we selected the checkpoint at epoch 700 for downstream anomaly scoring and detector ensemble<sup>6,18</sup>. Figure 12 shows the normalized light curve for object 2019soh mentioned earlier, before and after training. The reconstruction curves caught the pattern of the original curve, and the model performed well for the studied epochs.

### Reconstruction-error anomalies

At epoch  $\sim 700$ , the reconstruction-error threshold selects 398/3593 objects ( $\sim 11.1\%$ ) as anomalous with a threshold of 0.5. The threshold–anomaly curve exhibits a knee near this operating point, whereas lower thresholds increasingly include borderline cases<sup>6</sup>.

### HDBSCAN latent-space anomalies

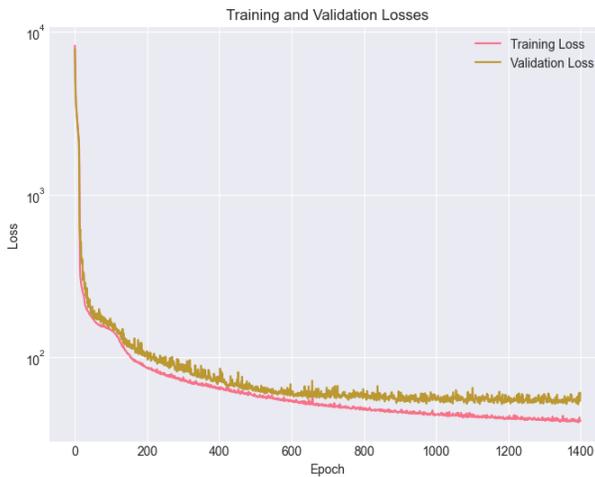
Using the selected HDBSCAN hyperparameters and outlier-score threshold 0.6, HDBSCAN identifies 256/3593 ( $\sim 7.1\%$ ) anomalies. The chosen setting lies in a stable regime of the sweep with high anomaly-set stability and low noise fraction<sup>8</sup>.

**Table 1** Summary: Anomaly Percentages by Epoch and Threshold

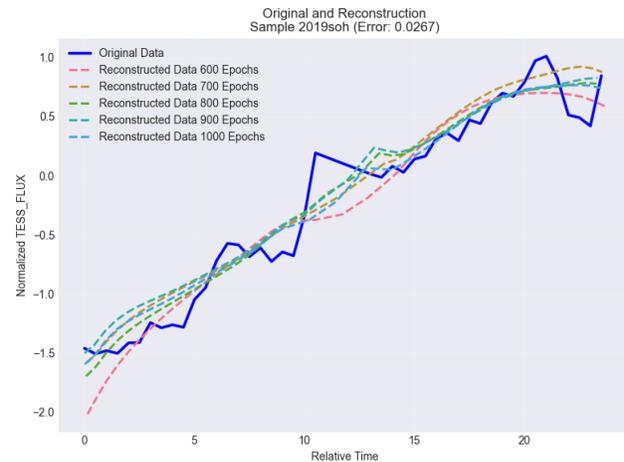
Epoch	0.2	0.3	0.4	0.5	0.6	0.7	0.9	Mean Error
200	84.00%	63.00%	45.40%	33.20%	25.00%	18.80%	12.60%	0.6815
400	51.90%	35.00%	23.90%	16.80%	12.90%	10.50%	7.50%	0.4216
500	45.30%	29.60%	19.50%	14.10%	10.80%	8.50%	5.90%	0.3643
600	41.90%	26.50%	17.20%	12.50%	9.20%	7.10%	5.00%	0.3309
700	38.40%	23.60%	15.60%	11.00%	8.10%	6.20%	4.80%	0.3010
800	38.30%	23.50%	15.40%	10.70%	7.80%	6.30%	4.60%	0.2948
900	38.70%	23.60%	14.90%	10.10%	7.00%	5.70%	4.00%	0.2856
1000	36.30%	21.50%	13.40%	9.20%	6.50%	5.20%	3.70%	0.2562
1100	35.40%	20.60%	12.70%	8.30%	6.20%	4.90%	3.30%	0.2463

**Table 2** Sensitivity study of HDBSCAN

mcs	ms	rank_score	anomaly_frac	persist_mean	Prob_clustered	jac_ms	jac_mcs
15	8	0.72	7.13	0.19	0.73	0.88	0.78
15	10	0.69	7.85	0.19	0.70	0.88	0.57
10	8	0.69	3.90	0.17	0.80	0.84	0.78
5	8	0.65	3.90	0.17	0.80	0.68	0.78
10	12	0.65	3.48	0.17	0.82	0.84	0.59
10	10	0.64	3.67	0.17	0.81	0.84	0.57
5	12	0.60	1.53	0.15	0.89	0.68	0.59
5	10	0.59	1.56	0.14	0.89	0.68	0.57



**Fig. 11** Training loss vs. epochs for the VAE model.



**Fig. 12** Normalized light curves before and after training for object 2019soh.

### Isolation Forest latent-space anomalies

Isolation Forest with the selected contamination identifies approximately 230/3593 (~ 6.4%) outliers. Score-distribution inspection and contamination sensitivity indicate a stable region before overly permissive settings begin to add borderline points<sup>9</sup>.

### Ensemble shortlist results

The three detectors partially overlap, reflecting complementary definitions of “unusual.” Pairwise overlaps are 38 ( $HDBSCAN \cap IF_{only}$ ), 63 ( $HDBSCAN \cap Reconstruction_{only}$ ), and 9 ( $IF \cap Reconstruction_{only}$ ). In total, 226 objects are flagged by at least 2-of-3 detectors,

---

and 116 objects are flagged by all 3 detectors, forming a high-confidence shortlist for prioritized follow-up<sup>11</sup>. Counts and fractions of light-curve objects flagged as anomalous under a three-detector ensemble consisting of VAE reconstruction-error thresholding, HDBSCAN latent-space outlier scoring, and Isolation Forest latent-space isolation. Pairwise overlaps are decomposed into “exactly two” agreements by subtracting the three-way intersection. We report both a moderate-consensus shortlist ( $\geq 2$  agreement) and a strict shortlist (3-of-3 agreement), which prioritize robustness to method-specific false positives.

## Discussion

We interpret the detector behaviors, discuss why consensus improves reliability in unlabeled settings, and outline practical validation steps and limitations for TESS survey light curves<sup>4,13</sup>.

### Why ensembling improves reliability in unsupervised settings

Unsupervised anomaly detection in survey data is vulnerable to detector-specific false positives. Reconstruction error may be inflated by noise bursts or detrending artifacts; density-based methods can be affected by local neighborhood structure; and Isolation Forest can isolate sparse points even when they are not scientifically interesting. An agreement-based ensemble reduces these risks by prioritizing candidates that remain anomalous across independent criteria. In this pipeline, consensus anomalies represent objects that are simultaneously difficult to reconstruct and isolated in latent structure, improving robustness for downstream vetting<sup>4</sup>.

### Interpreting the anomaly fractions

The anomaly fractions differ across detectors because they measure different properties and are calibrated differently. The reconstruction-error detector is intentionally broader (11.0%) to capture time-domain deviations, while latent-space detectors are slightly more conservative ( $\sim 6\text{--}7\%$ ) to focus on structural outliers in representation space. These rates are best interpreted as screening yields rather than estimates of true astrophysical anomaly prevalence. The ensemble step provides a principled way to reconcile differing detector yields into a stable shortlist<sup>13</sup>.

### Convergence and threshold stability as justification

The reconstruction-error sensitivity analysis provides two layers of justification: (i) the model converges around 700

epochs, and (ii) post-convergence anomaly fractions are stable under epoch changes for mid-range thresholds. Similarly, HDBSCAN and IF parameter selection is anchored to stability criteria rather than matching a desired anomaly fraction. This emphasis on stability reduces the risk that conclusions depend on fragile hyperparameter choices<sup>6</sup>.

### Latent-space quality and collapse considerations

A common VAE risk is posterior collapse, where latent variables carry little information, and anomalies become harder to separate in representation space. Mixed per-dimension variability (some latent dimensions near unit-scale and others smaller) suggests partial under-utilization but not full collapse. Importantly, the observed stability of HDBSCAN anomaly sets across sweeps indicates that latent structure is sufficiently informative for outlier detection under the chosen settings. Future work can quantify latent utilization more directly using KL-per-dimension diagnostics or mutual information estimates<sup>6</sup>.

### TESS catalogue cross-validation

#### TESS catalogue cross-validation

To contextualize the ensemble candidates, we cross-matched all 3,593 objects and each shortlist against public TESS catalog classification fields (public catalogs link: [https://tess.mit.edu/public/tesstransients/lc\\_bulk/count\\_transients.txt](https://tess.mit.edu/public/tesstransients/lc_bulk/count_transients.txt)). Across the full dataset, 479/3,593 objects (13.3%) are catalog-classified (i.e., not labeled “Unclassified”). Within the  $\geq 2$ -of-3 consensus shortlist ( $N = 226$ ), 28 objects (12.4%) are catalog-classified, and within the strict 3-of-3 shortlist ( $N = 116$ ), 12 objects (10.3%) are catalog-classified. The majority of ensemble-selected candidates therefore lack an existing catalog classification (198/226 and 104/116, respectively), suggesting that the consensus rule preferentially surfaces under-characterized light-curve morphologies; however, uncataloged objects can include both genuinely rare astrophysical variability and residual instrumental/systematic effects. For follow-up, we recommend screening candidates using pipeline data-quality and contamination indicators<sup>3,24</sup>. It is important to note that the training population used in this study is a subset of objects with available processed light curves and, therefore, does not represent the full population of sources present in the public TESS catalogues. Consequently, “anomalous” scores are defined relative to this limited reference set: objects that are well characterized (i.e., catalog-classified) may still appear unusual with respect to the subset’s dominant morphology distribution and instrument/noise properties. In that sense, catalog-classified objects flagged as anomalies may reflect genuine rarity within the

**Table 3** Detector agreement structure and ensemble shortlist sizes

Combination (flagged by...)	Count	% of dataset
HDBSCAN $\cap$ Isolation Forest $\cap$ Reconstruction (3-of-3)	116	3.23%
HDBSCAN $\cap$ Isolation Forest only (exactly 2-of-3)	38	1.06%
HDBSCAN $\cap$ Reconstruction only (exactly 2-of-3)	63	1.75%
Isolation Forest $\cap$ Reconstruction only (exactly 2-of-3)	9	0.25%
At least 2-of-3 (ensemble shortlist)	226	6.29%
All 3-of-3 (strict shortlist)	116	3.23%

analyzed population rather than errors in the catalogue. A key next step is to expand the training and scoring population to include a substantially larger (ideally complete) set of TESS light curves processed under consistent criteria. With a broader reference distribution, the learned embedding and detector decision boundaries should better reflect the true survey population and may reduce the incidence of catalog-classified objects being selected as anomalies.

### Limitations and future validation

The primary limitation is the absence of ground-truth labels, which prevents direct precision/recall evaluation. Future validation will use injection-and-recovery tests (synthetic transits/flares and controlled artifacts) and systematic crossmatching to quantify detection performance. Additionally, manual vetting can introduce subjective bias; future work will define stricter, reproducible inspection criteria and expand review to larger samples<sup>5,13,25</sup>. Recent work on training anomaly detectors under label noise and contamination includes latent outlier exposure, which could be adapted to light-curve representations<sup>26</sup>.

### Conclusions

We developed an unsupervised anomaly detection pipeline for TESS light curves that integrates reconstruction error from an uncertainty-aware RNN-VAE with two complementary latent-space detectors (HDBSCAN and Isolation Forest). We selected detector operating points using sensitivity analyses and stability-based criteria, yielding screening sets of 398 (11.0%) reconstruction-error candidates, 256 (7.125%) HDBSCAN candidates, and  $\sim 230$  (6.4%) Isolation Forest candidates on objects. We then combined these signals via an agreement-based ensemble, producing both a broad  $\geq 2$ -of-3 shortlist and a strict 3-of-3 shortlist for high-confidence anomalies. This approach demonstrates that stability-driven calibration and consensus ensembling can generate defensible anomaly shortlists in large photometric surveys without labeled data, forming a foundation for catalog cross-matching and targeted scientific follow-up.

### Appendix A. t-SNE latent-space visualization

*t-SNE* is used only as a qualitative visualization of the standardized 16-*D* latent means. Unless otherwise stated, we use *perplexity* = 30, *learning\_rate* = 200, *n\_iter* = 1000, *init* = 'pca', and *random\_state* = 42; we verified that the global structure is similar under  $\pm 10$  changes in perplexity and alternate random seeds (visual check only)<sup>27,28</sup>. We include a *t-SNE* projection of the VAE latent mean vectors as a qualitative visualization of global structure in representation space; all anomaly scoring is performed in the full latent space. Replace the placeholder below with a plot generated from the epoch  $\sim 700$  embeddings (e.g., color by detector flags or ensemble membership).

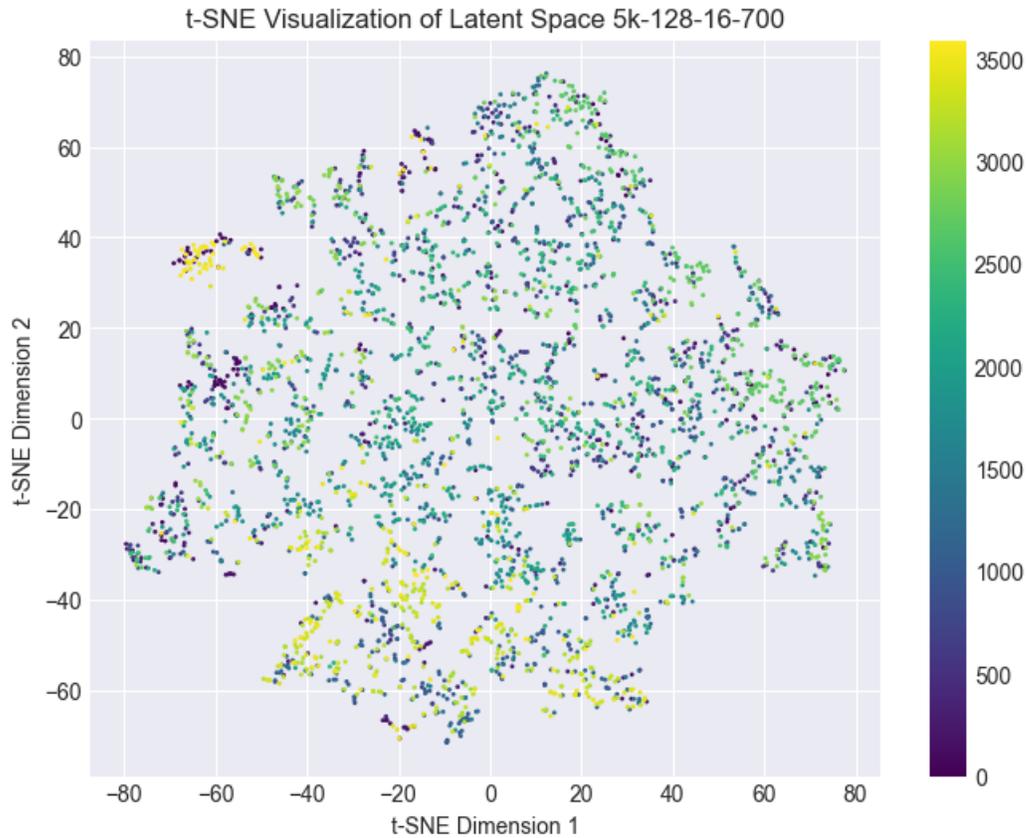
### Appendix B. Detected anomalies in the studied TESS Dataset

#### Detected by 3 detectors

2018eny, 2018eph, 2018fdx, 2018fhw, 2018fwm, 2018fxj, 2018glq, 2018gsg, 2018gum, 2018hql, 2018jjd, 2018jkg, 2018jms, 2018kao, 2018kjp, 2019agu, 2019ajf, 2019alv, 2019bgp, 2019cqi, 2019dtv, 2019erz, 2019esl, 2019gn, 2019hbl, 2019iiv, 2019kli, 2019nfv, 2019nng, 2019obp, 2019pco, 2019rj, 2019rvh, 2019saz, 2019sld, 2019sse, 2019tjl, 2019tlu, 2019tsu, 2019vzy, 2020aayk, 2020abbx, 2020acpv, 2020afq, 2020bxc, 2020ddi, 2020ftl, 2020fyf, 2020hgp, 2020mic, 2020rhn, 2020rkb, 2020rld, 2020rme, 2020tir, 2020tld, 2020uic, 2020wnf, 2020xgz, 2021aal, 2021acet, 2021afte, 2021afth, 2021ageo, 2021agoa, 2021bll, 2021bma, 2021buu, 2021cmf, 2021cxp, 2021dvt, 2021efq, 2021hpf, 2021jhb, 2021ksl, 2021odw, 2021qpb, 2021qut, 2021soz, 2021wue, 2021xeq, 2021xet, 2021xgv, 2021xgy, 2021xhf, 2021xhm, 2021xhq, 2021xkr, 2021xmj, 2021xob, 2021xqg, 2021xrc, 2021xrd, 2021xuk, 2021xzu, 2021ybe, 2021yga, 2021ygg, 2021yib, 2021yjr, 2021yln, 2021ylu, 2021yps, 2021ysk, 2021yuh, 2021yyu, 2021yzz, 2021zct, 2021zdo, 2021zeu, 2021zf, 2022chy, 2022cmr, 2022dsv, 2022duk, 2022ejx

#### Detected by 2 or 3 detectors

2018eny, 2018eph, 2018fdw, 2018fdx, 2018fhw, 2018fwm, 2018fxj, 2018glq, 2018gsg, 2018gum, 2018gxi, 2018hib, 2018hka, 2018hka, 2018hks, 2018hps, 2018hql, 2018izr, 2018jjd, 2018jkg, 2018jms, 2018kao, 2018kjp, 2019agu, 2019ahk,



**Fig. 13** Example normalized light curve (object 2019kli) under different MAD factors  $k$ .

2019ajf, 2019alv, 2019aay, 2019bcp, 2019bgp, 2019caa, 2019cmb, 2019cqi, 2019dfr, 2019dvt, 2019erz, 2019esl, 2019fir, 2019gn, 2019hbl, 2019hnp, 2019iiv, 2019kag, 2019kli, 2019lru, 2019mqc, 2019nfv, 2019nng, 2019nvm, 2019obp, 2019osx, 2019pco, 2019pzj, 2019qck, 2019qoj, 2019rj, 2019rvh, 2019saz, 2019sho, 2019slid, 2019src, 2019sse, 2019sts, 2019tjl, 2019tlu, 2019tng, 2019tsu, 2019tsz, 2019ufy, 2019umq, 2019vzy, 2019xe, 2019zdo, 2019zej, 2020aayk, 2020abbx, 2020acpv, 2020adgm, 2020adsx, 2020afq, 2020bol, 2020bsb, 2020bvg, 2020bxc, 2020ddi, 2020dsl, 2020fit, 2020fyf, 2020hgp, 2020iwn, 2020kav, 2020kur, 2020kif, 2020mic, 2020npl, 2020nrl, 2020nyb, 2020qit, 2020rhn, 2020rkb, 2020rld, 2020rme, 2020rxw, 2020tir, 2020tld, 2020uic, 2020ut, 2020uvg, 2020vku, 2020wnf, 2020wvh, 2020xgu, 2020xgz, 2020xjp, 2020xoy, 2020zo, 2020zui, 2021aal, 2021aanc, 2021aarc, 2021aaxr, 2021aazx, 2021abfd, 2021acet, 2021afte, 2021afth, 2021afvw, 2021ageo, 2021agoa, 2021bll, 2021bma, 2021buu, 2021ckj, 2021cmf, 2021cxp, 2021dvt, 2021efq, 2021flg, 2021gfy, 2021hit, 2021hpf, 2021jhb, 2021kei, 2021koi, 2021ksl, 2021odw, 2021oim, 2021qpb, 2021qut, 2021soz, 2021tod, 2021ttz, 2021wue, 2021wzp, 2021xeq, 2021xet, 2021xgv, 2021xgy, 2021xhb, 2021xhf, 2021xhm, 2021xhq, 2021xiw, 2021xjq, 2021xkr, 2021xku, 2021xmj, 2021xnb, 2021xni, 2021xns, 2021xob, 2021xpe, 2021xph, 2021xpn, 2021xqg, 2021xrc, 2021xrd, 2021xuk, 2021xvu, 2021xxt, 2021xzu, 2021ybe, 2021ybj, 2021yga, 2021ygg, 2021yhq, 2021yib, 2021yic,

2021yiq, 2021yjr, 2021yln, 2021ylu, 2021ynq, 2021yys, 2021yysk, 2021yuh, 2021yvo, 2021yvr, 2021yxf, 2021yyu, 2021yzy, 2021zba, 2021zbt, 2021zct, 2021zdo, 2021zeu, 2021zf, 2022amx, 2022ark, 2022cdv, 2022chy, 2022cmr, 2022djv, 2022dsv, 2022dtj, 2022dto, 2022dtp, 2022duk, 2022duz, 2022ejx, 2022ely, 2022emm, 2022exc, 2022eyw, 2022gaz, 2022kh, 2022oz

**Unclassified objects after cross-checking with public TESS catalogue (by 3 detectors, 104/116)**

2018eny, 2018fwm, 2018fxj, 2018glq, 2018gsg, 2018gum, 2018hql, 2018jkg, 2018jms, 2018kao, 2018kjp, 2019agu, 2019ajf, 2019alv, 2019bgp, 2019cqi, 2019dvt, 2019erz, 2019esl, 2019gn, 2019hbl, 2019kli, 2019nfv, 2019nng, 2019obp, 2019pco, 2019rj, 2019rvh, 2019slid, 2019sse, 2019tjl, 2019tsu, 2019vzy, 2020aayk, 2020abbx, 2020acpv, 2020bxc, 2020ddi, 2020fyf, 2020hgp, 2020mic, 2020rhn, 2020rkb, 2020rld, 2020rme, 2020tir, 2020uic, 2020wnf, 2020xgz, 2021aal, 2021acet, 2021afte, 2021afth, 2021ageo, 2021agoa, 2021bll, 2021buu, 2021cmf, 2021cxp, 2021dvt, 2021efq, 2021hpf, 2021jhb, 2021ksl, 2021odw, 2021qpb, 2021qut, 2021soz, 2021wue, 2021xeq, 2021xet, 2021xgv, 2021xgy, 2021xhf, 2021xhm, 2021xhq, 2021xkr, 2021xmj, 2021xob, 2021xqg, 2021xrc, 2021xrd, 2021xuk, 2021xzu, 2021ybe, 2021yga, 2021ygg, 2021yib, 2021yjr, 2021yln, 2021ylu, 2021yys,

2021ysk, 2021yuh, 2021yzz, 2021zct, 2021zdo, 2021zeu, 2021zf, 2022chy, 2022cmr, 2022dsv, 2022duk, 2022ejx

## Unclassified objects after cross-checking with public TESS catalogue (by 2 or 3 detectors, 198/226)

2018eny, 2018fdw, 2018glq, 2018gsg, 2018fxj, 2018fwm, 2018gum, 2018hql, 2018hps, 2018izr, 2018jms, 2018jkg, 2018kjp, 2019gn, 2019rj, 2018kao, 2019agu, 2019ajf, 2019alv, 2019xe, 2019ahk, 2019ayy, 2019caa, 2019bgp, 2019cqi, 2019cmb, 2019dfr, 2019dvt, 2019erz, 2019esl, 2019fir, 2019hnp, 2019hnb, 2019kag, 2019kli, 2019mqc, 2019nfv, 2019pco, 2019qck, 2019lru, 2019nng, 2019osx, 2019pzj, 2019qoj, 2019rvh, 2019obp, 2019tjl, 2019sld, 2019tng, 2019tsu, 2019tsz, 2019sse, 2019sts, 2019sho, 2019src, 2019umq, 2019vzy, 2019ufy, 2019zej, 2019zdo, 2020zo, 2020bsb, 2020bxc, 2020bvg, 2020ddi, 2020dsl, 2020fyf, 2020hgp, 2020iwn, 2020kav, 2020kur, 2020mic, 2020lkf, 2020nrl, 2020rld, 2020qit, 2020rxw, 2020rhn, 2020wnf, 2020rkb, 2020uic, 2020rme, 2020vku, 2020vwh, 2020tir, 2020zui, 2020xgu, 2020xgz, 2020xoy, 2020xjp, 2020aayk, 2020abbx, 2020acpv, 2020adgm, 2021zf, 2020adsx, 2021bil, 2021buu, 2021cxp, 2021dvt, 2021efq, 2021gfy, 2021cmf, 2021hit, 2021hpf, 2021flg, 2021koi, 2021ksl, 2021jhb, 2021kei, 2021oim, 2021qpb, 2021qut, 2021odw, 2021tod, 2021soz, 2021ttz, 2021xku, 2021xns, 2021xob, 2021xuk, 2021xeq, 2021yiq, 2021zbt, 2021xet, 2021xgv, 2021xgy, 2021xhb, 2021xiw, 2021zba, 2021wzp, 2021xnb, 2021yvo, 2021yvr, 2021xhf, 2021xhm, 2021xxt, 2021yga, 2021yib, 2021yic, 2021ysk, 2021aarc, 2021xjq, 2021xrc, 2021xrd, 2021ynq, 2021xni, 2021yhc, 2021yxf, 2021yzz, 2021ybe, 2021aal, 2021abfd, 2021zct, 2021zdo, 2021aaxr, 2021afvw, 2021afte, 2021afth, 2021ageo, 2021acet, 2021agoa, 2022ark, 2022oz, 2022kh, 2022cmr, 2022amx, 2022dto, 2022chy, 2022cdv, 2022djv, 2022dtj, 2022duk, 2022gaz, 2022ely, 2022duz, 2022ejx, 2022dsv, 2022dtp

## Acknowledgments

I would like to thank Cari Cesarotti for her help and support throughout the process and Tyler Moulton for useful conversations.

## References

- 1 G. Ricker, J. Winn, R. Vanderspek, D. Latham, G. Bakos, J. Bean, Z. Berta-Thompson, T. Brown, L. Buchhave, N. Butler, R. Butler, W. Chaplin, D. Charbonneau, J. Christensen-Dalsgaard, M. Clampin, D. Deming, J. Doty, N. Lee, C. Dressing, E. Dunham, M. Endl, F. Fressin, J. Ge, T. Henning, M. Holman, A. Howard, S. Ida, J. Jenkins, G. Jernigan, J. Johnson, L. Kaltenegger, N. Kawai, H. Kjeldsen, G. Laughlin, A. Levine, D. Lin, J. Lissauer, P. MacQueen, G. Marcy, P. McCullough, T. Morton, N. Narita, M. Paegert, E. Palle, F. Pepe, J. Pepper, A. Quirrenbach, S. Rinehart, D. Sasselov, B. Sato, S. Seager, A. Sozzetti, K. Stassun, P. Sullivan, A. Szentgyorgyi, G. Torres, S. Udry and J. Villaseñor, *Journal of Astronomical Telescopes, Instruments, and Systems*, **1**, 014003.
- 2 J. Jenkins, J. Twicken, S. McCauliff, J. Campbell, D. Sanderfer, D. Lung, M. Mansouri-Samani, F. Girouard, P. Tenenbaum, T. Klaus, J. Smith, D. Caldwell, A. Chacon, C. Henze and C. Heiges, *Proceedings of SPIE*, p. 99133E.
- 3 J. Twicken, D. Caldwell, J. Jenkins, P. Tenenbaum, J. Smith, B. Wohler, M. Rose, E. Ting, R. Vanderspek, E. Morgan, A. Rudat, M. Fausnaugh,

- S. Fleming and E. Quintana, *Rev F*). *NASA/MST technical documentation*.
- 4 V. Chandola, A. Banerjee and V. Kumar, *ACM Computing Surveys*, **41**, 15.
- 5 R. Chalopathy and S. Chawla, *Deep learning for anomaly detection: A survey*, arXiv:1901.03407,.
- 6 D. Kingma and M. Welling, *Auto-encoding variational Bayes*, arXiv:1312.6114,.
- 7 J. Chung, K. Kastner, L. Dinh, K. Goel, A. Courville and Y. Bengio, *A recurrent latent variable model for sequential data*, arXiv:1506.02216,.
- 8 L. McInnes, J. Healy and S., *Journal of Open Source Software*, **2**, 205.
- 9 F. Liu, K. Ting and Z. Zhou, 2008 IEEE International Conference on Data Mining, pp. 413–422,.
- 10 J. An and S. Cho, *Variational autoencoder-based anomaly detection using reconstruction probability*, Snu data mining center technical report (snudm-tr-2015-03),.
- 11 M. Lochner and B. Bassett, *Astronomy and Computing*, **36**, 100481.
- 12 M. Lochner and B. Bassett, *A Hitchhiker™s Guide to Anomaly Detection with Astronomy*, arXiv:2201.10189,.
- 13 S. Fotopoulou, *Astronomy and Computing*, **47**, 100851,.
- 14 F. Hampel, *Journal of the American Statistical Association*, **69**, 383–393.
- 15 D. Rezende, S. Mohamed and D. Wierstra, *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pp. 1278–1286,.
- 16 K. Cho, B. Merrinboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, *Proceedings of EMNLP*, pp. 1724–1734.
- 17 P. Malhotra, L. Vig, G. Shroff and P. Agarwal, *European Symposium on Artificial Neural Networks (ESANN)*.
- 18 D. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv:1412.6980,.
- 19 R. Campello, D. Moulavi and J. Sander, *Advances in Knowledge Discovery and Data Mining (PAKDD)*, pp. 160–172.
- 20 S. Webb, M. Lochner, D. Muthukrishna, J. Cooke, C. Flynn, A. Mahabal, S. Goode, I. Andreoni, T. Pritchard and T. Abbott, *Monthly Notices of the Royal Astronomical Society*, **498**, 3077–3094,.
- 21 A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. K[ö]pf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala, *Advances in Neural Information Processing Systems (NeurIPS)*.
- 22 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. VanderPlas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, *Journal of Machine Learning Research*, **12**, 2825–2830.
- 23 J. Lightkurve Collaboration, C. Hedges, M. Gully-Santiago, N. Saunders, A. Cody, T. Barclay, A. Hall, T. Sagar, J. Turtelboom, W. Zhang, A. Tzanidakis, K. Mighell, J. Coughlin, E. Bell, Z. Berta-Thompson, P. Williams, A. Dotson and G. Barentsen, ascl:1812.013.
- 24 D. Caldwell, P. Tenenbaum, J. Twicken, J. Jenkins, E. Ting, J. Smith, C. Hedges, M. Fausnaugh, M. Rose and C. Burke, *Research Notes of the AAS*, **4**, 201.
- 25 W.Zhang, *TESS-light-curve-anomaly-detection*, GitHub repository, 2026.
- 26 C. Qiu, A. Li, M. Kloft, M. Rudolph and S. Mandt, *International Conference on Machine Learning (ICML)*, pp. 18153–18167.
- 27 L. Maaten and G. Hinton, *Journal of Machine Learning Research*, **9**, 2579–2605.
- 28 L. Maaten, *Journal of Machine Learning Research*, **15**, 3221–3245.