UniTED: A Unified Time Series Event Detection Repository

Janio Lima[1,2], Hélio Castro[1], Luiz Oliveira[1,2], Ellen Paixão[1], Lais Baroni[1], Rebecca Salles[3], Ricardo Vargas[2], Eduardo Ogasawara[1]

[1]Federal Center for Technological Education of Rio de Janeiro (CEFET/RJ)

[2]Petróleo Brasileiro S.A. (Petrobras)

[3]INRIA

{janio.lima, helio.castro, luiz.oliveira.7}@aluno.cefet-rj.br,

{ellenpaixao, lais.baroni}@aluno.cefet-rj.br, rebecca.pontes-salles@inria.fr,

ricardo.vargas@petrobras.com.br, eogasawara@ieee.org

Abstract. *Event detection in time series is essential for numerous real-world applications, from monitoring industrial systems to identifying health anomalies. Public annotated datasets are crucial for benchmarking, training, and validating detection models. Despite recent advances in the field, there is a lack of a standardized and unified repository for evaluating different event types, which limits progress in reproducibility, comparability, and model development. This paper presents the UniTED, a Unified Event Detection Dataset for time series. UniTED consolidates annotated series from diverse domains and offers a common format and protocol for evaluation. The repository supports three event types: anomalies, change points, and motifs. UniTED fosters reusability and reproducibility, contributing to improved performance assessment and model generalization across data analysis tasks. However, existing datasets have limitations, including poor standardization, a lack of annotation guidelines, limited support for different event types, and difficulties in automating performance evaluation. UniTED presents a harmonized ETL process, label and annotation conventions, and an open-source implementation. Three use cases are presented to demonstrate the applicability of the dataset.*

1. Introduction

The availability of datasets remains a critical factor in data analysis tasks, as data volume, class balance, domain representativeness, and diversity of data types directly influence the performance assessment and generalization of scientific and technical results. In the context of time series, an analytical challenge is event detection, which involves identifying points or subsequences whose behavior deviates from expected patterns. This task encompasses three main categories: anomalies, change points, and motifs [Ogasawara et al., 2025].

The evaluation of detection methods relies on classification-based metrics, such as accuracy, precision, recall, and F_1-score, which require labeled time series [Han et al., 2012; Ahmad et al., 2017]. However, existing repositories rely on heterogeneous file structures, naming conventions, and annotation protocols, which require manual preprocessing and hinder experimental reproducibility.

To address these limitations, this work introduces the UniTED repository, a publicly available repository of labeled time series designed to support reproducible research on event detection. UniTED results from an ETL process that standardizes data formats, harmonizes event annotations, and ensures consistent access across series. Unlike previous resources, UniTED integrates datasets annotated for anomalies, change points, and motifs under a common interface, enabling benchmarking across event types and application domains.

This repository currently consolidates six curated datasets from diverse domains, including water quality monitoring, network traffic, cloud computing infrastructure, health signals, oil extraction systems, and financial indicators. All series are publicly accessible, and their annotations have been adapted or consolidated to ensure consistency in event types. Additionally, this repository incorporates data cleaning procedures, annotation harmonization, and a simplified access interface to support experimentation and facilitate experiment execution. As UniTED supports three event types (anomalies, change points, and motifs), handling both univariate and multivariate time series, and enabling temporal-tolerant evaluation through standardized annotations, it distinguishes itself from prior repositories such as TimeEval [Wenig et al., 2022] and UCR Anomaly Archive (UCR AA) [Wu and Keogh, 2023], which are typically limited to anomalies and univariate data.

In addition to this introduction, the paper is organized as follows. Section 2 presents the literature review. Section 3 describes the method used to organize, standardize, and label the series included in the repository. Section 4 presents case studies and demonstrations of how this repository can be used to evaluate event detection methods. Finally, Section 5 provides concluding remarks.

2. Literature Review

Time Series events are classified into three types: (i) anomalies, defined as isolated observations that deviate from expected behavior; (ii) change points, defined as transition timestamps between stationary regimes; and (iii) motifs, defined as recurring subsequences of interest [Ogasawara et al., 2025]. In event detection, each observation is assessed to determine whether it corresponds to an event of interest, producing a logical decision associated with a timestamp [Chandola et al., 2009]. Evaluation of event detections requires reference labels, typically annotated by domain experts, enabling the evaluation of metrics such as precision, recall, and F_1-score derived from a confusion matrix [Han et al., 2012].

While most comparisons rely on traditional classification metrics, others incorporate time-sensitive variants tailored to the temporal nature of the problem [Salles et al., 2020; Lima et al., 2024; Salles et al., 2024]. This includes the Numenta Anomaly Benchmark (NAB) [Ahmad et al., 2017]. It provides a complete evaluation framework, including a dataset, scoring function (NAB Score), and baseline performance results for several detection methods.

Studies such as Lomio et al. [2020] and Wu and Keogh [2023] highlight biases in widely used benchmarks, including unrealistic anomaly densities and mislabeling. Wu and Keogh [2023] proposes the UCR AA as a curated alternative, addressing issues of trivial event scenarios, label density inconsistency, mislabeling of true/false segments, and sampling bias. Lomio et al. [2020] further notes the scarcity of datasets targeting resource monitoring in computing systems and contributes a new dataset in this domain.

Several public repositories have supported comparative studies, including Yahoo Labs [webscope, 2015], NAB [Ahmad et al., 2017], GECCO Challenge 2018 (GECCO) [Moritz et al., 2018], and UCR AA [Wu and Keogh, 2023]. These datasets differ substantially in event types, labeling protocols, and file structures. For instance, Yahoo Labs comprises four subgroups: A1, which features real data, and A2-A4, which include synthetic series with contextual anomalies and trend reversals [Duraj et al., 2025]. The GECCO dataset focuses on water quality monitoring, offering multivariate signals and labeled events annotated as anomalies to support complex detection scenarios. The MIT-BIH Arrhythmia Database (MIT-BIH) [Moody and Mark, 2001], one of the earliest resources for physiological data, includes over 110,000 expert-annotated ECG events from 47 patients. The 3W Dataset [Vargas et al., 2019], derived from Petrobras oil well monitoring systems, comprises real, simulated, and manually crafted time series sampled at 1 Hz, labeled with eight event classes related to abnormal extraction processes. Additionally, TimeEval [Wenig et al., 2022] provides a standardized and extensive set of anomaly detection datasets, facilitating consistent benchmarking across multiple scenarios.

Despite their utility, these datasets present heterogeneous formats and annotation definitions, which often require the development of distinct preprocessing procedures to enable their joint use in evaluation pipelines. This lack of standardization limits the reproducibility and scalability of experimental workflows. Most existing repositories focus exclusively on one event type, typically anomalies or change points, and do not support integrated evaluation of multiple types. Additionally, inconsistencies in labeling conventions and data structure hinder comparative analysis.

The UniTED repository addresses these gaps by harmonizing datasets across domains and standardizing their format and annotation schema. This initiative enables reproducible benchmarking for diverse event types, including anomalies, change points, and motifs, under a unified data model.

3. Method

The construction of the UniTED repository was guided by the need to support reproducible research in event detection across diverse application domains. To achieve this goal, six publicly available datasets were selected based on three criteria: (i) the presence of event labels, (ii) diversity in domains and data sources, and (iii) representation of different event types (anomalies, change points, and motifs). The selected datasets are listed in Table 1.

Each dataset was restructured through an ETL process to conform to a standardized data model. The transformation pipeline included label normalization rules, such as converting interval-based annotations to point-wise markers at the first timestamp of each segment. This design choice enables compatibility with classification-based evaluation metrics (e.g., precision, recall, F_1-score), simplifies annotation consistency, and facilitates the application of binary event detectors. To mitigate potential semantic loss, the original segment length is preserved through the `seqlen` field, and temporal-tolerant metrics such as SoftED [Salles et al., 2024] can be employed to evaluate detection accuracy within a permissible delay window. All transformations preserved the semantics of original event annotations while enabling unified storage and access. A detailed description of these rules is available in the repository documentation[1]. The series are stored as `list`-type objects in the R programming language, with harmonized fields and labeling conventions according to event type.

Table 1 summarizes the datasets included in the repository. The column Data describes the number of series groups, whether the series are univariate or multivariate, and whether the labels were adapted. The column Annotation indicates the suitability of each dataset for detecting anomalies (AN) and change points (CP), as well as the discovery of motifs (M). The structure of the series is described in Figure 1, the details and method of use of which can be accessed in the repository and Section 4.

Table 1. Content of the UniTED Repository

	Data			Annotation		
Dataset	Groups	Type	Labels	AN	CP	M
3W [Vargas et al., 2019]	7	Multivariate	Adapted		X	
UCR AA [Wu and Keogh, 2023]	4	Univariate	Adapted	X		
MIT-BIH [Moody and Mark, 2001]	1	Univariate	Adapted			X
Yahoo [webscope, 2015]	4	Univariate	Unchanged	X	X	
NAB [Ahmad et al., 2017]	6	Univariate	Unchanged	X		
GECCO [Moritz et al., 2018]	1	Multivariate	Unchanged	X		

All univariate series follow a common structure with four columns: `idx`, `value`, `event`, and `type`. Here, `idx` is a sequential integer index representing the timestamp. Additionally, if there are actual dates in a timestamp column in the original data, this is represented by `timestamp`. Column `value` is the observed measurement, `event` is a logical attribute indicating whether the observation is labeled as an event (TRUE) or not (FALSE), and `type` specifies the event type (anomaly, change point, or motif). In the case of sequence anomalies or motifs, an additional column, named `seqlen`, is included to characterize the sequence size.

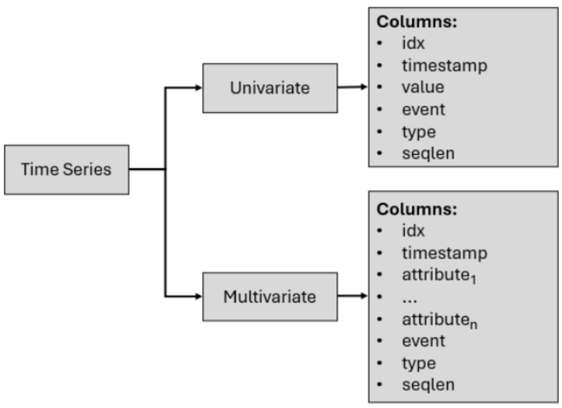

Figure 1. Standard structure for labeled time series

For multivariate datasets, the structure is similar, but the column `value` is replaced by multiple columns, each named according to the variable they represent (e.g.,

[1]https://github.com/cefet-rj-dal/united

`attribute`$_1$, $\cdots$, `attribute`$_n$). The other columns remain consistent and carry the same semantics as in the univariate case.

In both cases, the `list` object is named according to the datasets[1] in Table 1 and contains the series as objects of the `data.frame` type of the R language, whose columns follow the pattern in Figure 1.

The transformation pipeline included additional steps to ensure usability and consistency: (i) all measurements were converted to a numeric format; (ii) all interval-based labels were converted to single-point labels with a sequence length (seqlen), and (iii) labels were adapted to support the expected event detection task.

4. Experimental Evaluation

This section presents the application of three representative event detection tasks (anomaly detection, change point detection, and motif discovery) on the UniTED repository. Although not exhaustive, these tasks illustrate how the unified structure and standardized labeling of the repository enable reproducible evaluations across distinct event types. Time series are loaded using standardized examples[1], which require minimal preprocessing and enable the direct application of detection methods.

4.1. Anomaly Detection Case Study

This case study employs a univariate series from the UCR AA [Wu and Keogh, 2023], which provides labeled intervals for domain-specific anomaly events, primarily in health monitoring. To align with the UniTED format, these interval-based labels were transformed into logical vectors, where each anomaly interval is represented by a single marker at its first index, recorded as `TRUE` in the `event` column. This preserves the original semantics while enabling point-wise evaluation and reducing ambiguity in performance metrics.

Detections were evaluated using precision, recall, F_1-score, and accuracy, highlighting the repository's ability to support standardized experimentation. Figure 2 illustrates the result of applying a baseline anomaly detector based on a simple moving average to the selected series, using the Harbinger framework [Salles et al., 2020]. The red dot at index 17,024 represents the detected anomaly, while the blue dot at index 16,900 marks the ground truth event label.

Although the exact index of the ground truth was not detected, the result is considered valid due to its proximity to the true event. This evaluation yields an accuracy of 0.99 and an F_1-score of 0.39. Moreover, the detected point at index 17,024 falls within the valid range [16,900, 17,100] defined in the UCR AA documentation [Wu and Keogh, 2023], further supporting its correctness under interval-based assessment. In this context, metrics such as SoftED [Salles et al., 2024] enable approximate matching between predictions and ground truth within a temporal window, allowing for tolerant evaluation strategies that complement the standard point-wise approach.

4.2. Change Point Detection Case Study

This case study explores a multivariate series from the 3W Dataset [Vargas et al., 2019], which comprises both simulated and real sensor data from oil well systems. The task

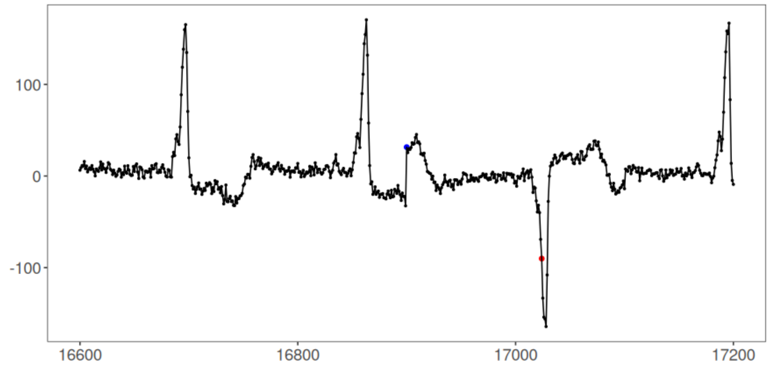

Figure 2. AN case study: UCR - ECG - series 096_ECG2

involves identifying behavioral transitions in a selected variable using a change point detector. The original labels in the 3W Dataset denote contiguous operational segments. For change point detection, these are redefined to mark the onset of each new segment. Such points are encoded in the `event` column with the event type CP, representing a distributional change in the observed process.

Figure 3 illustrates the result of applying a change point detector with the Harbinger framework [Salles et al., 2020] to the variable P-TPT. The dashed vertical lines indicate the detections, and the blue dots mark the ground truth change points.

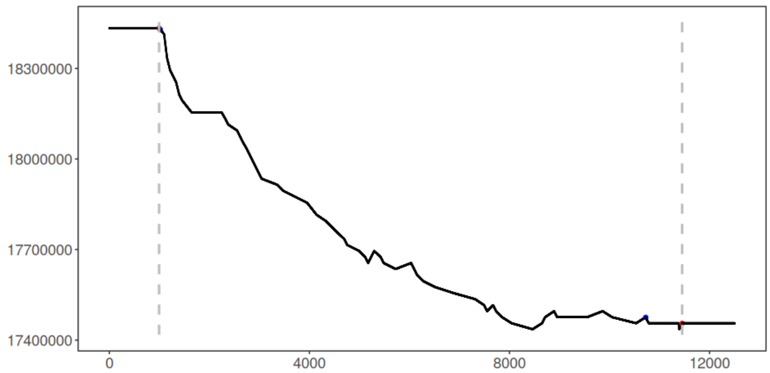

Figure 3. CP case study: 3W - Group 1 - First well - Variable P-TPT

Although the exact change points were not achieved, the detections are considered valid under SoftED due to their temporal proximity to the labeled transitions. This evaluation yields an accuracy of 0.99 and an F_1-score of 0.38, demonstrating the repository's suitability for tolerant and latency-sensitive assessment of change point detectors.

4.3. Motif Discovery Case Study

For motif discovery, a segment of the MIT-BIH dataset [Moody and Mark, 2001] is selected, containing repeated occurrences of a specific heartbeat morphology. A pattern-based motif detector is applied to identify repeated structures in the ECG waveform.

The dataset comprises ECG recordings from 47 individuals, sampled at 360 Hz, with over 110,000 annotations manually labeled by cardiologists. In this context, motifs are defined as recurring cardiac events of clinical relevance. The ETL pipeline extracts the first five series for each sensor and annotation. Annotations are aligned to the start index of each repeated pattern with a `seqlen` equal to 50, and motif instances are encoded as `event` entries with the event type motif. The resulting structure enables motif discovery using consistent indexing and semantic annotations.

Figure 4 presents the results obtained from applying the SAX-based motif discovery algorithm implemented in the Harbinger framework [Salles et al., 2020] to the variable `value`. The red dots indicate the detections, while the blue dots mark the ground-truth motifs.

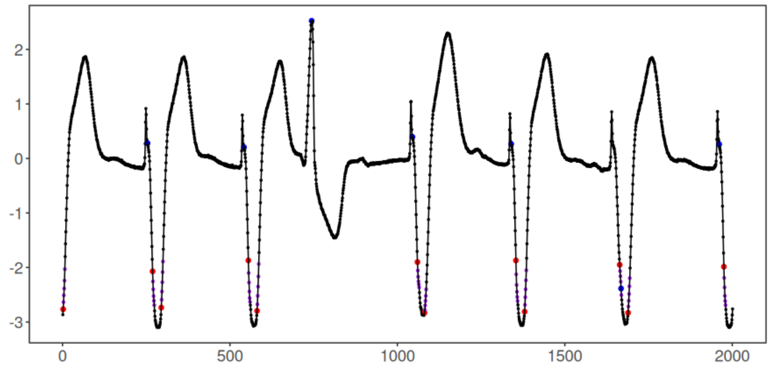

Figure 4. Motif case study: MIT-BIH - r102 - Variable value

Even though the exact positions of cardiac events were not precisely identified, the detections are considered meaningful under the SoftED evaluation framework due to their temporal proximity to the annotated motifs. This assessment yields an accuracy of 0.99 and an F_1-score of 0.55, underscoring the repository's effectiveness in evaluating motif discovery algorithms with temporal tolerance in physiological time series analysis.

4.4. Summary of Results

These case studies underscore the repository's capacity to support diverse event detection scenarios within a unified framework. Its harmonized data format, explicit labeling by event type, and domain variety minimize preparation overhead and enable consistent evaluation of detection algorithms.

Although the case studies employ simple baselines, they illustrate how the repository's structure promotes fair comparison and straightforward metric computation. This design makes it especially valuable for research on online detection, weak supervision, and time-aware evaluation.

5. Conclusion

This paper presents the UniTED, a unified repository of labeled time series designed to support reproducible research in event detection. UniTED addresses limitations of existing resources by offering a standardized interface, harmonized annotation structure, and multi-domain coverage. It supports anomalies, change points, and motifs using a consistent annotation schema and access method. This repository consolidates six publicly available datasets from various domains, including sensor monitoring, industrial systems, financial data, and biomedical signals. Each dataset undergoes a structured ETL process to unify formats, clean and adapt labels, and ensure compatibility with a common interface in the R environment, preserving data semantics while standardizing annotations. The standardized structure also facilitates experimentation and uniform access to metadata and annotations.

Acknowledgments

The authors gratefully acknowledge the maintainers of the original datasets adopted by UniTED and the partial support received from CNPq, CAPES, and FAPERJ.

References

Ahmad, S., Lavin, A., Purdy, S., and Agha, Z. (2017). Unsupervised real-time anomaly detection for streaming data. *Neurocomputing*, 262:134 – 147.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3).

Duraj, A., Szczepaniak, P. S., and Sadok, A. (2025). Detection of anomalies in data streams using the lstm-cnn model. *Sensors*, 25(5).

Han, J., Kamber, M., and Pei, J. (2012). *Data Mining: Concepts and Techniques*. Elsevier.

Lima, J., Tavares, L. G., Pacitti, E., Ferreira, J. E., Santos, I., Siqueira, I. G., Carvalho, D., Porto, F., Coutinho, R., and Ogasawara, E. (2024). Online Event Detection in Streaming Time Series: Novel Metrics and Practical Insights. In *Proceedings of the IJCNN 2024*.

Lomio, F., Baselga, D. M., Moreschini, S., Huttunen, H., and Taibi, D. (2020). RARE: A labeled dataset for cloud-native memory anomalies. In *MaLTeSQuE 2020*, pages 19 – 24.

Moody, G. and Mark, R. (2001). The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50.

Moritz, S., Rehbach, F., Chandrasekaran, S., Rebolledo, M., and Bartz-Beielstein, T. (2018). GECCO Industrial Challenge 2018 Dataset. Technical report, https://zenodo.org/record/3884398.

Ogasawara, E., Salles, R., Porto, F., and Pacitti, E. (2025). *Event Detection in Time Series*. Synthesis Lectures on Data Management. Springer Nature Switzerland, Cham, 1 edition.

Salles, R., Escobar, L., Baroni, L., Zorrilla, R., Ziviani, A., Kreischer, V., Delicato, F., Pires, P. F., Maia, L., Coutinho, R., Assis, L., and Ogasawara, E. (2020). Harbinger: Um framework para integração e análise de métodos de detecção de eventos em séries temporais. In *Anais do Simpósio Brasileiro de Banco de Dados (SBBD)*, pages 73–84. SBC.

Salles, R., Lima, J., Reis, M., Coutinho, R., Pacitti, E., Masseglia, F., Akbarinia, R., Chen, C., Garibaldi, J., Porto, F., and Ogasawara, E. (2024). SoftED: Metrics for soft evaluation of time series event detection. *Computers and Industrial Engineering*, 198.

Vargas, R. E. V., Munaro, C. J., Ciarelli, P. M., Medeiros, A. G., do Amaral, B. G., Barrionuevo, D. C., de Araújo, J. C. D., Ribeiro, J. L., and aes, L. P. M. (2019). A realistic and public dataset with rare undesirable real events in oil wells. *Journal of Petroleum Science and Engineering*, 181.

webscope (2015). S5 - A Labeled Anomaly Detection Dataset, version 1.0. Technical report, https://webscope.sandbox.yahoo.com/catalog.php?datatype=s&did=70.

Wenig, P., Schmidl Sebastian, S., and Papenbrock, T. (2022). TimeEval: A Benchmarking Toolkit for Time Series Anomaly Detection Algorithms. *Proceedings of the VLDB Endowment*, 15(12):3678 – 3681.

Wu, R. and Keogh, E. J. (2023). Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress. *IEEE Transactions on Knowledge and Data Engineering*, 35(3):2421 – 2429.