

Benchmarking Anomaly Detection in Streaming Time Series: Current Landscape and Where to Go Next

Magali Parrino^{1,2}, Antoine Ajenjo¹, Emmanuel Remy¹, Pierre Stephan¹,
Pierre Senellart², Paul Boniol²

¹ EDF R&D, PRISME, Chatou, France (first.last@edf.fr)

² Inria, ENS, CNRS, Project-team VALDA, Paris, France (first.last@inria.fr)

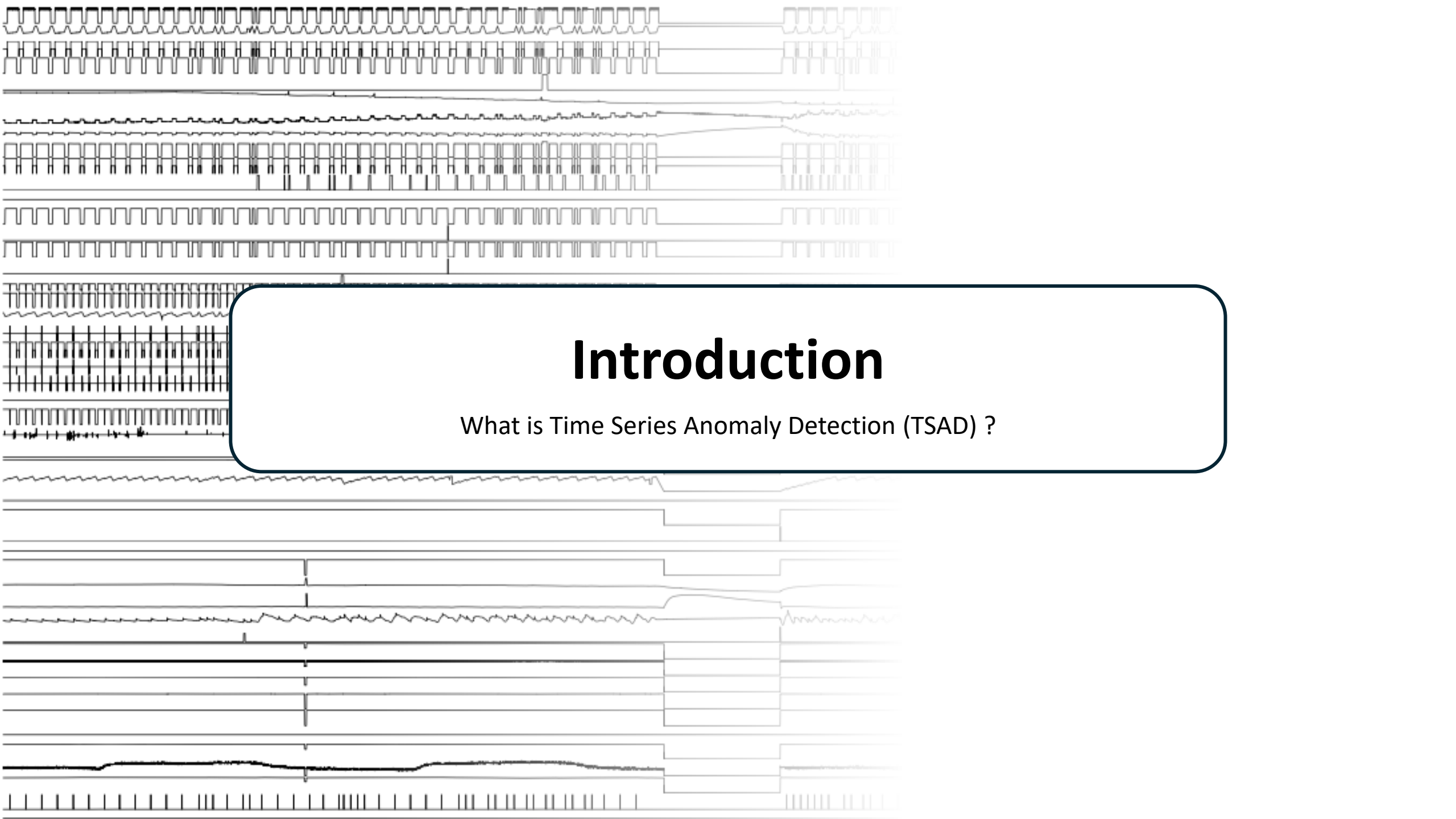




Trigger Warning



(et en plus
c'est un ppt)

The background of the slide is composed of a dense, overlapping collage of various time series plots. These include line graphs with different trends (linear, step, fluctuating), bar charts, and plots with vertical markers. The plots are rendered in a light, semi-transparent style, creating a technical and data-driven atmosphere.

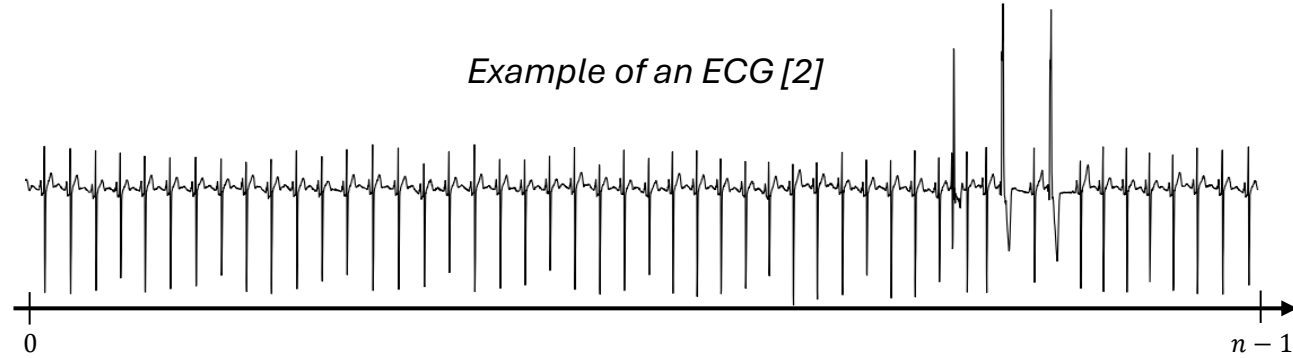
Introduction

What is Time Series Anomaly Detection (TSAD) ?

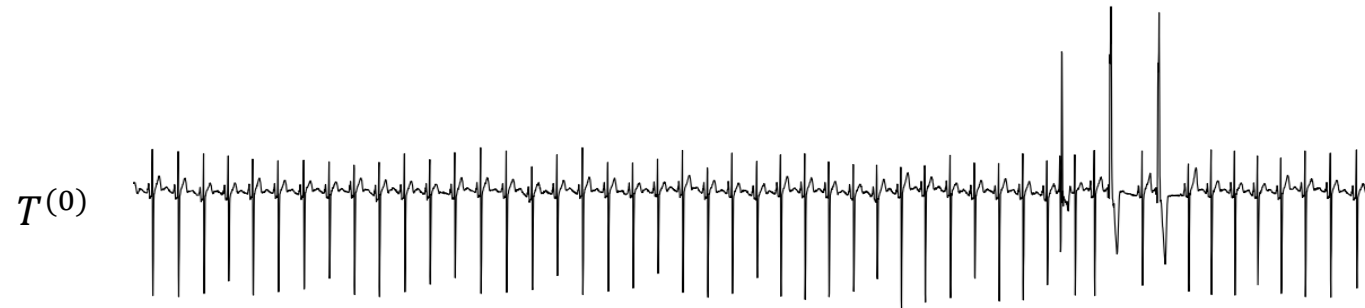
Introduction: Time Series

An **univariate time series** T of length n is an ordered sequence of n values

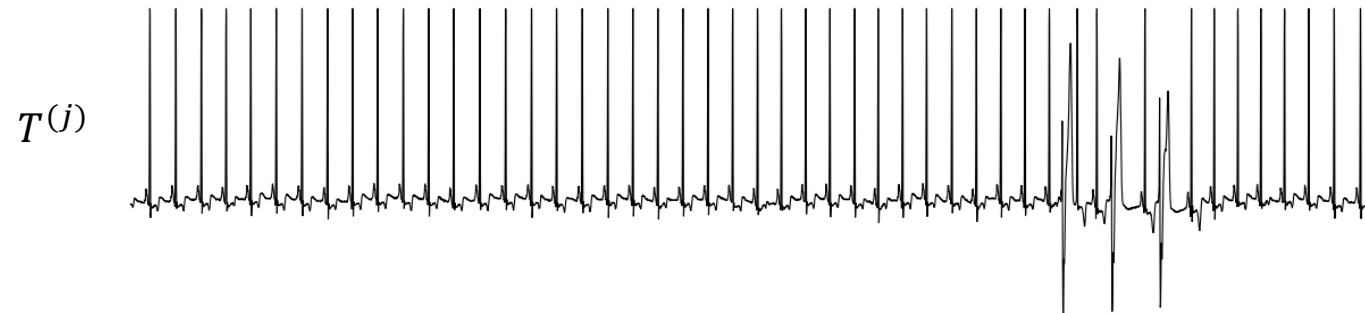
Example of an ECG [2]



Introduction: Time Series



...



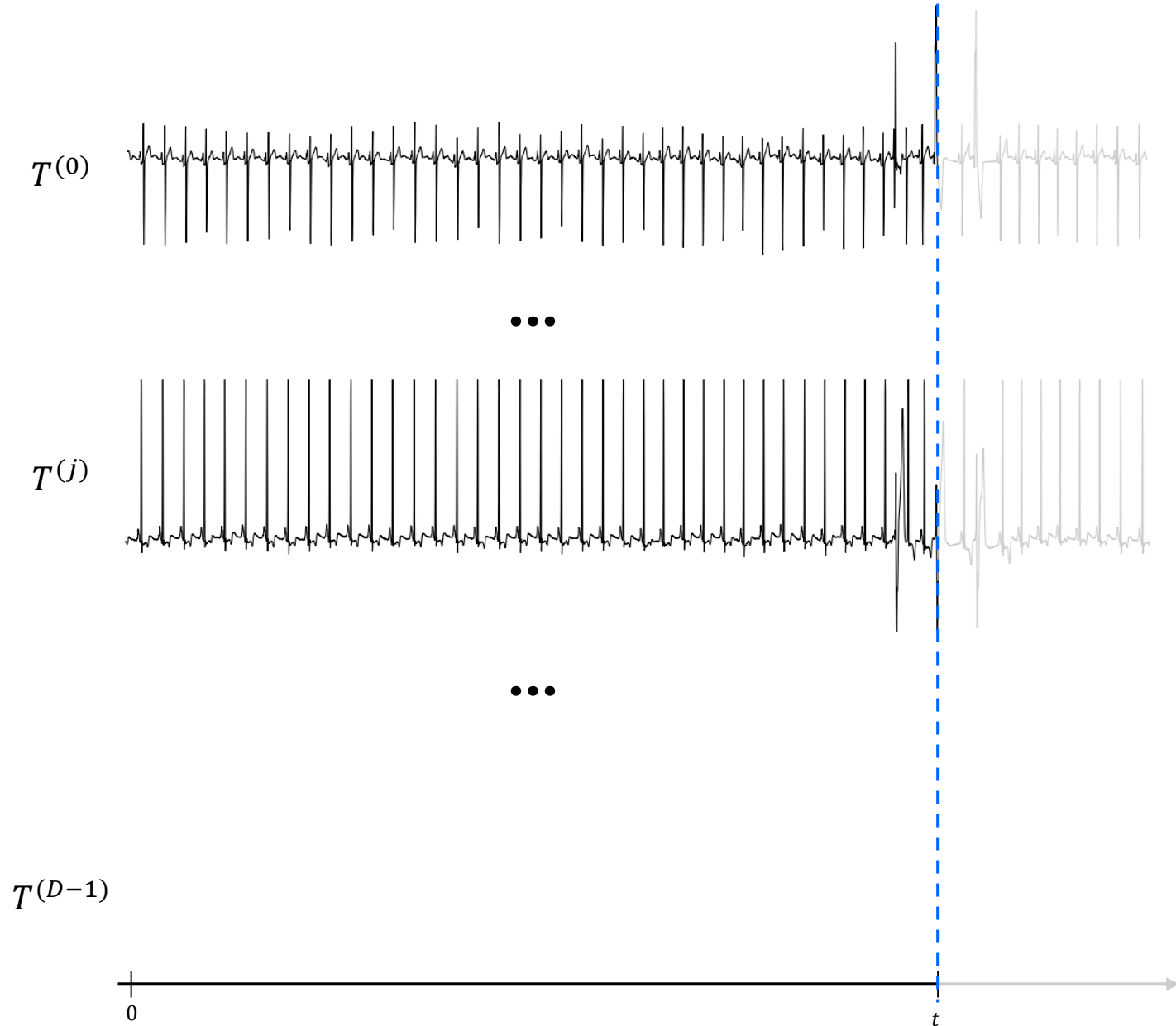
...



A **multivariate time series** T is
a collection of D univariate
time series

Introduction: Time Series

A **streaming** time series T is a collection of D ordered, **unbounded**, sequences of values



Introduction: Time Series Anomalies

Time Series Anomaly : an observation that deviates so significantly it raises suspicions.

Introduction: Time Series Anomalies

Time Series Anomaly : an observation that deviates so significantly it raises suspicions.

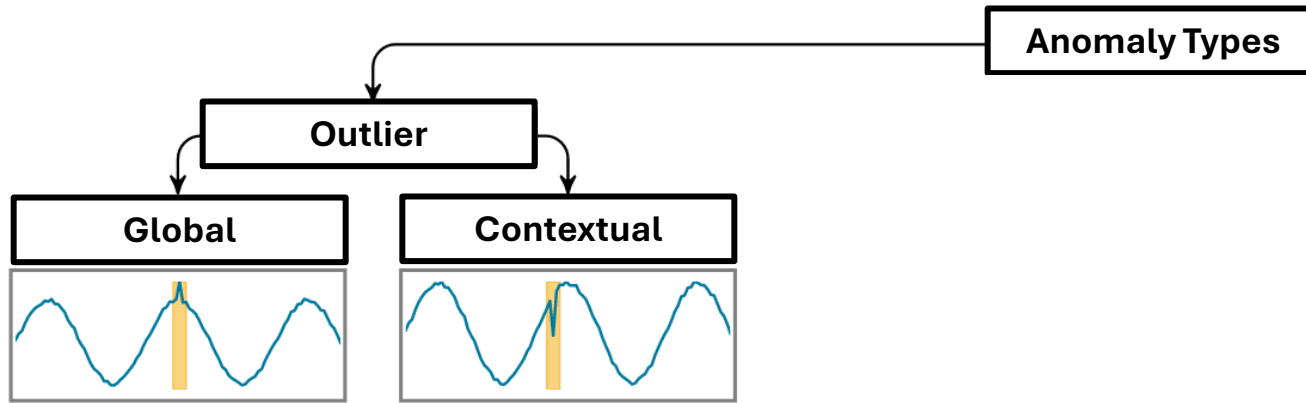


Figure 2: *Anomaly taxonomy and illustrations [3]*

Introduction: Time Series Anomalies

Time Series Anomaly : an observation that deviates so significantly it raises suspicions.

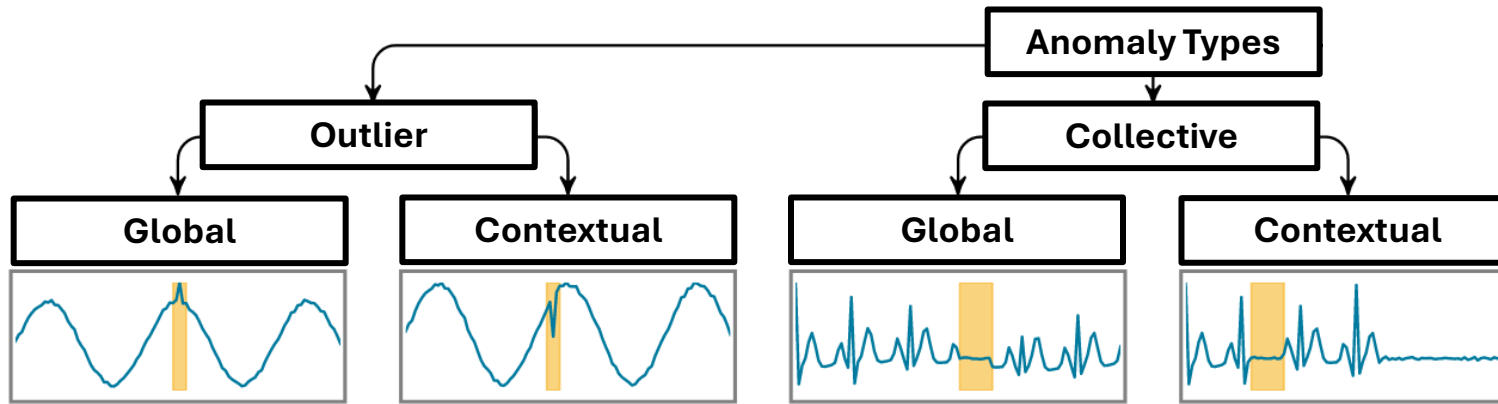


Figure 2: *Anomaly taxonomy and illustrations [3]*

Introduction: Time Series Anomalies

Time Series Anomaly : an observation that deviates so significantly it raises suspicions.

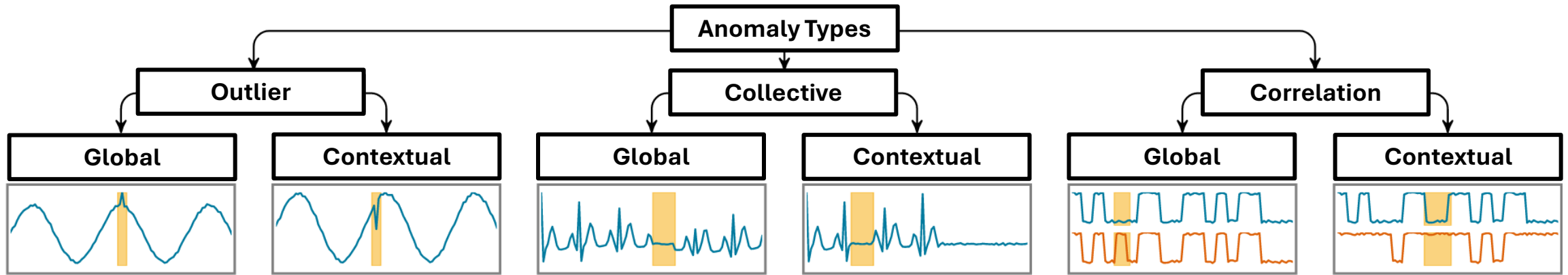


Figure 2: Anomaly taxonomy and illustrations [3]

Introduction: Time Series Anomalies

Time Series Anomaly : an observation that deviates so significantly it raises suspicions.

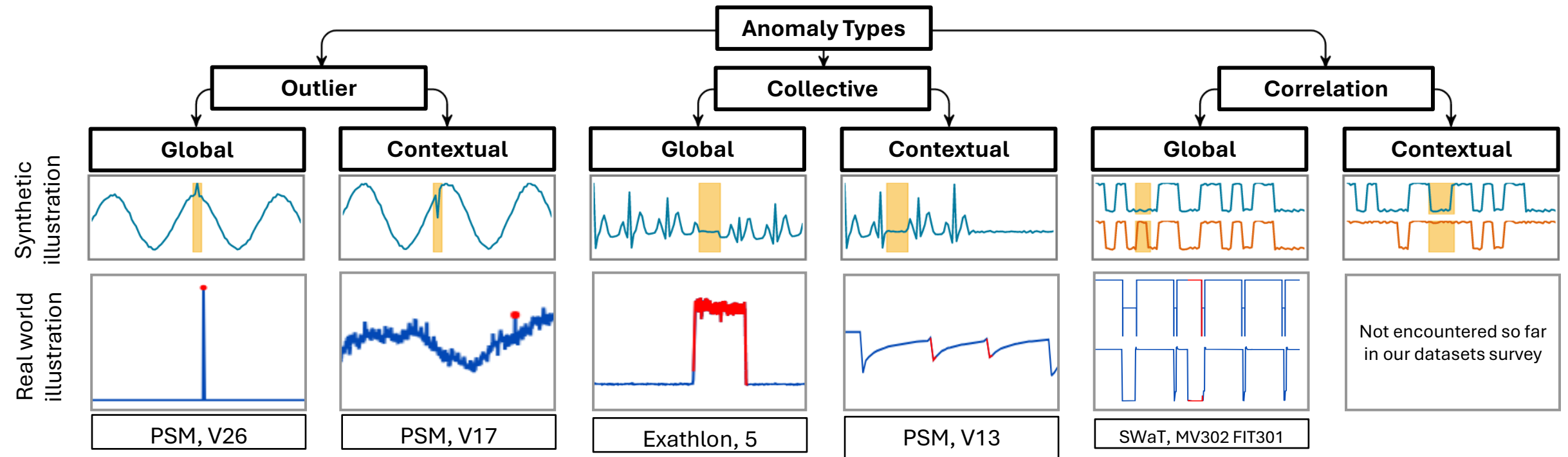
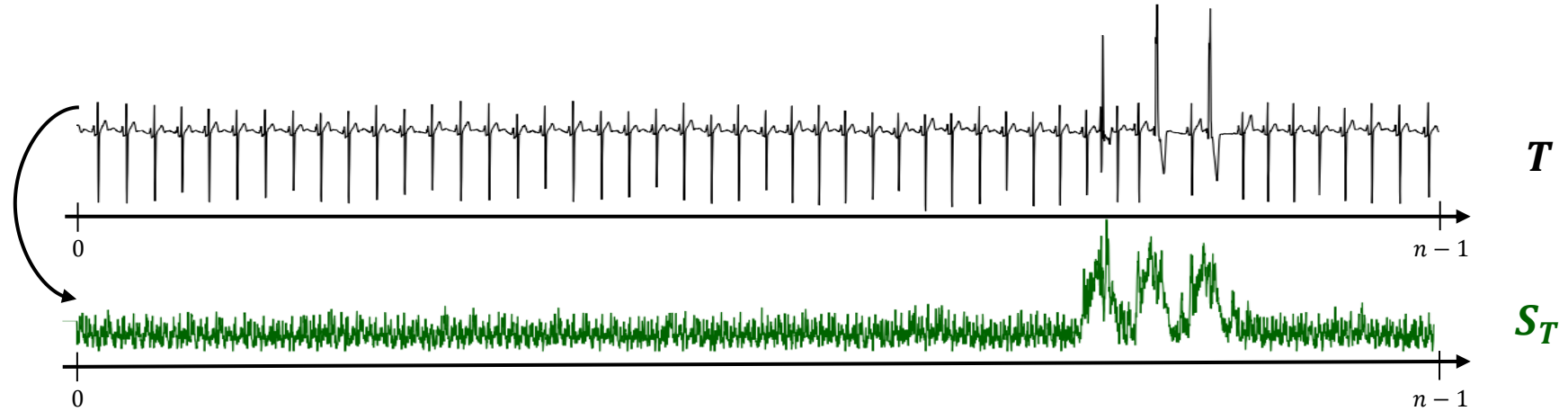


Figure 2: Anomaly taxonomy and illustrations [3]

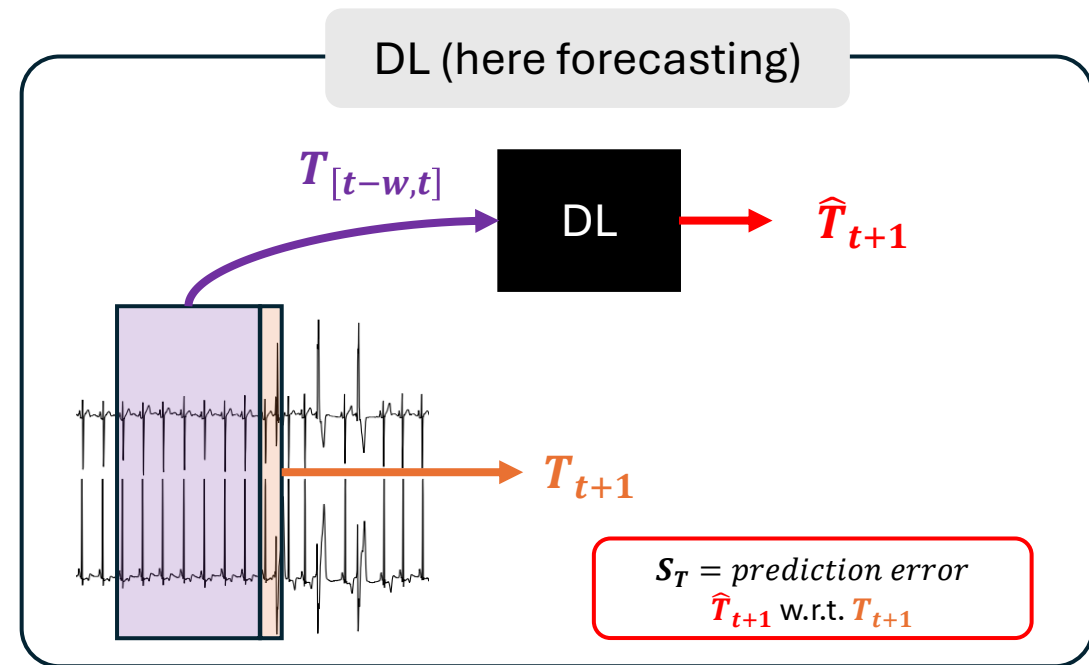
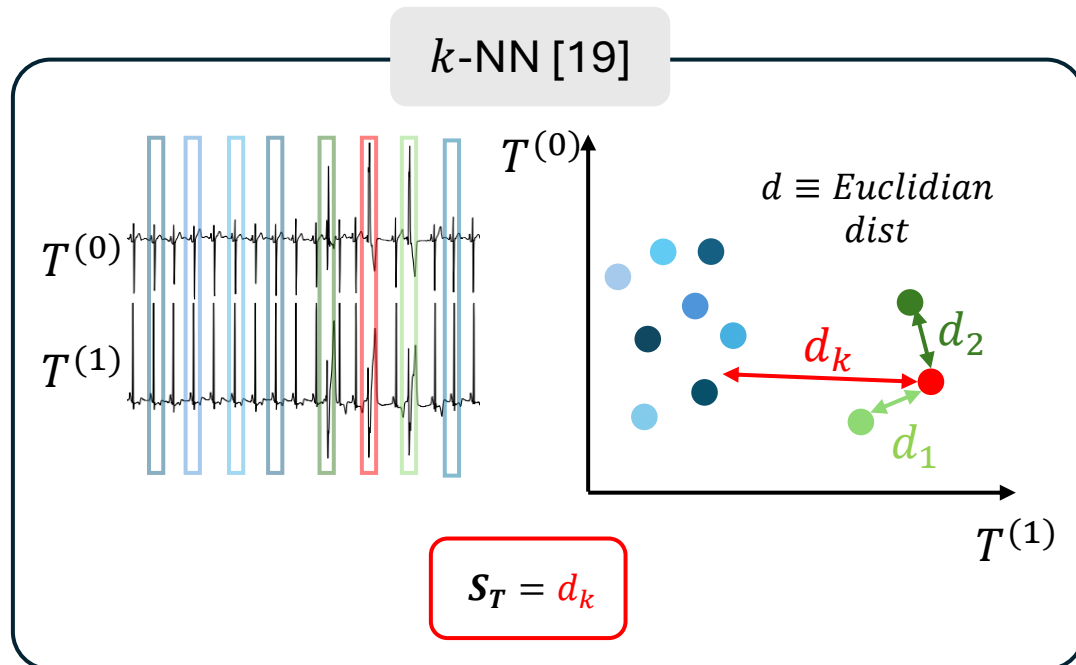
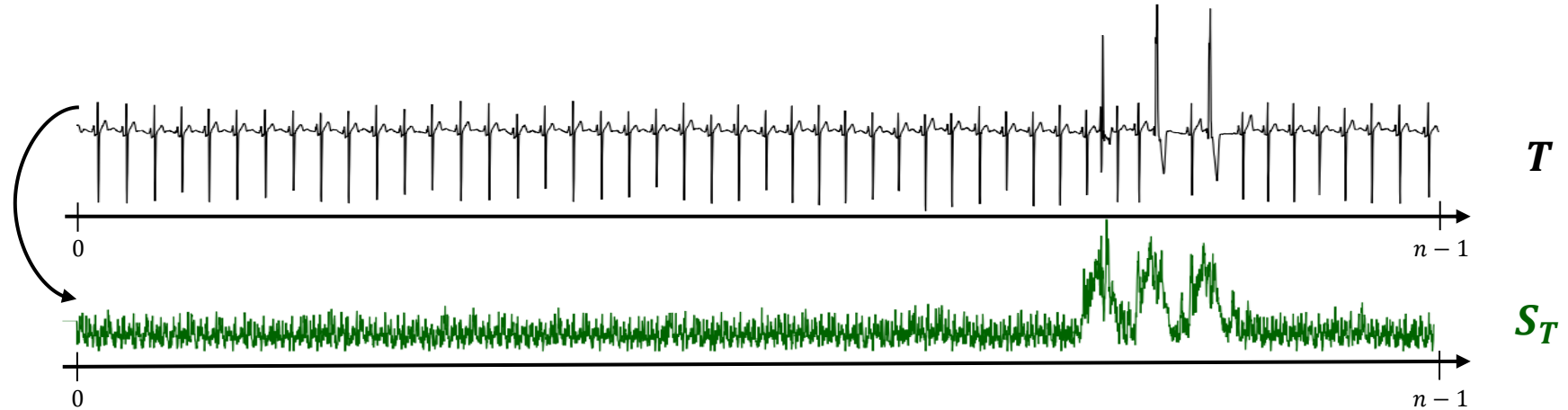
Introduction: Time Series Anomaly Detection (TSAD)

An **Anomaly Detection algorithm** takes a time series T and returns an **Anomaly Score S_T**

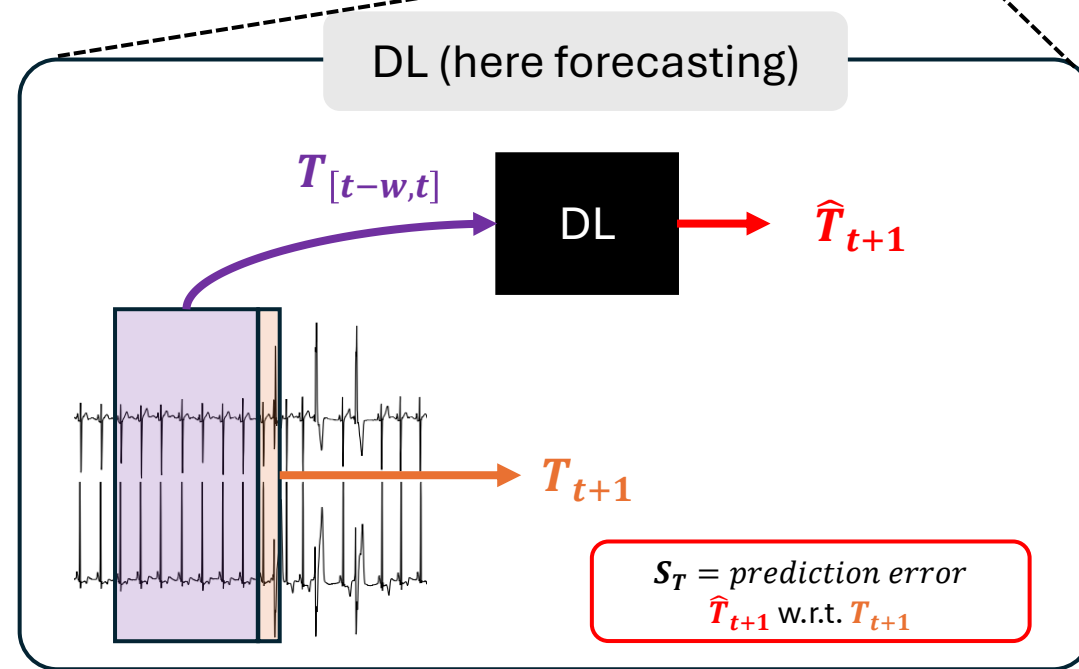
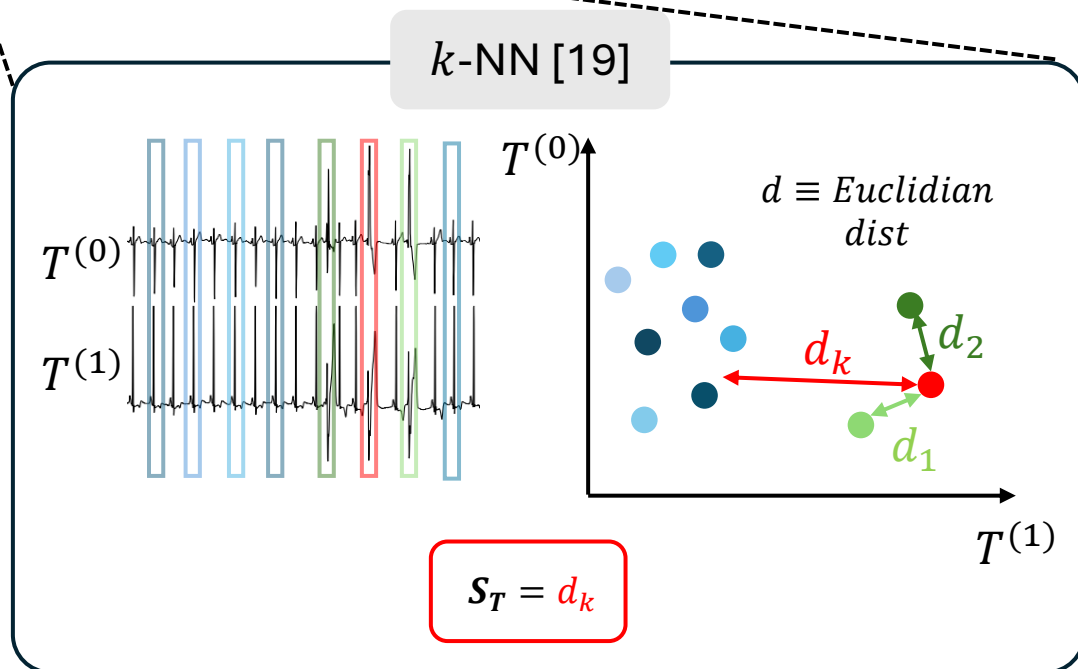
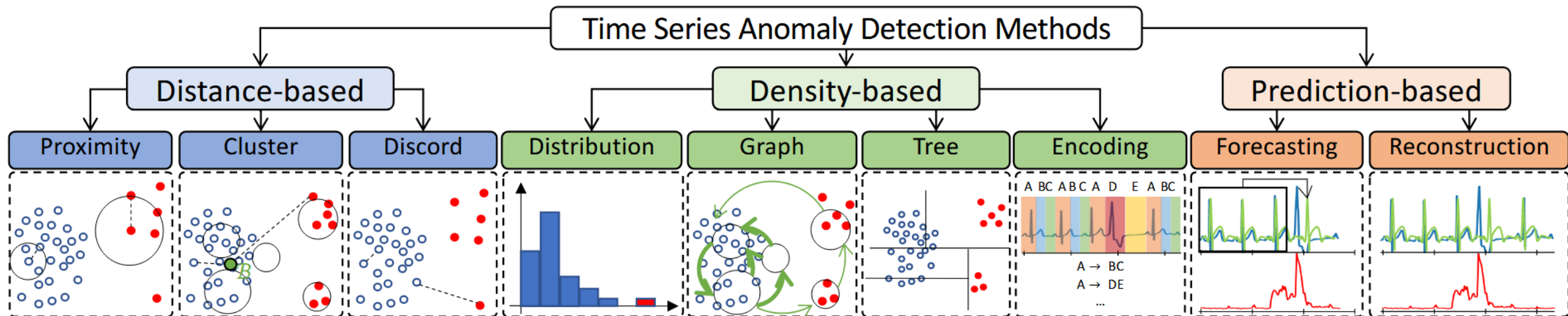


Introduction: Time Series Anomaly Detection (TSAD)

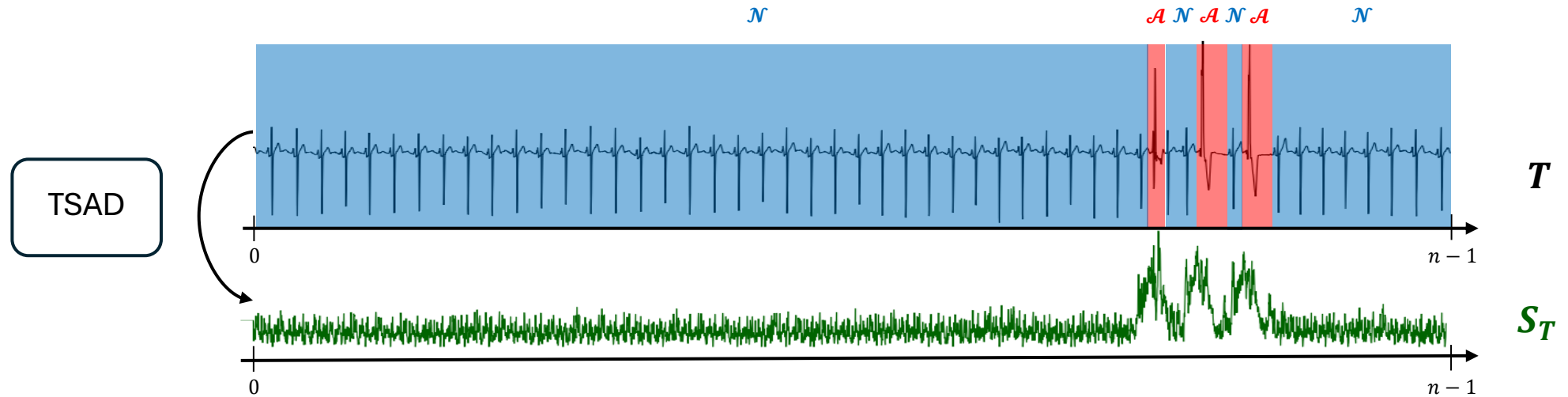
An **Anomaly Detection algorithm** takes a time series T and returns an **Anomaly Score S_T**



Introduction: Time Series Anomaly Detection (TSAD)



Introduction: Time Series Anomaly Detection (TSAD)

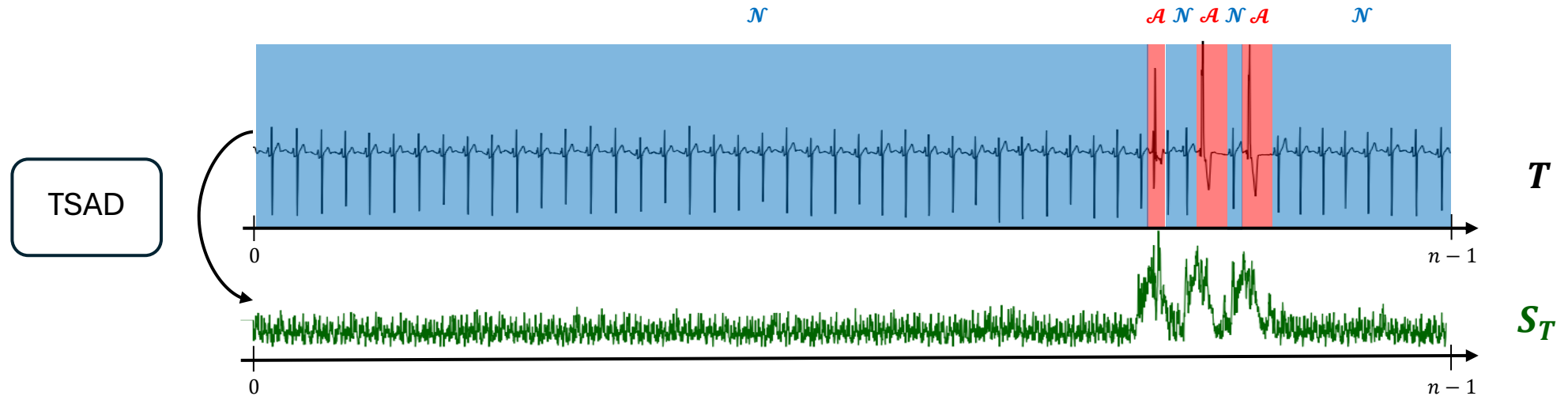


Supervised TSAD

Needs to train on both **Normal** and **Abnormal** data

Supervision Modes

Introduction: Time Series Anomaly Detection (TSAD)



Supervised TSAD

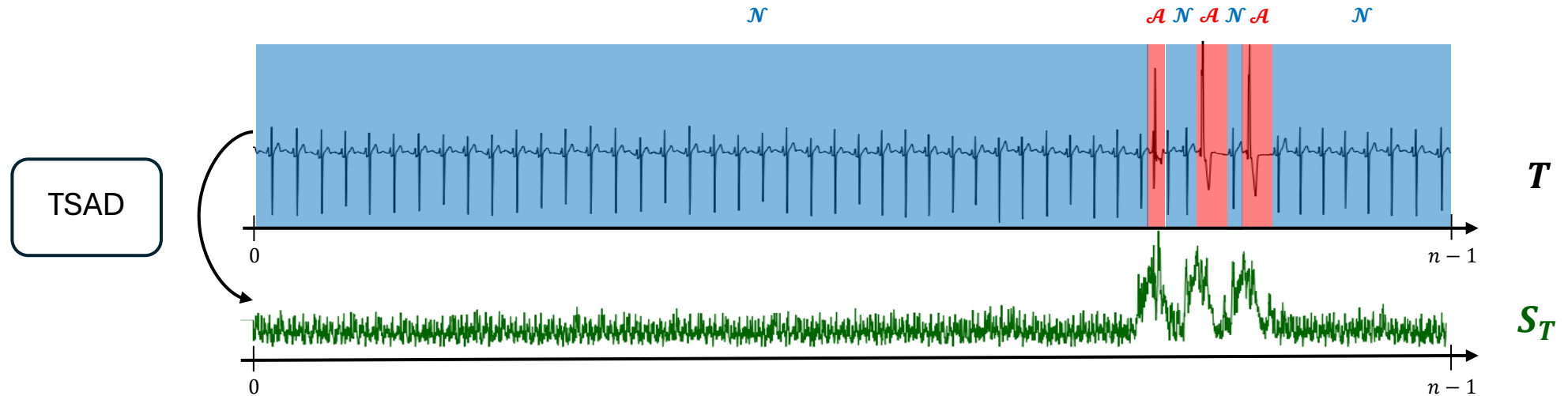
Needs to train on both **Normal** and **Abnormal** data

Semi Supervised TSAD

Needs to train on **Normal** data

Supervision Modes

Introduction: Time Series Anomaly Detection (TSAD)



Supervised TSAD

Needs to train on both **Normal** and **Abnormal** data

Semi Supervised TSAD

Needs to train on **Normal** data

Unsupervised TSAD

Doesn't need labeled data

Supervision Modes

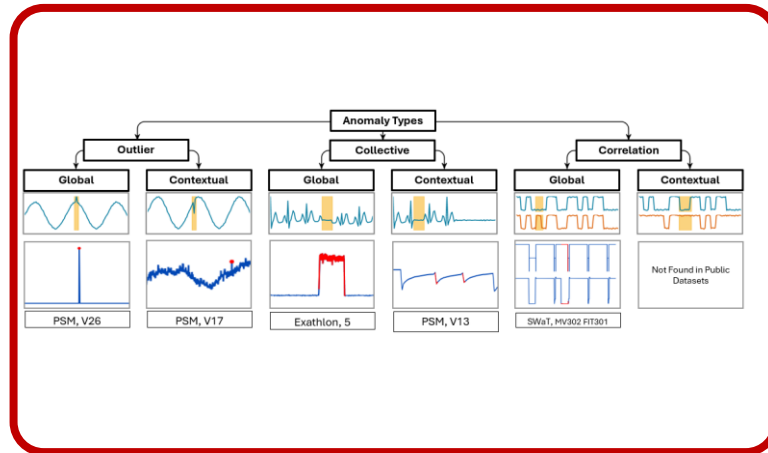
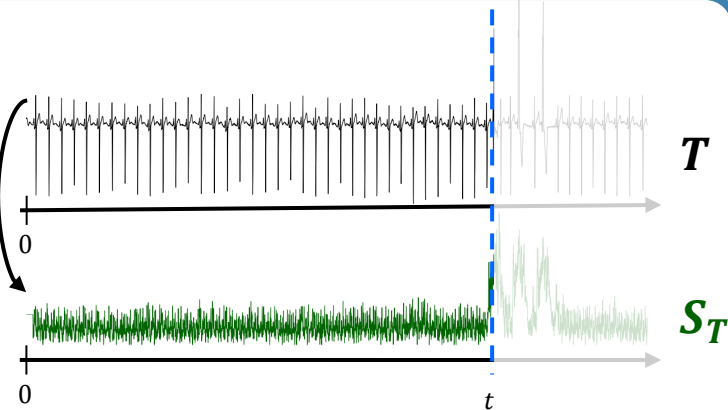
Introduction: Scope

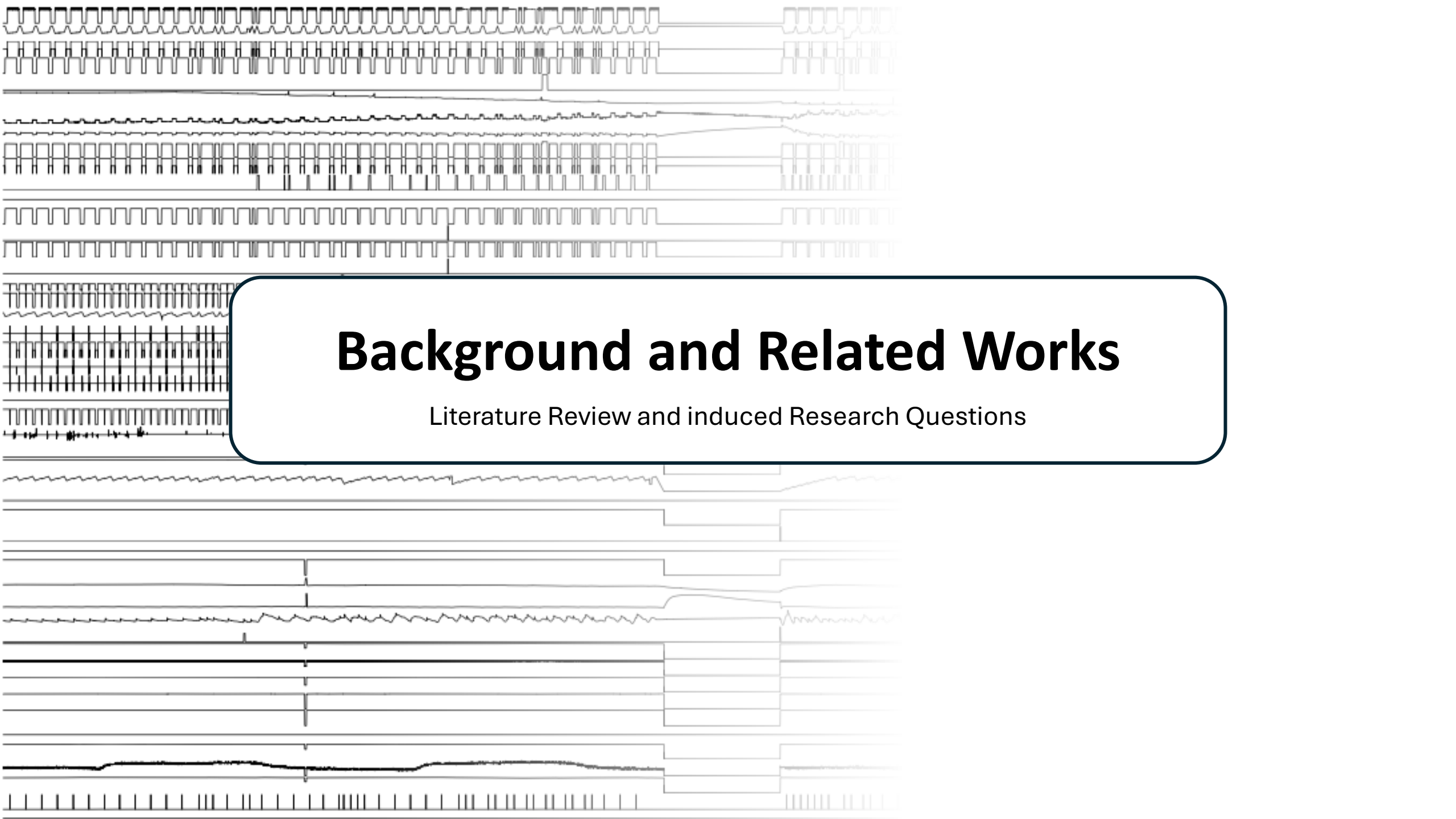
Object of the study:
Unsupervised **Streaming** Anomaly Detection on **real world data**

Unsupervised TSAD

Doesn't need labeled data

Streaming
AD
algorithm

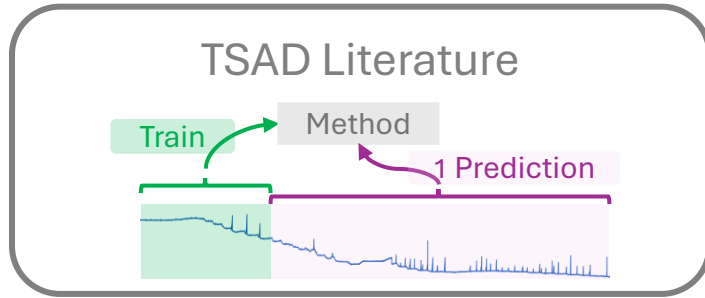


The background of the slide is a faded, grayscale image of multiple EEG (electroencephalogram) waveforms. These waveforms are arranged in horizontal rows, showing various patterns of electrical activity, including sharp spikes, rhythmic oscillations, and slower, more irregular waves. The waveforms are typical of what might be seen in a clinical or research setting, used to study brain function and activity.

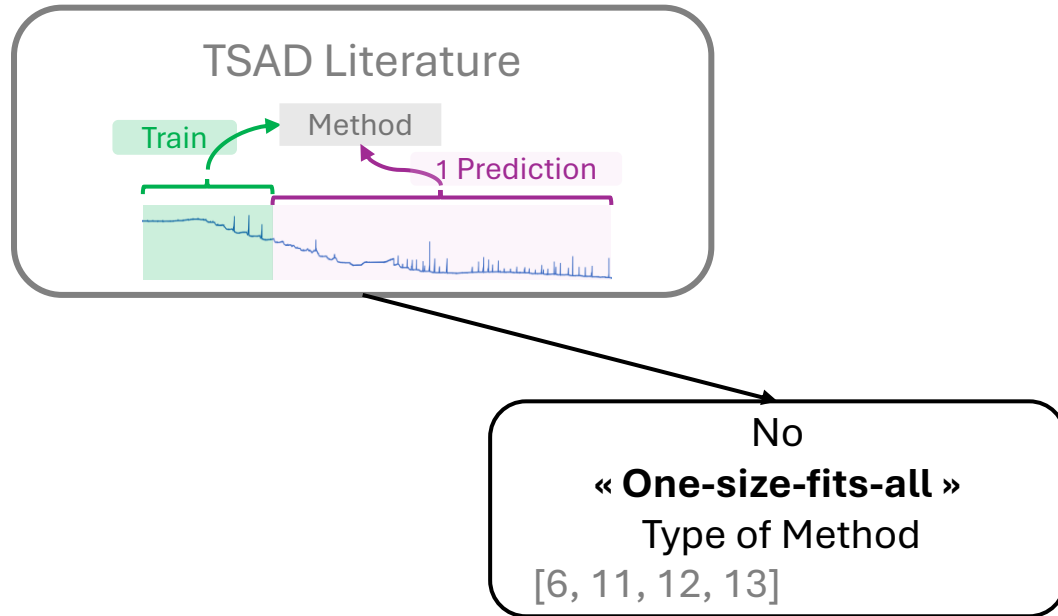
Background and Related Works

Literature Review and induced Research Questions

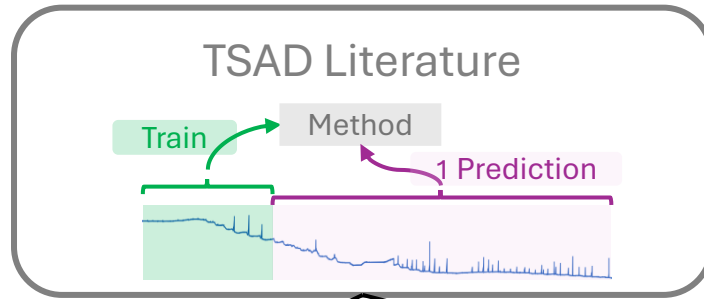
Literature Review: TSAD



Literature Review: TSAD



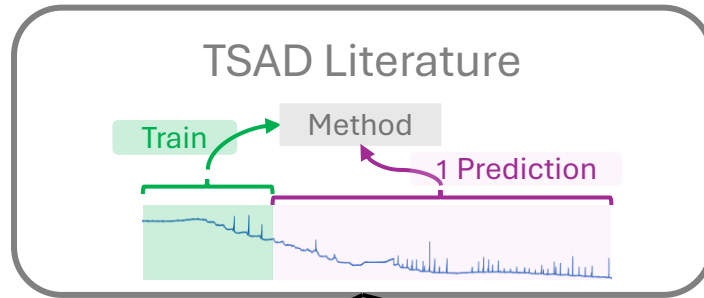
Literature Review: TSAD



Adapted for
Subsequences
[4, 17, 18]

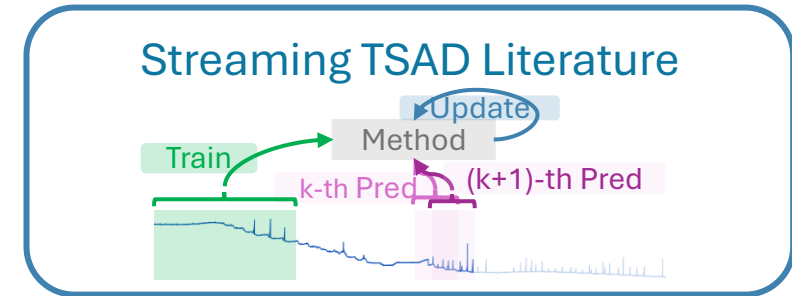
No
« **One-size-fits-all** »
Type of Method
[6, 11, 12, 13]

Literature Review: TSAD & Streaming TSAD

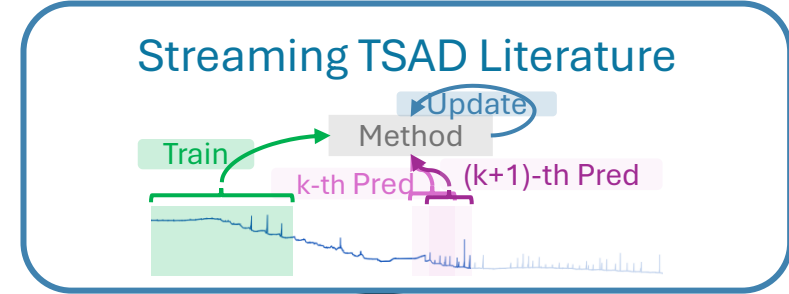
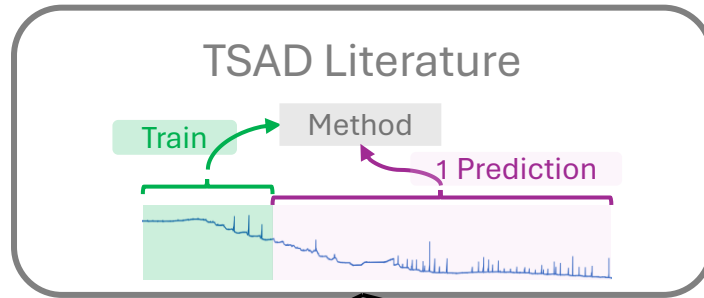


Adapted for
Subsequences
[4, 17, 18]

No
« **One-size-fits-all** »
Type of Method
[6, 11, 12, 13]



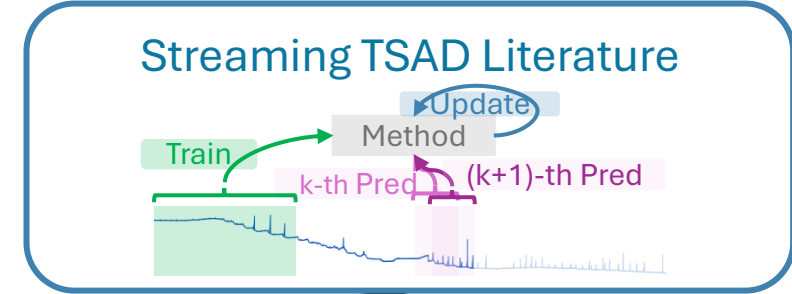
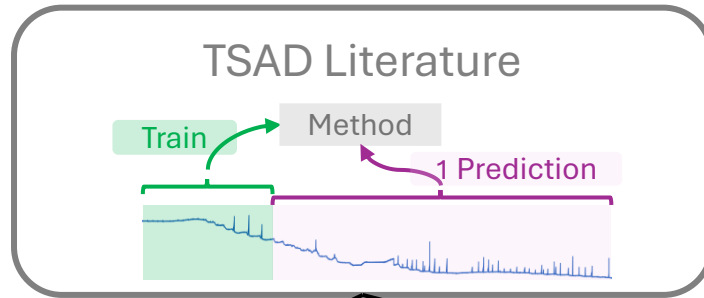
Literature Review: TSAD & Streaming TSAD



Adapted for
Subsequences
[4, 17, 18]

No
« **One-size-fits-all** »
Type of Method
[6, 11, 12, 13, 7, 14, 15, 16]

Literature Review: TSAD & Streaming TSAD

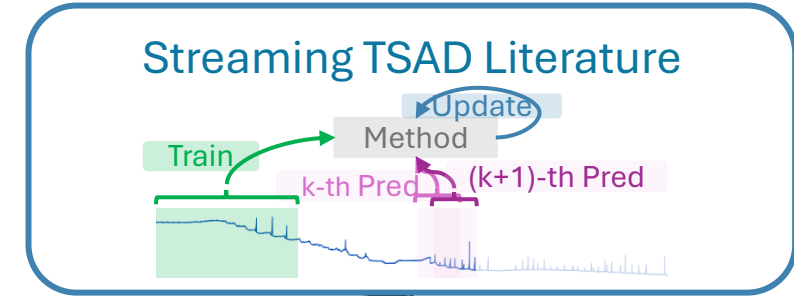
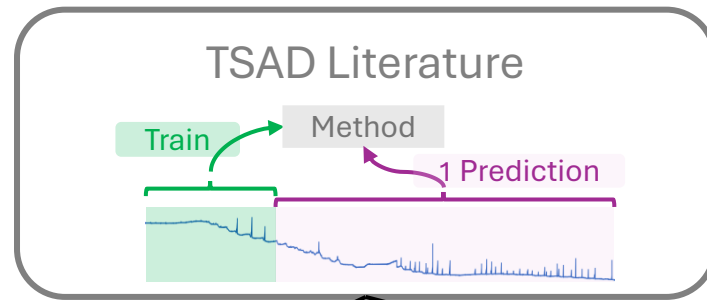


Adapted for
Subsequences
[4, 17, 18]

No
« **One-size-fits-all** »
Type of Method
[6, 11, 12, 13, 7, 14, 15, 16]

Focused on **Outlier
Detection**
[7, 14, 15, 16]

Literature Review: TSAD & Streaming TSAD



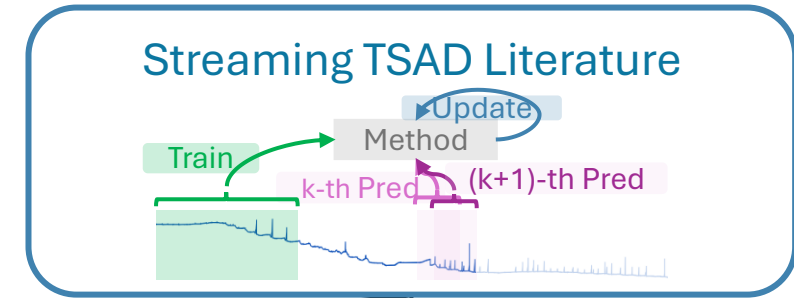
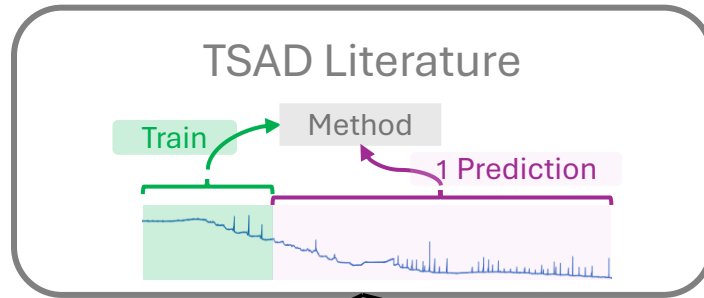
Adapted for
Subsequences
[4, 17, 18]

No
« **One-size-fits-all** »
Type of Method
[6, 11, 12, 13, 7, 14, 15, 16]

Focused on **Outlier
Detection**
[7, 14, 15, 16]

Streams May Induce
Concept Drift [23]

Literature Review: TSAD & Streaming TSAD



Adapted for
Subsequences
[4, 17, 18]

No
« **One-size-fits-all** »
Type of Method
[6, 11, 12, 13, 7, 14, 15, 16]

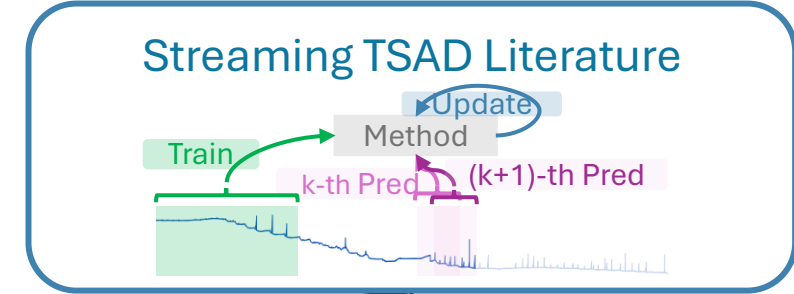
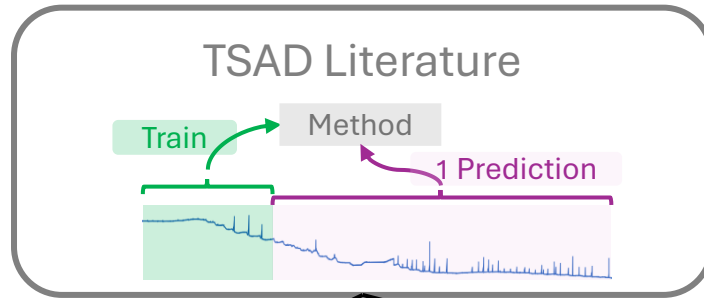
Focused on **Outlier
Detection**
[7, 14, 15, 16]

Streams May Induce
Concept Drift [23]

**Update
Mechanism**
[21, 22]

**Forgetting
Mechanism**
[21, 22]

Literature Review: TSAD & Streaming TSAD



Adapted for
Subsequences
[4, 17, 18]

No
« **One-size-fits-all** »
Type of Method
[6, 11, 12, 13, 7, 14, 15, 16]

Focused on **Outlier
Detection**
[7, 14, 15, 16]

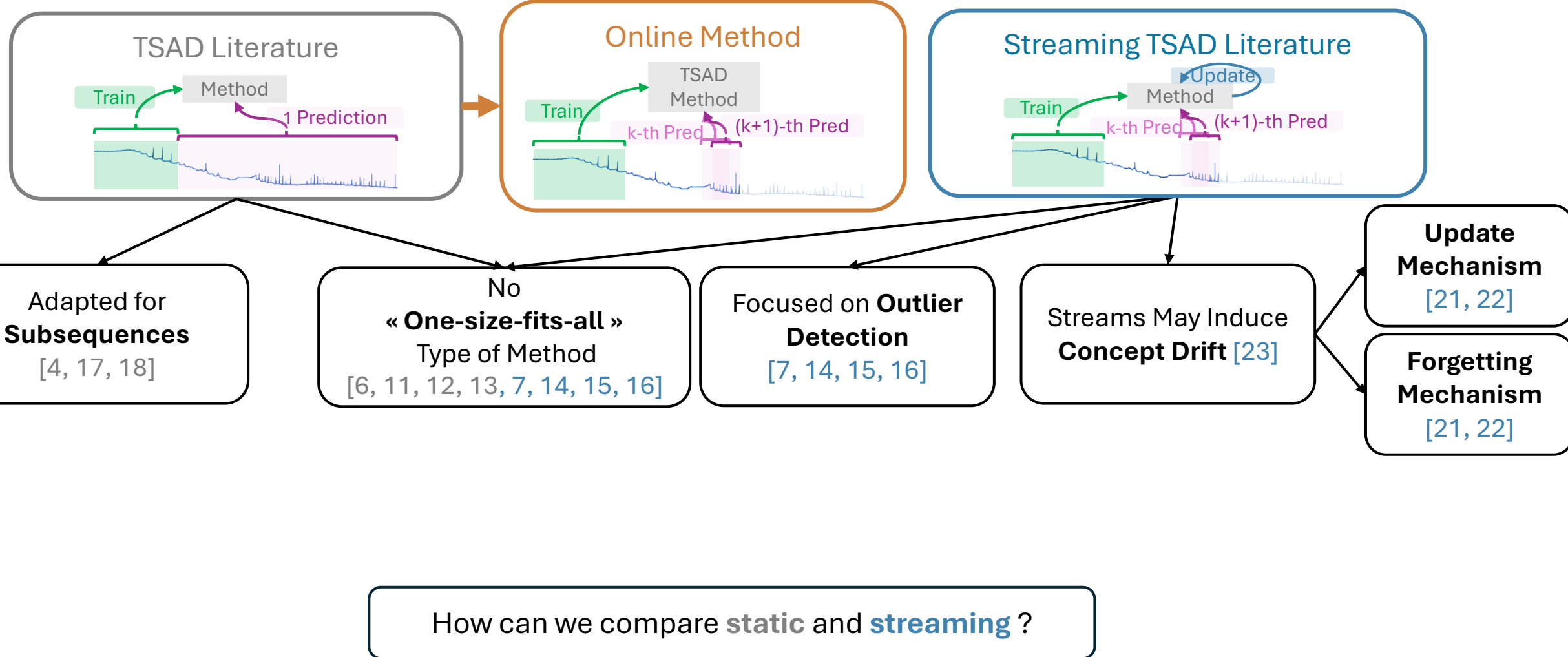
Streams May Induce
Concept Drift [23]

**Update
Mechanism**
[21, 22]

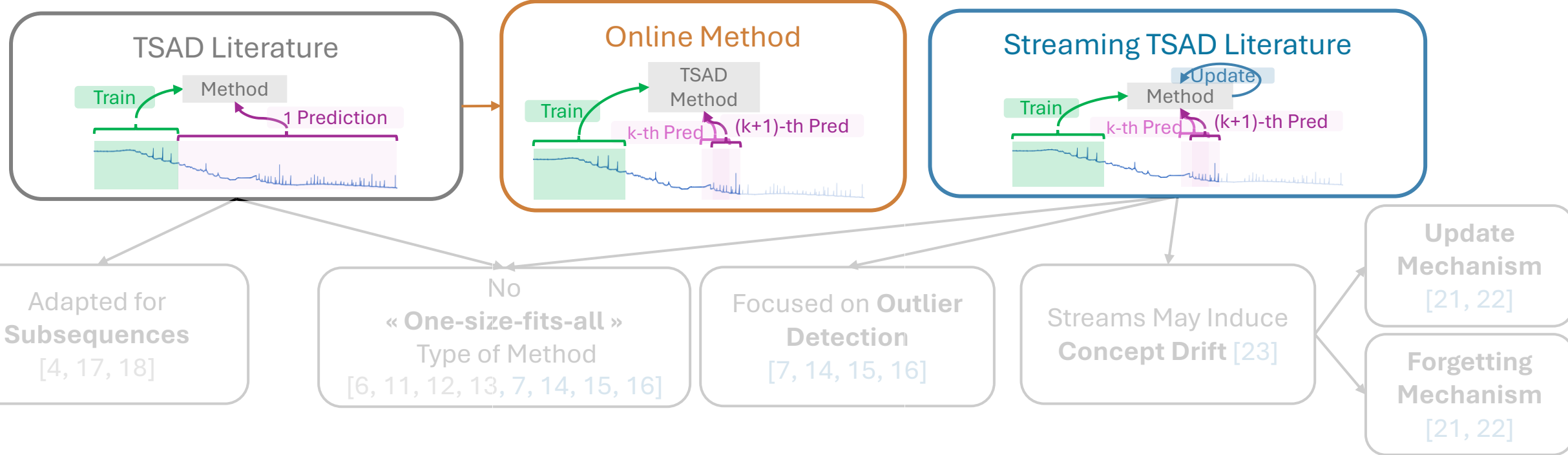
**Forgetting
Mechanism**
[21, 22]

How can we compare **static** and **streaming** ?

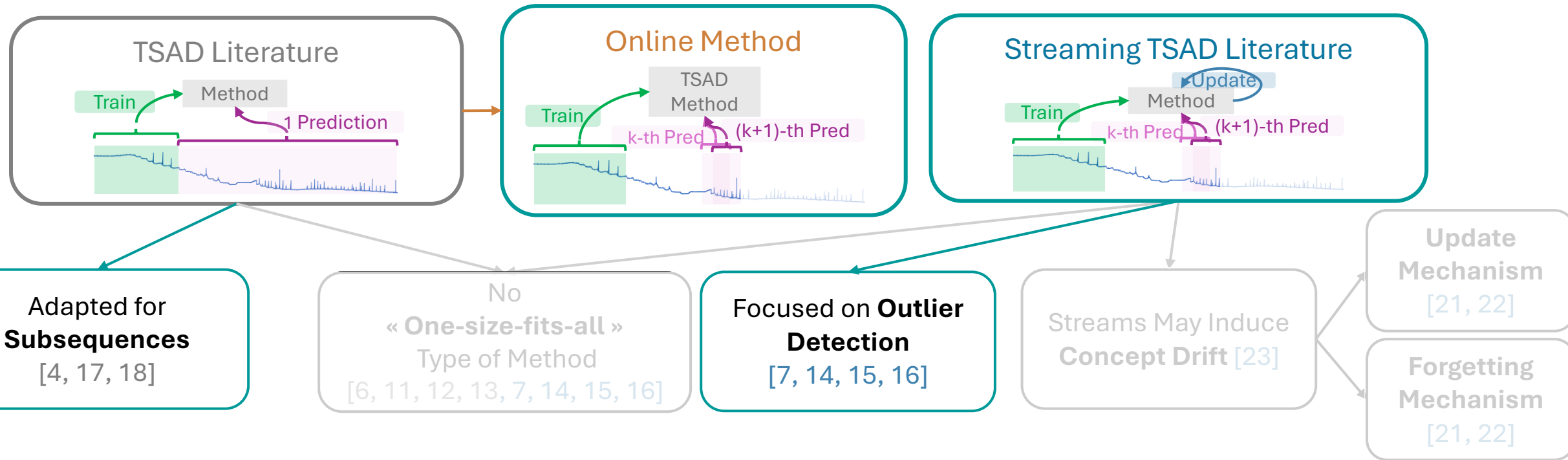
Literature Review: TSAD & Streaming TSAD



Literature Review: Main Research Questions

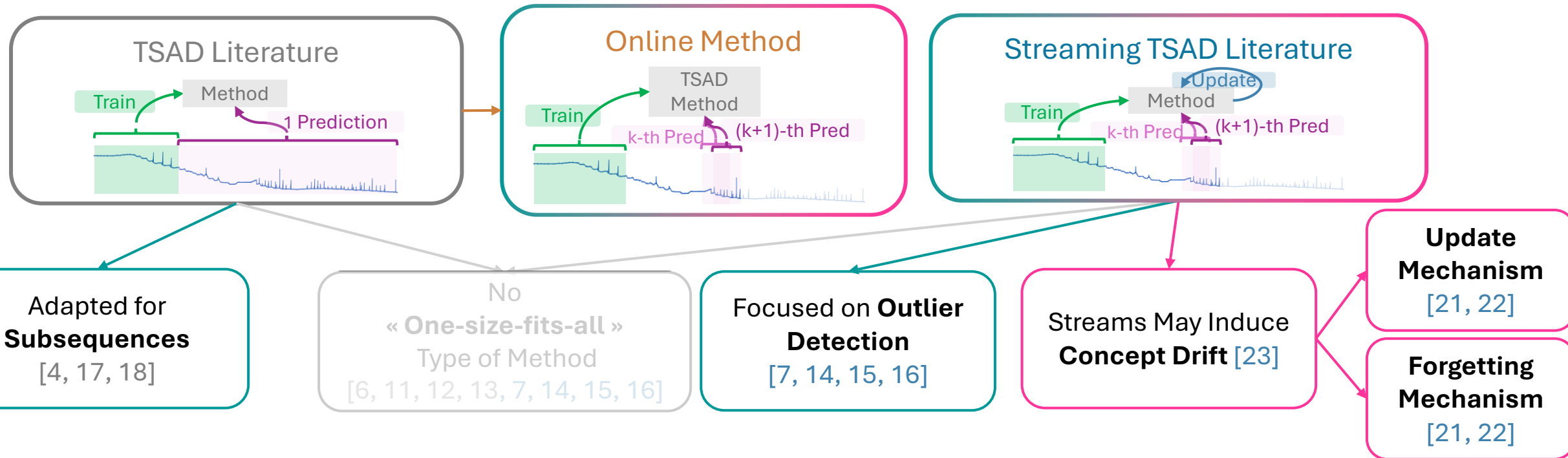


Literature Review: Main Research Questions



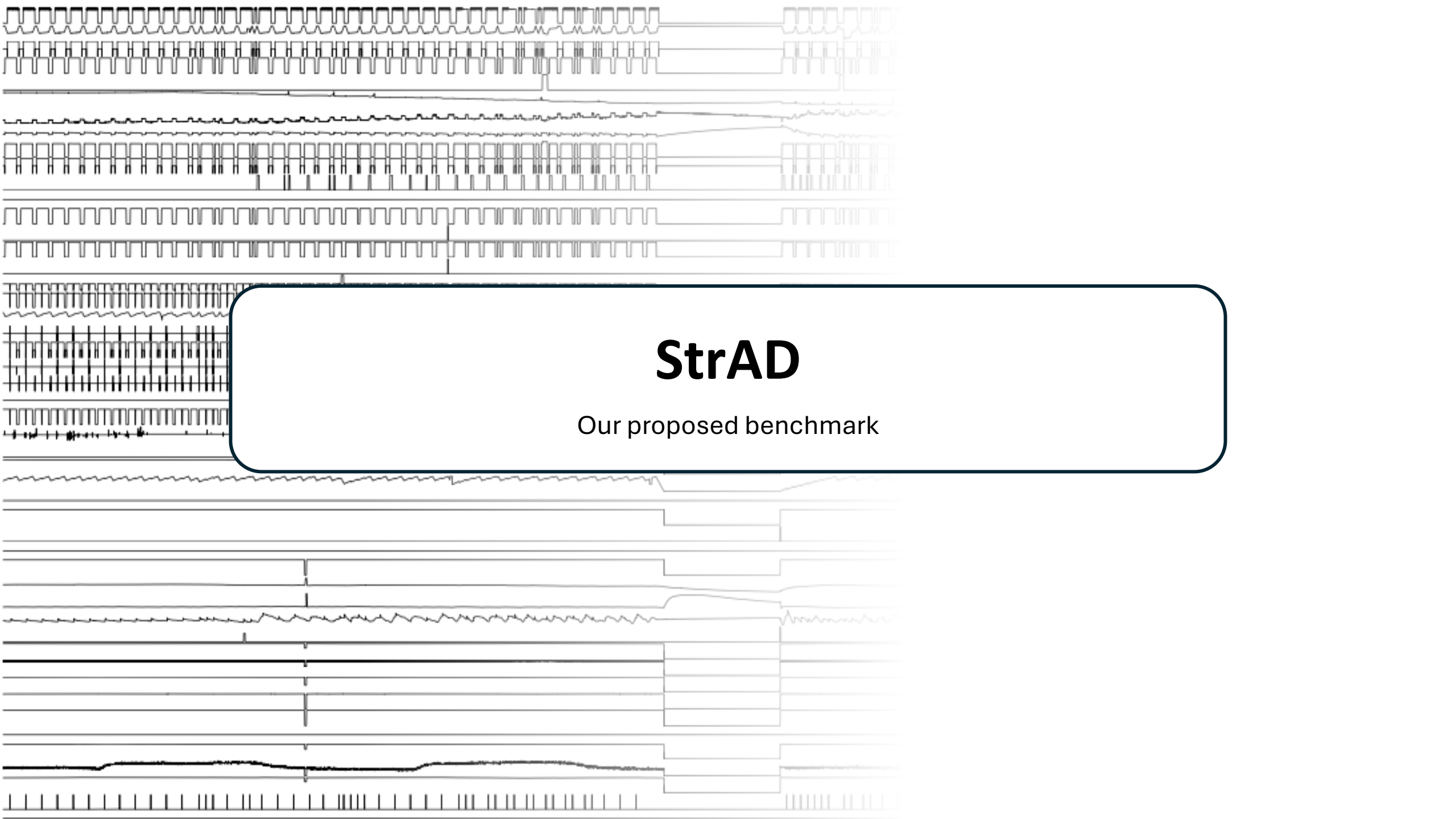
(Q1) Are **Streaming** approaches more **relevant** than **Online** approaches in **streaming settings**?

Literature Review: Main Research Questions



(Q1) Are **Streaming** approaches more **relevant** than **Online** approaches in **streaming settings**?

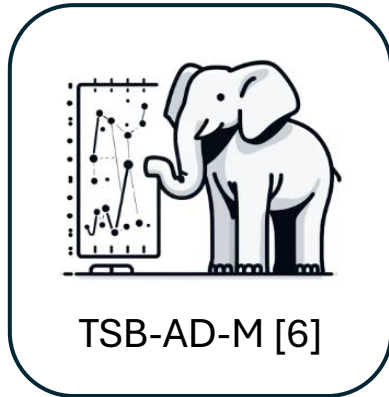
(Q2) Are **Streaming** approaches more appropriate than **Online** ones in case of **Concept Drift**?

The background of the slide is filled with various black and white waveform patterns, including square waves, sine waves, and complex digital signals, creating a technical, electronic aesthetic.

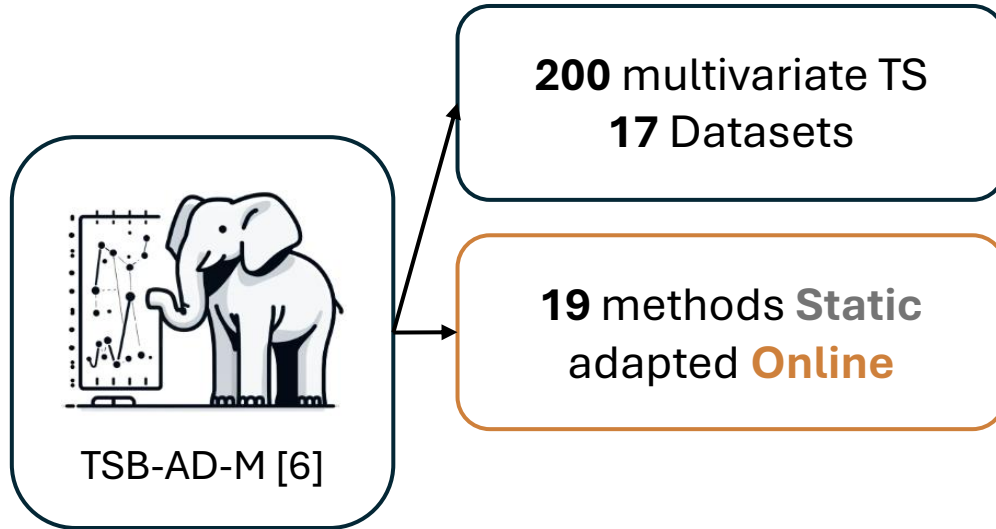
StrAD

Our proposed benchmark

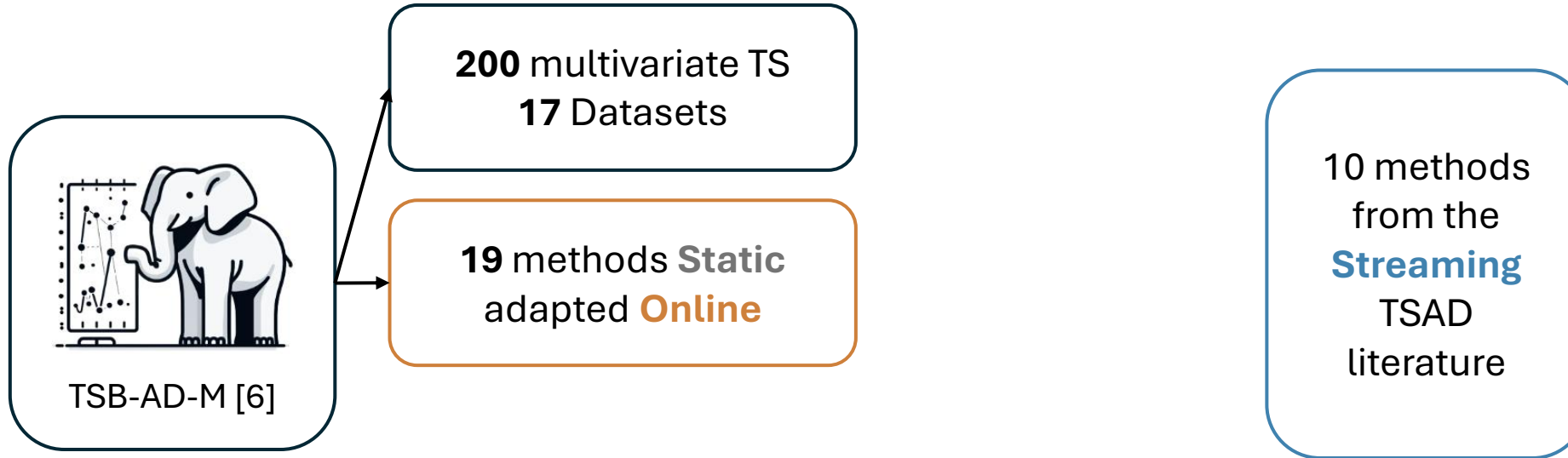
StrAD: Framework Description



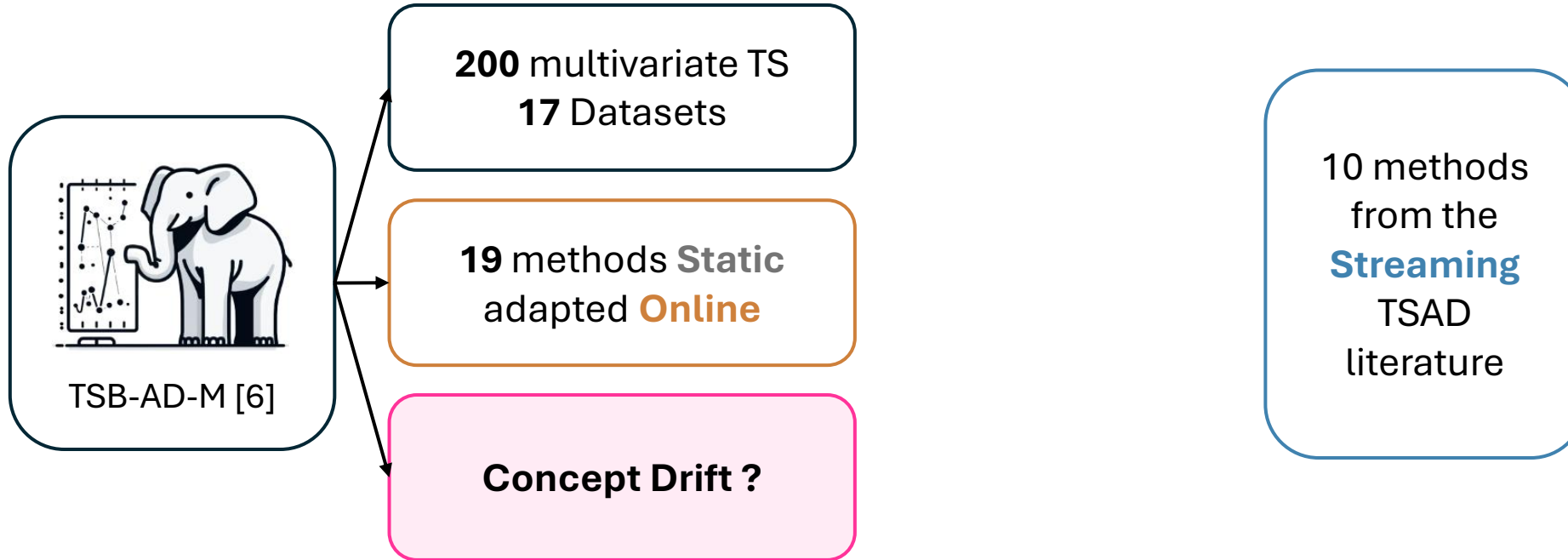
StrAD: Framework Description



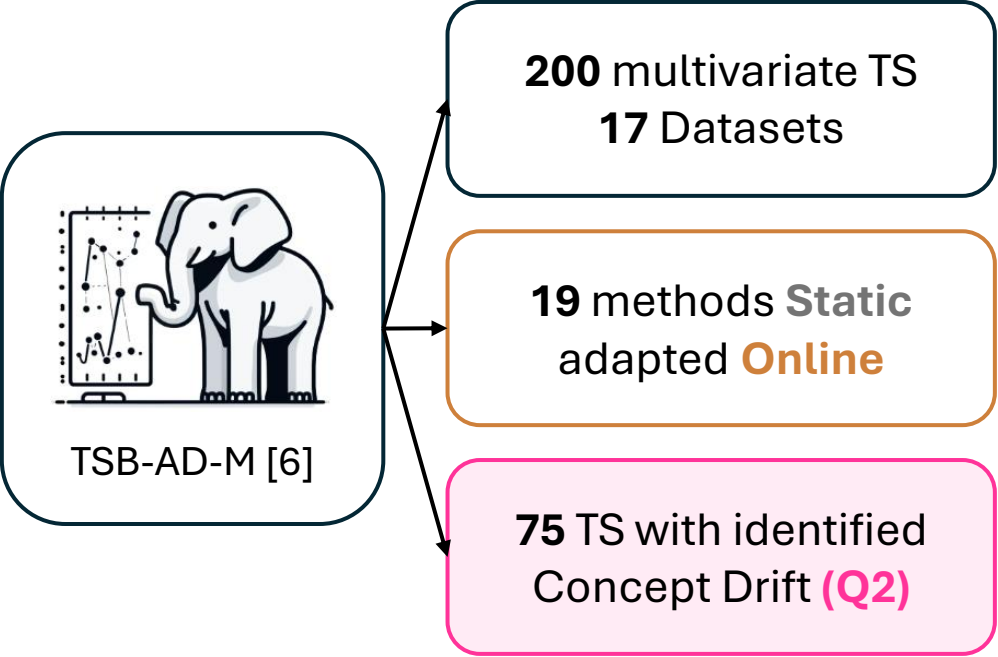
StrAD: Framework Description



StrAD: Framework Description



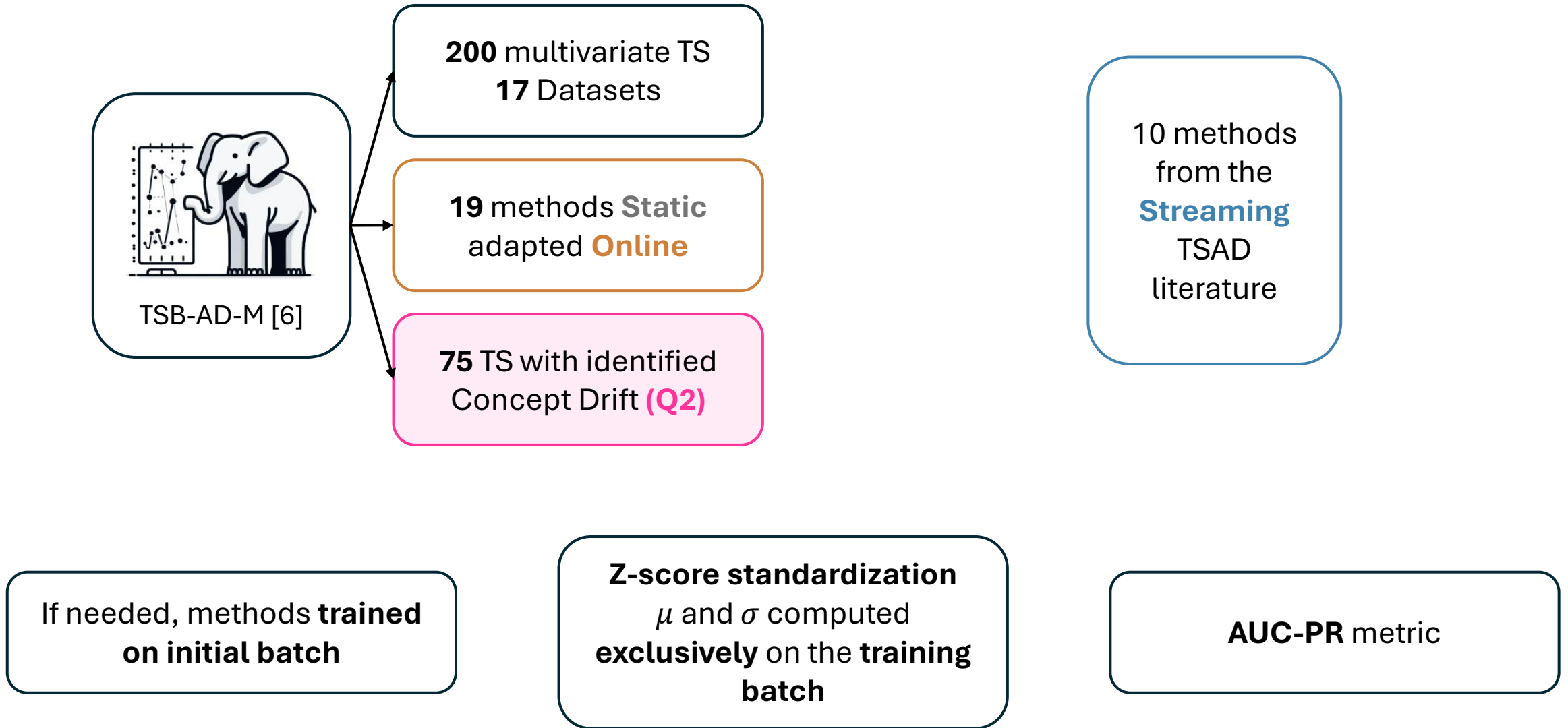
StrAD: Framework Description

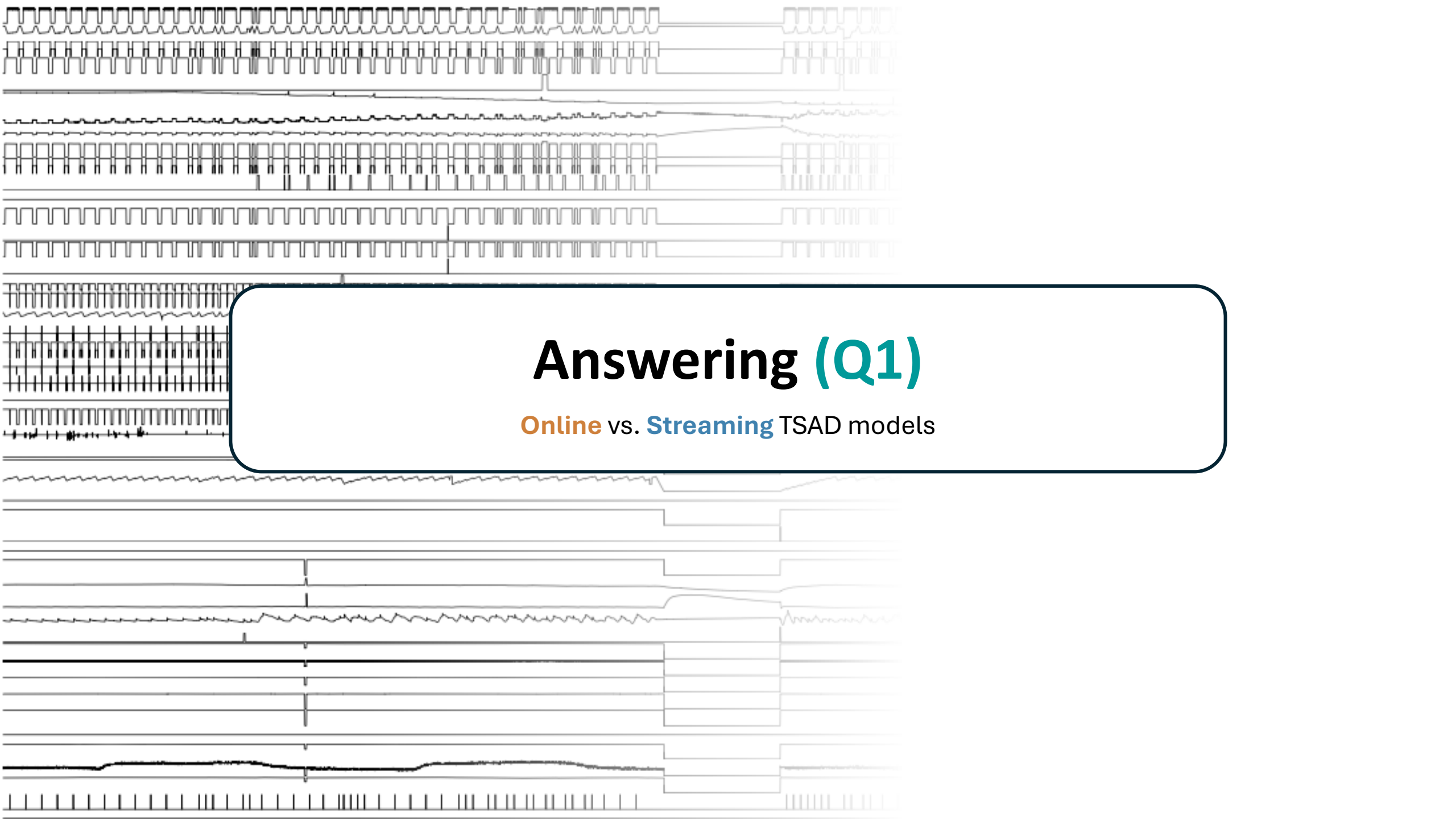


10 methods
from the
Streaming
TSAD
literature

Matrix Pattern	Time Series	Concept Drift	Occurrence Examples
		Continuous (C)	Sensor Degradation
		Change Point (CP)	Dynamic Allocation
		Periodic drift (P)	Seasonality
		Random Walk (RW)	Physiological Variability
		No drift	Medical (ECG)

StrAD: Framework Description





Answering (Q1)

Online vs. Streaming TSAD models

Answering (Q1): Online vs Streaming

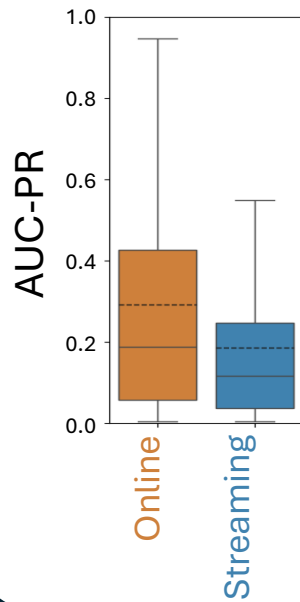
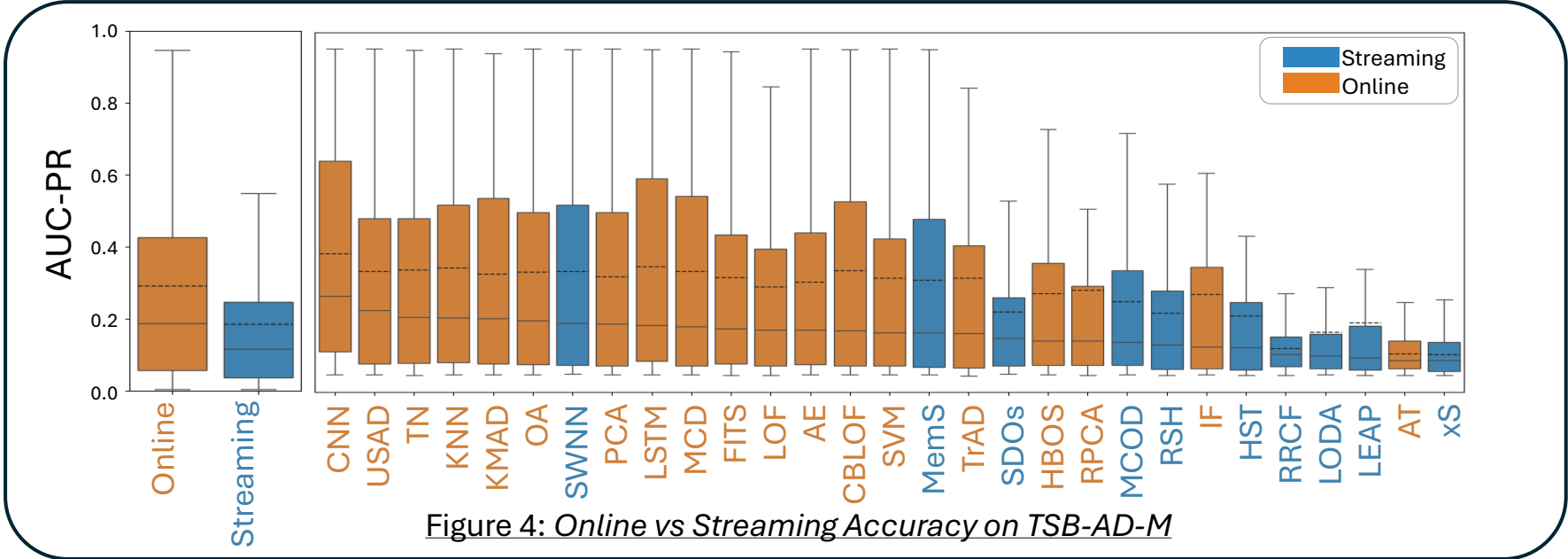
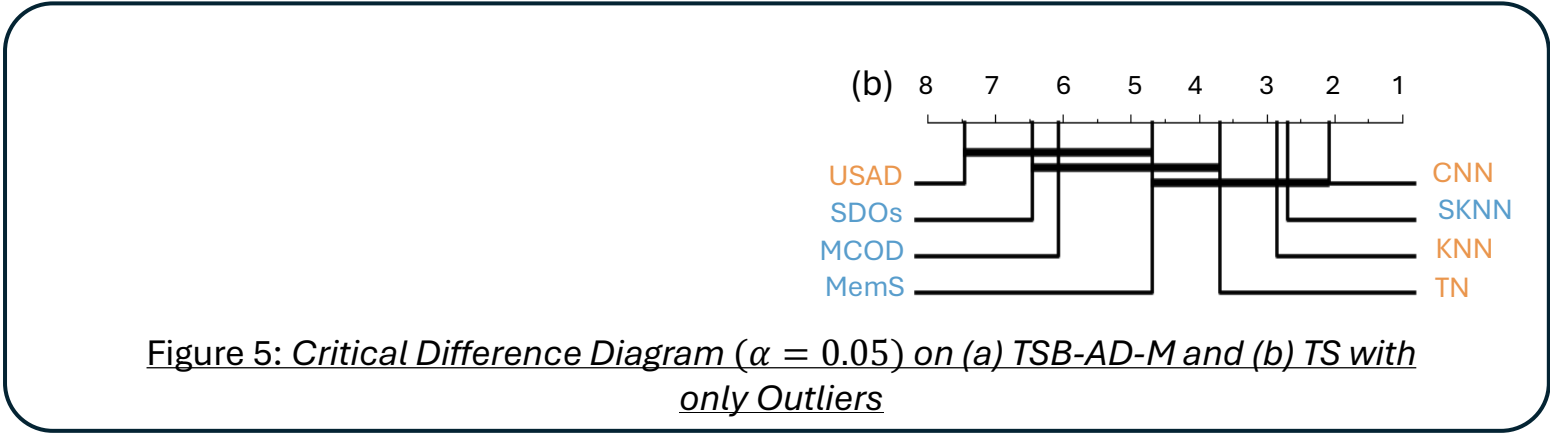
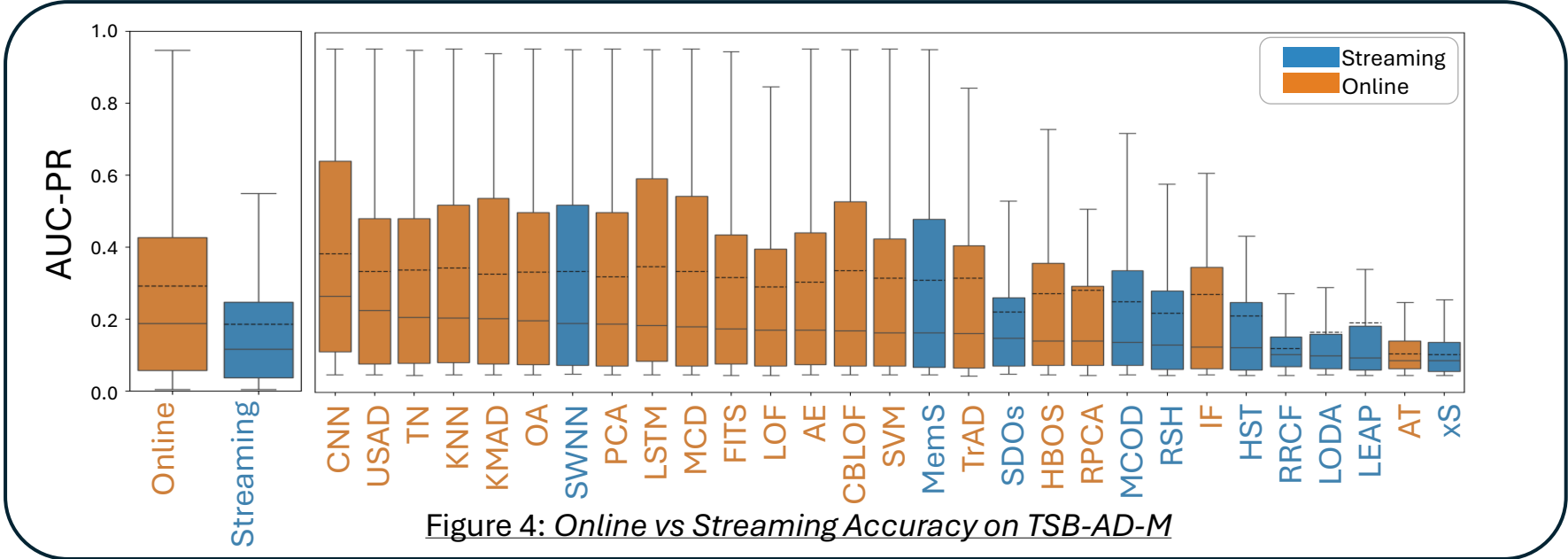


Figure 4: *Online vs Streaming Accuracy on TSB-AD-M*

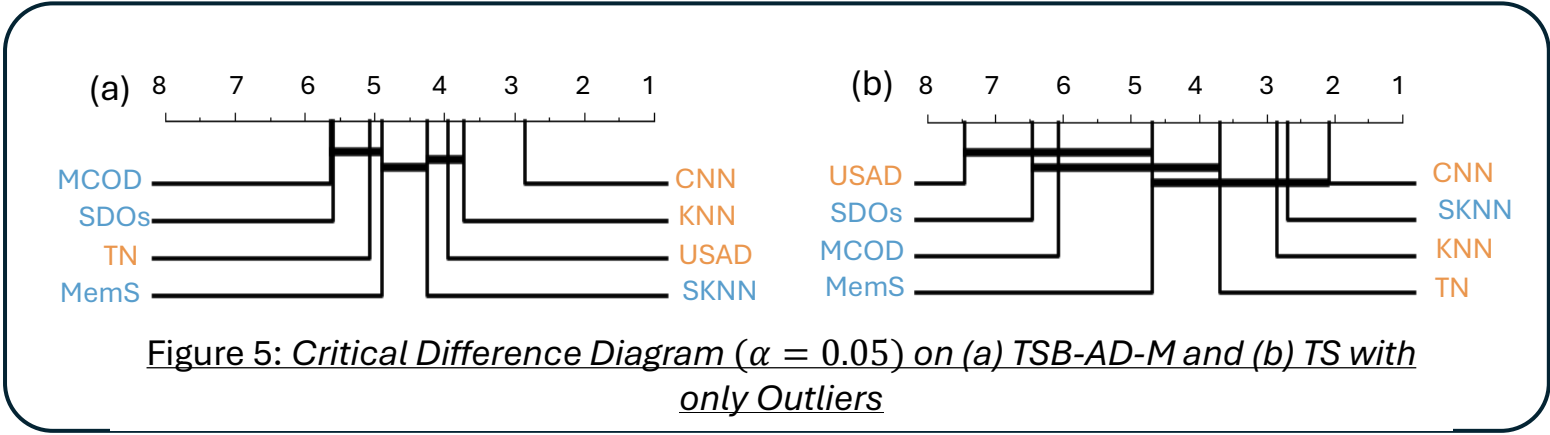
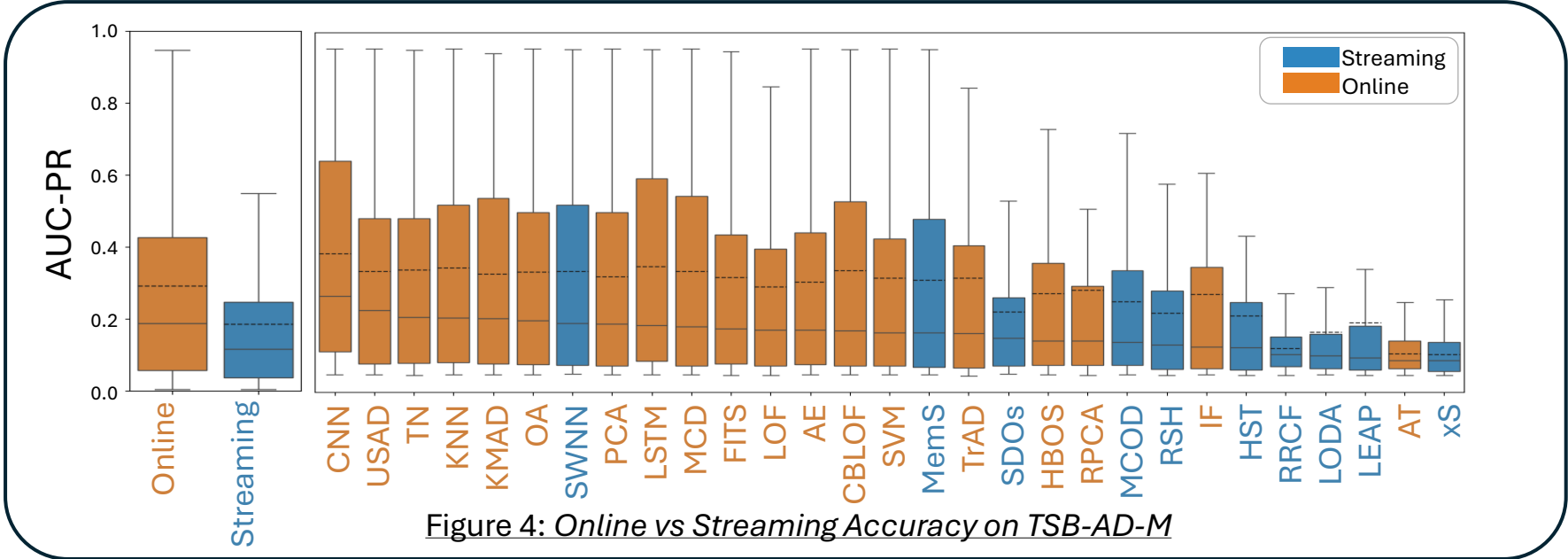
Answering (Q1): Online vs Streaming



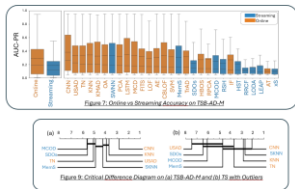
Answering (Q1): Online vs Streaming



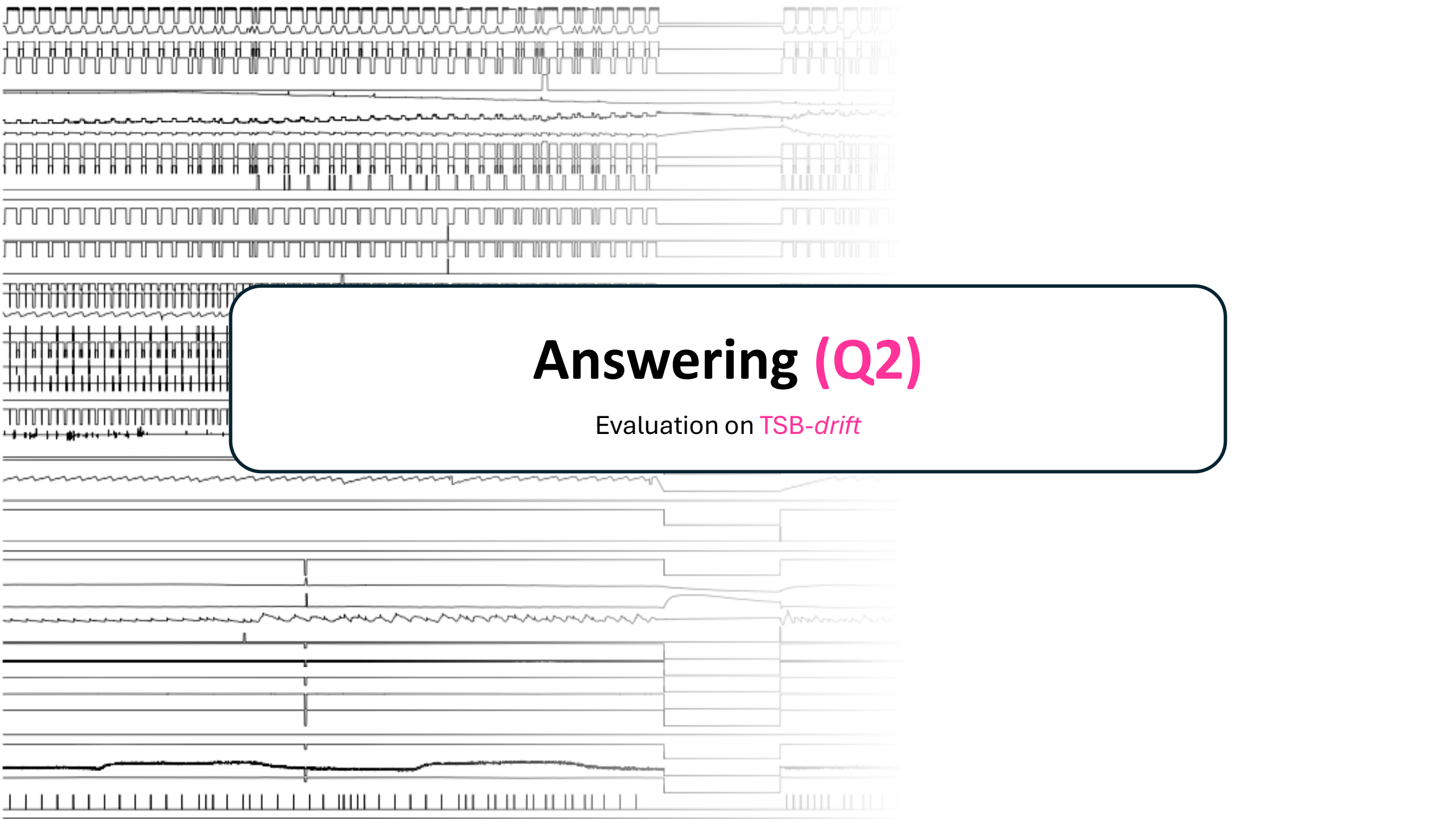
Answering (Q1): Online vs Streaming



Answering (Q1): **Online** vs **Streaming**



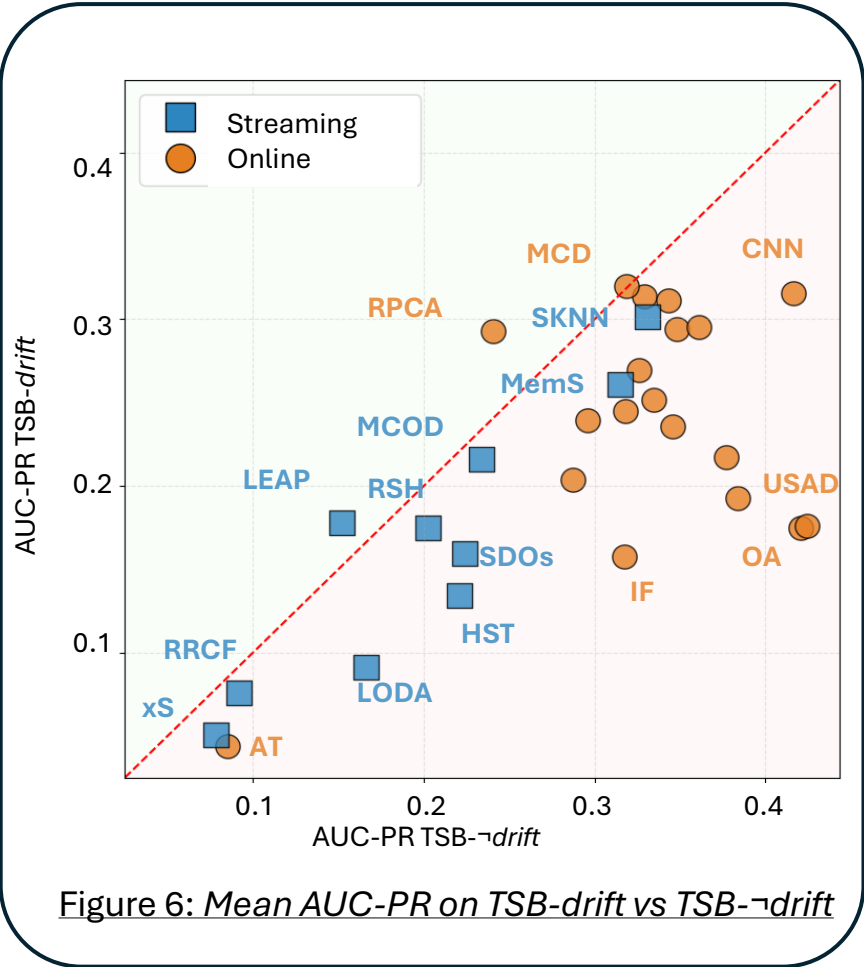
(Q1) **Streaming** approaches are **less accurate** than **Online** models on **real-world data**. **Streaming** methods suffer from a **focus on point outliers** and struggle to detect temporal nuances.



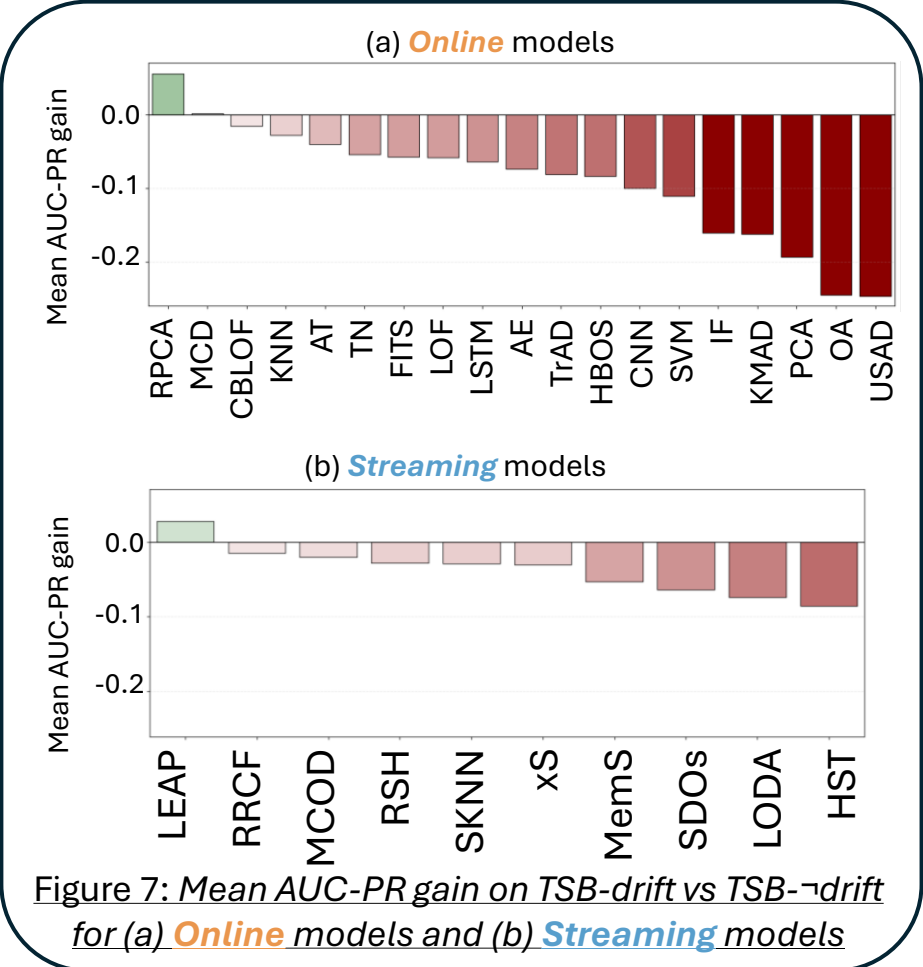
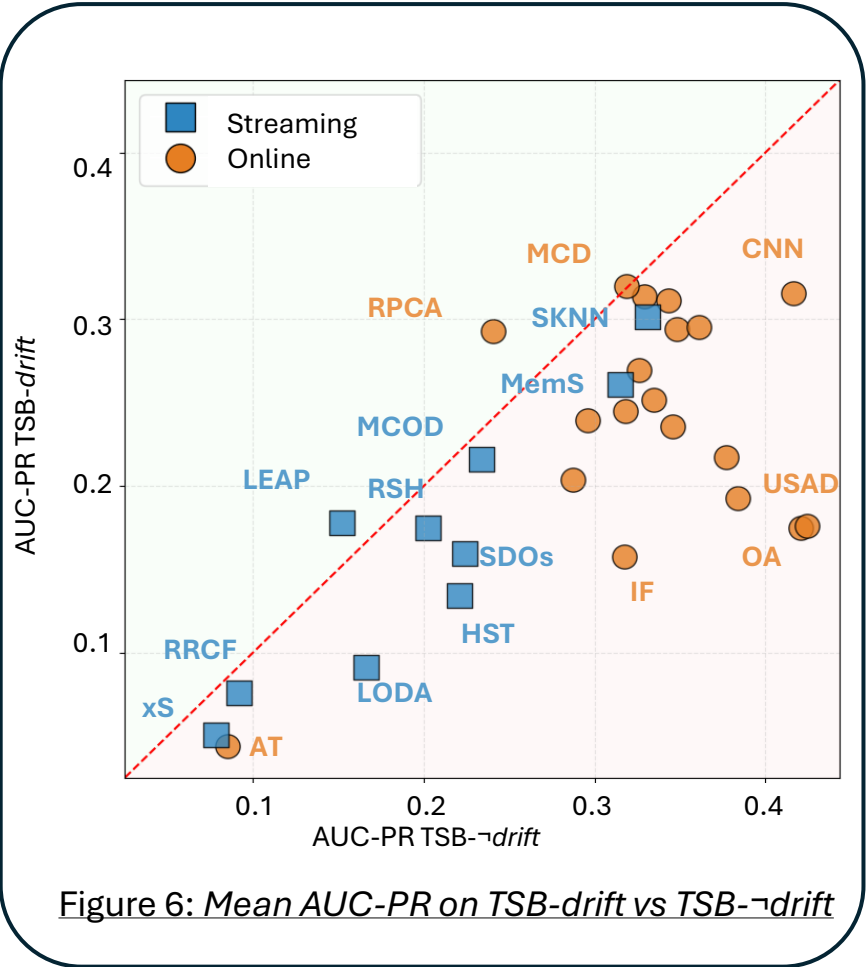
Answering (Q2)

Evaluation on *TSB-drift*

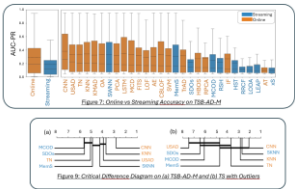
Answering (Q2): Evaluation on TSB-drift



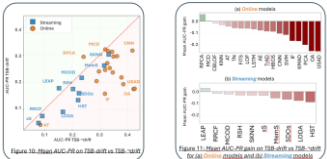
Answering (Q2): Evaluation on TSB-drift



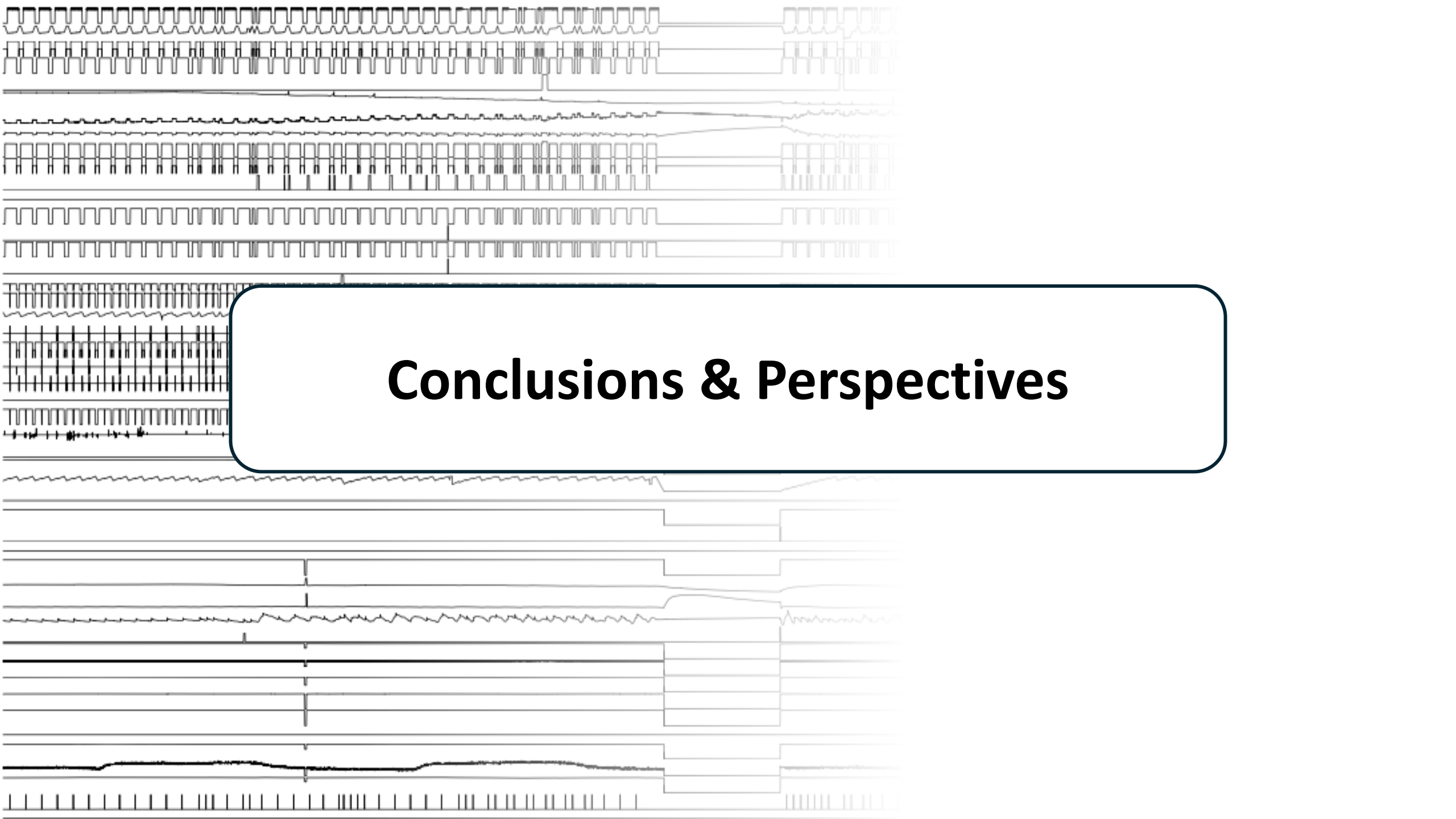
Answering (Q2): Evaluation on *TSB-drift*



(Q1) **Streaming** approaches are **less accurate** than **Online** models on **real-world data**. **Streaming** methods suffer from a **focus on point outliers** and struggle to detect temporal nuances.

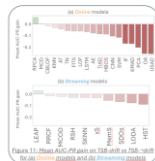
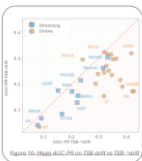
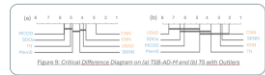
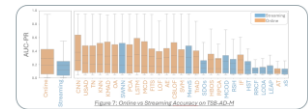


(Q2) While **Streaming** approaches demonstrate **superior resilience to concept drifts**, they are **not necessarily the more appropriate** in terms of **absolute accuracy**.



Conclusions & Perspectives

Conclusions of the paper & Perspectives



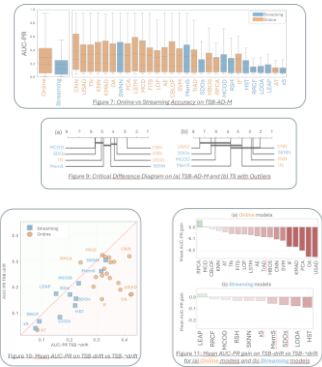
Limitations of Current Streaming TSAD:

the emphasis on **point-wise outliers**. In practice, **anomalies are often collective patterns** rather than isolated points. **Streaming** methods based on distance or density **fail to capture such patterns**.

Necessary shift in Streaming TSAD:

there is a need to develop **sequence-aware Streaming** methods

Conclusions of the paper & Perspectives

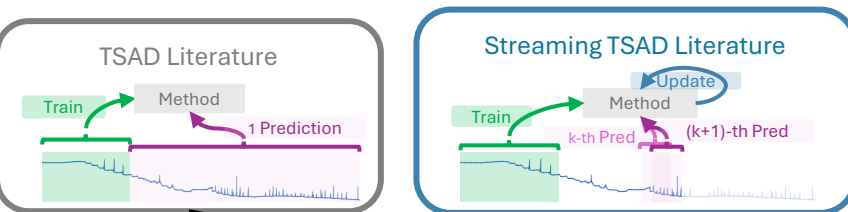


Limitations of Current Streaming TSAD:

the emphasis on **point-wise outliers**. In practice, **anomalies are often collective patterns** rather than isolated points. **Streaming** methods based on distance or density **fail to capture such patterns**.

Necessary shift in Streaming TSAD:

there is a need to develop **sequence-aware Streaming** methods



No

« One-size-fits-all »

Type of Method

[6, 11, 12, 13, 7, 14, 15, 16]

Streaming AutoAD as a possible Solution:

AutoAD **effective** in **Static** settings but largely **unexplored** in **Streaming**

Thank you for your attention!
Any questions?

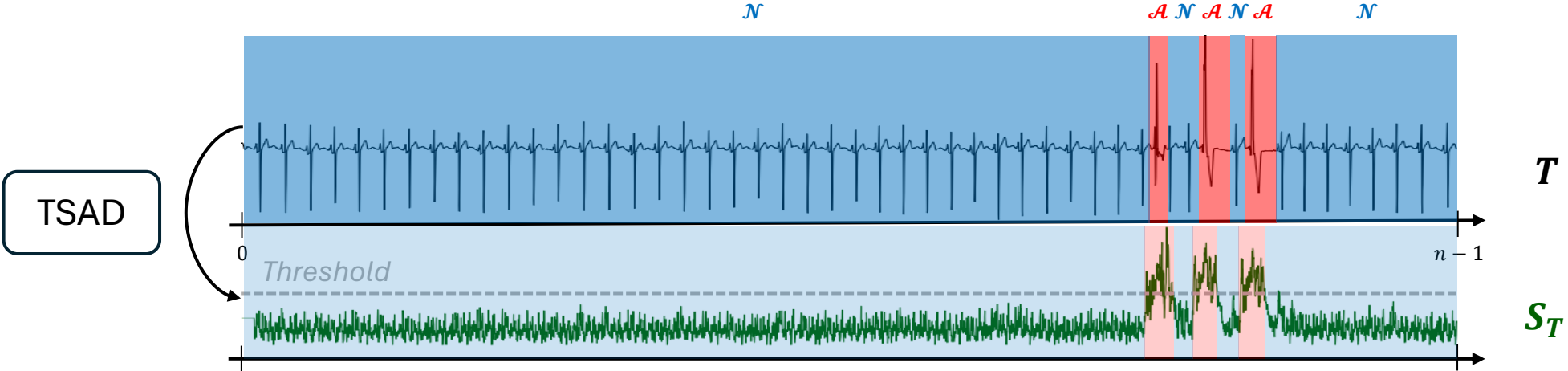


Paper

References

- [1] Paul Boniol, Anomaly Detection in Time Series, keynote for AALTD workshop at ECML / PKDD 2024 (September 2024).
- [2] A. L. Goldberger, et al. Physiobank, physiotookit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.
- [3] Wenig, P. et al. (2024). Anomaly Detectors for Multivariate Time Series: The Proof of the Pudding is in the Eating. *ICDEW*, 96-101.
- [4] Paparrizos, J. et al. (2025). Advances in Time-Series Anomaly Detection: Algorithms, Benchmarks, and Evaluation Measures, *SIGKDD*, 6151–6161.
- [5] Paul Boniol, et al. VUS: effective and efficient accuracy measures for time-series anomaly detection. *VLDB J.* 34, 3 (2025), 32. doi:10.1007/S00778-025-00907-X
- [6] Liu, L. and Paparrizos, J. (2024). The Elephant in the Room: Towards A Reliable Time-Series Anomaly Detection Benchmark, *NeurIPS*.
- [7] Cao, Y. et al. (2024). Revisiting streaming anomaly detection: benchmark and evaluation. *Artif. Intell. Rev.*, 58(1).
- [8] Parrino, M, et al. In a StreamingWorld, Should You Stand Still? A Comprehensive Benchmark of Anomaly Detection in Streams. 2026, *KDD*
- [9] Qinghua Liuet al. TSB-AutoAD: Towards Automated Solutions for Time-Series Anomaly Detection [E, A & B]. *Proc. VLDB (2025)* doi:10.14778/3749646.3749699
- [10] Emmanouil Sylligardos, et al. MSAD: A deep dive into model selection for time series anomaly detection. *VLDB J.* (2025), 72. doi:10.1007/S00778-025-00949-1
- [11] Sebastian Schmidl et al. Anomaly Detection in Time Series: A Comprehensive Evaluation. *Proc. VLDB Endow.* (2022) 10.14778/3538598.3538602
- [12] Dennis Wagner et al. 2023. TimeSeAD: Benchmarking Deep Multivariate Time-Series Anomaly Detection. *Trans. Mach. Learn. Res.* 2023 (2023).
<https://openreview.net/forum?id=iMmsCI0JsS>
- [13] Julien Audibert, et al. 2022. Do deep neural networks contribute to multivariate time series anomaly detection? *Pattern Recognit.* 132 (2022), 108945. doi:10.1016/J.PATCOG.2022.108945
- [14] Antonios Ntroumpogiannis, et al. 2023. A meta-level analysis of online anomaly detectors. *VLDB J.* 32, 4 (2023), 845–886. doi:10.1007/S00778-022-00773-X
- [15] Rebecca Salles, et al. 2025. Scalable and accurate online multivariate anomaly detection. *Inf. Syst.* 131 (2025), 102524. doi:10.1016/J.IS.2025.102524
- [16] Félix Iglesias Vázquez, et al. 2023. Anomaly detection in streaming data: A comparison and evaluation study. *Expert Syst. Appl.* 233 (2023), 120994. doi:10.1016/J.ESWA.2023.120994
- [17] Eamonn J. Keogh, et al. 2005. HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence. In *Proceedings of the 5th IEEE International Conference on Data Mining (ICDM) 2005*, 27-30 November 2005, 226–233. doi:10.1109/ICDM.2005.79
- [18] Mohsin Munir, et al. 2019. DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series. *IEEE Access* 7 (2019), 1991–2005. doi:10.1109/ACCESS.2018.2886457
- [19] D. M. Hawkins. 1980. Identification of Outliers. Springer. doi:10.1007/978-94-015-3994-4
- [20] Markus Goldstein, et al. 2012. Histogram-based Outlier Score (HBOS): A fast Unsupervised Anomaly Detection Algorithm. In *KI*.
- [21] Sudipto Guha, et al. 2016. Robust Random Cut Forest Based Anomaly Detection on Streams. In *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016*. JMLR.org, 2712–2721. <http://proceedings.mlr.press/v48/guha16.html>
- [22] Saket Sathe, et al. 2018. Subspace histograms for outlier detection in linear time. *Knowl. Inf. Syst.* 56, 3 (2018), 691–715. doi:10.1007/S10115-017-1148-8
- [23] João Gama, et al. 2014. A survey on concept drift adaptation. *ACM Comput. Surv.* 46, 4 (2014), 44:1–44:37. doi:10.1145/2523813

Appendix: Time Series Anomaly Detection (TSAD) evaluation



F1-score

		Ground Truth	
		$\mathcal{A}$	$\mathcal{N}$
Model Prediction	$\mathcal{A}$	TP	FP
	$\mathcal{N}$	FN	TN

$$\text{Precision} = \frac{TP}{TP + FP}$$

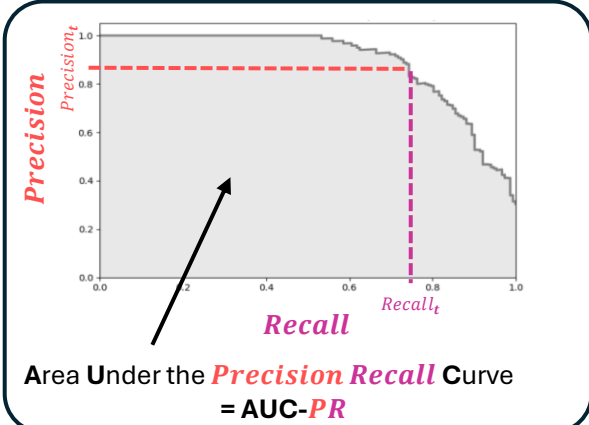
$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

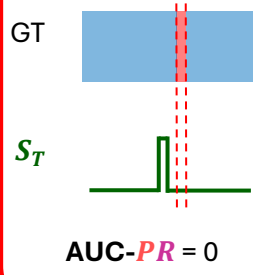
Threshold dependent !

AUC-PR

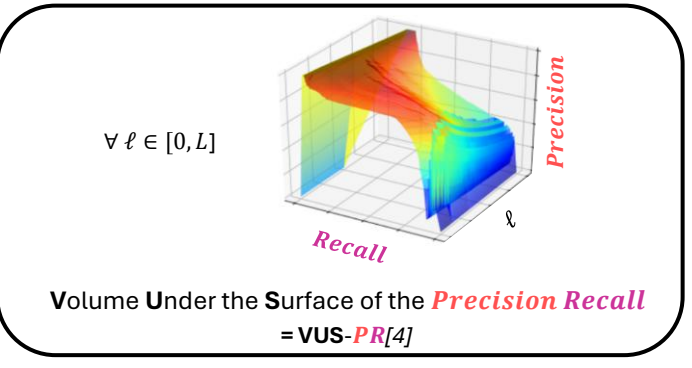
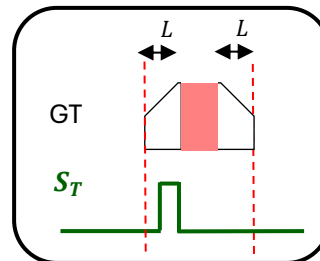
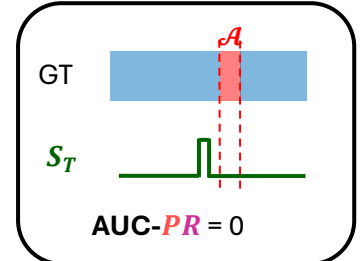
$$\text{Precision}_t = \frac{TP_t}{TP_t + FP_t}$$
$$\text{Recall}_t = \frac{TP_t}{TP_t + FN_t}$$
$$\forall t \in [S_{tmin}, S_{tmax}]$$



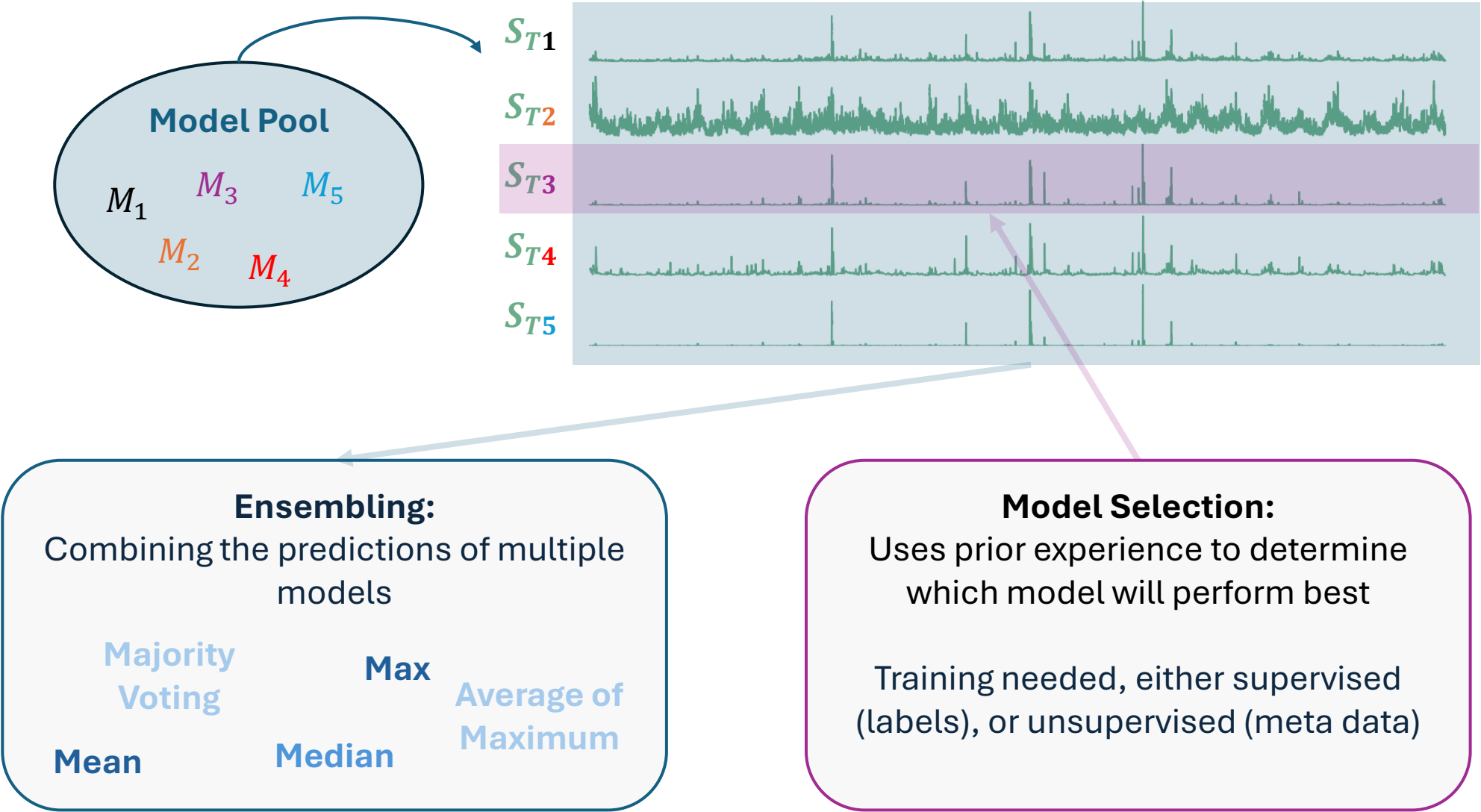
Can be overly penalizing!



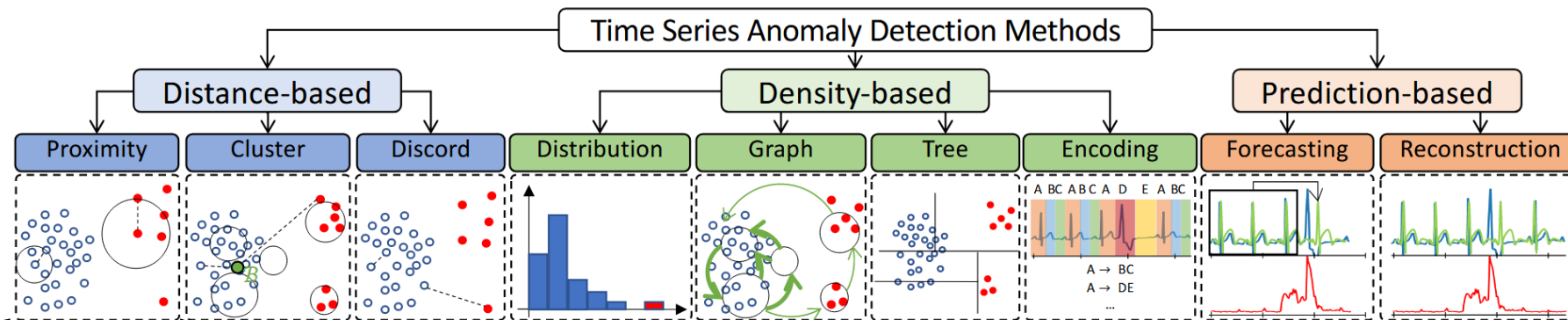
VUS-PR



Appendix: Automated Anomaly Detection (AutoAD)

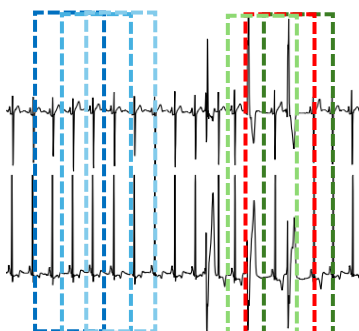


Appendix: TSAD methods examples



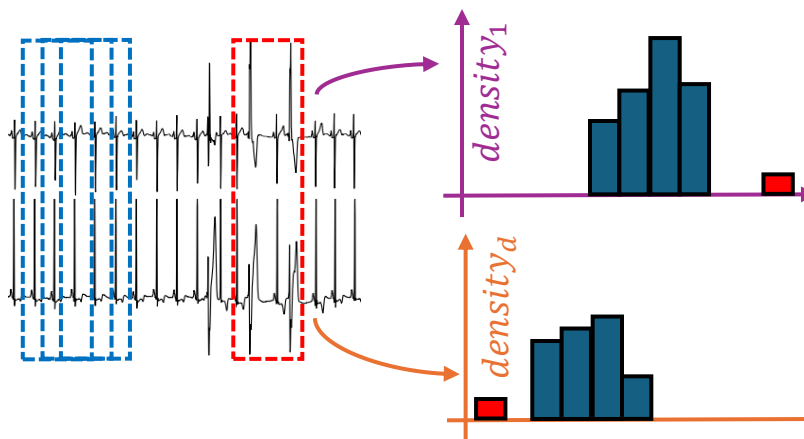
***k*-NN [19]**

$d \equiv \text{Euclidian dist}$



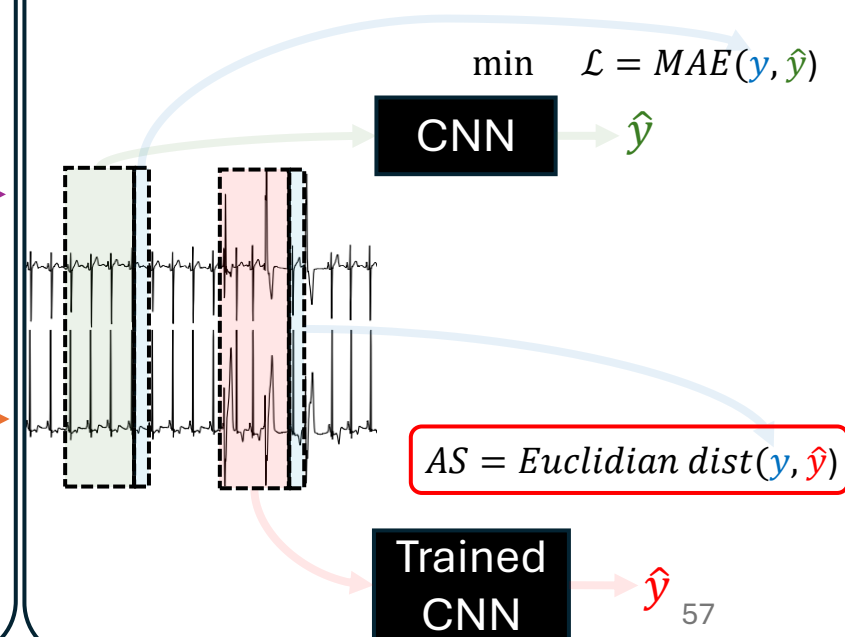
$$AS = d_k \text{ or } AS = \text{mean}(d_1, \dots, d_k)$$

HBOS [20]



$$AS = \sum_{i=1}^d \log \left(\frac{1}{\text{density}_i + \alpha} \right)$$

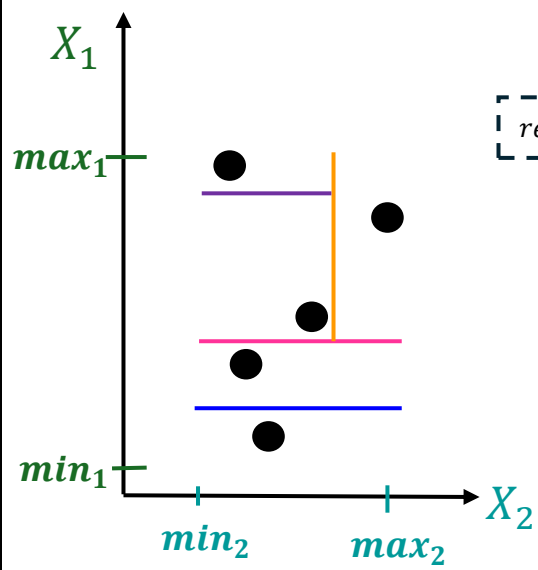
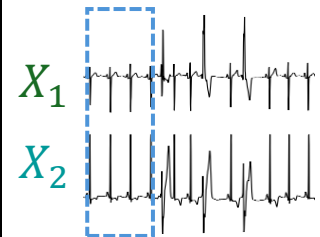
CNN [18]



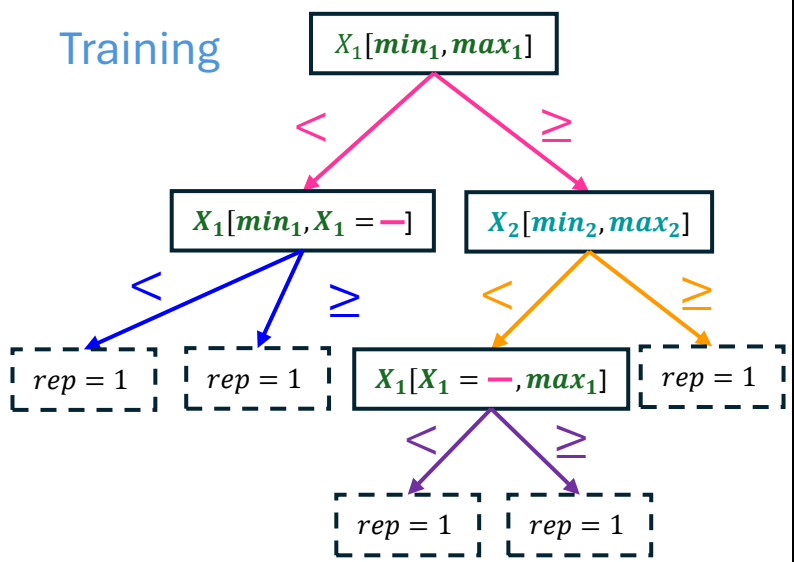
Appendix: Streaming TSAD methods examples

Structural Update:
RRCF [21]

Train



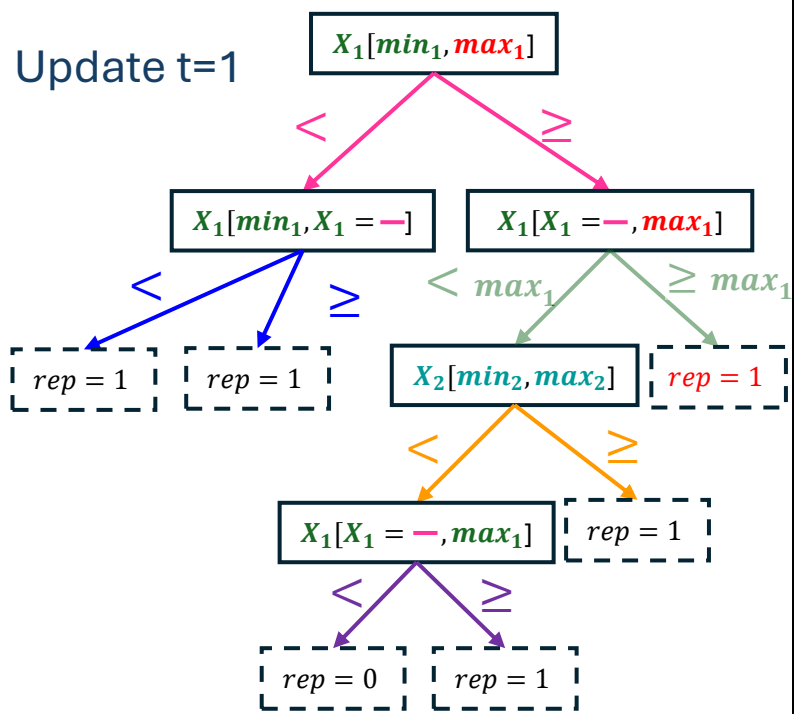
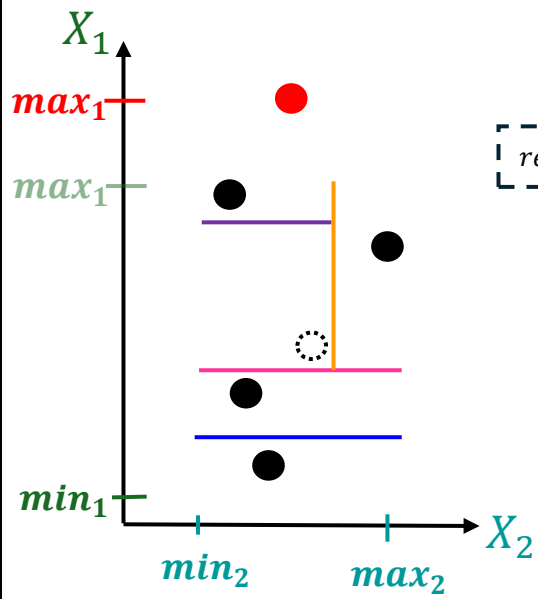
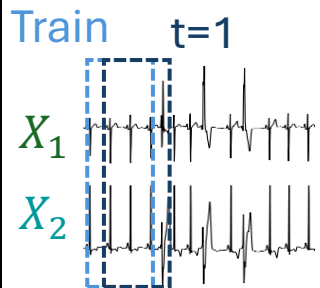
Training



Numerical Update:
RSHash [22]

Appendix: Streaming TSAD methods examples

Structural Update:
RRCF [21]



$AS = CoDisp(\bullet)$

Numerical Update:
RSHash [22]

RRCF Memory Management: **Hard Forgetting**
Sliding Window (FIFO)

StrAD: Framework Description

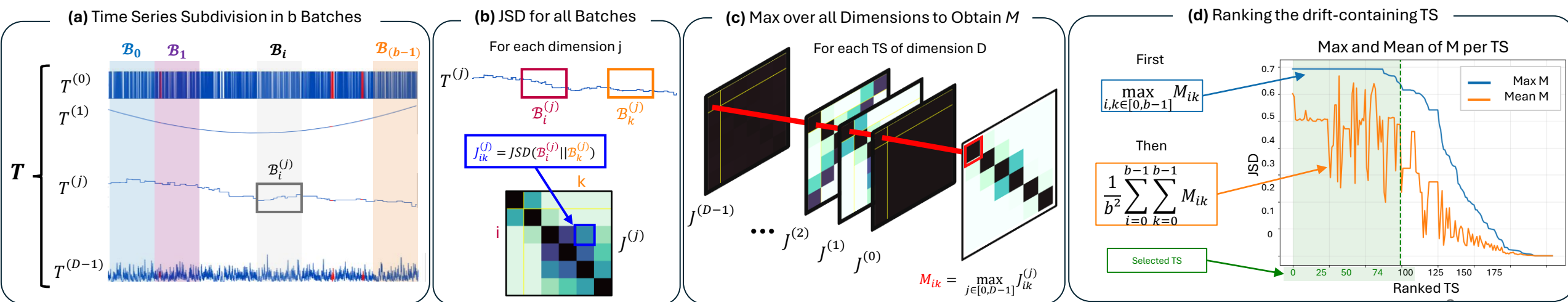
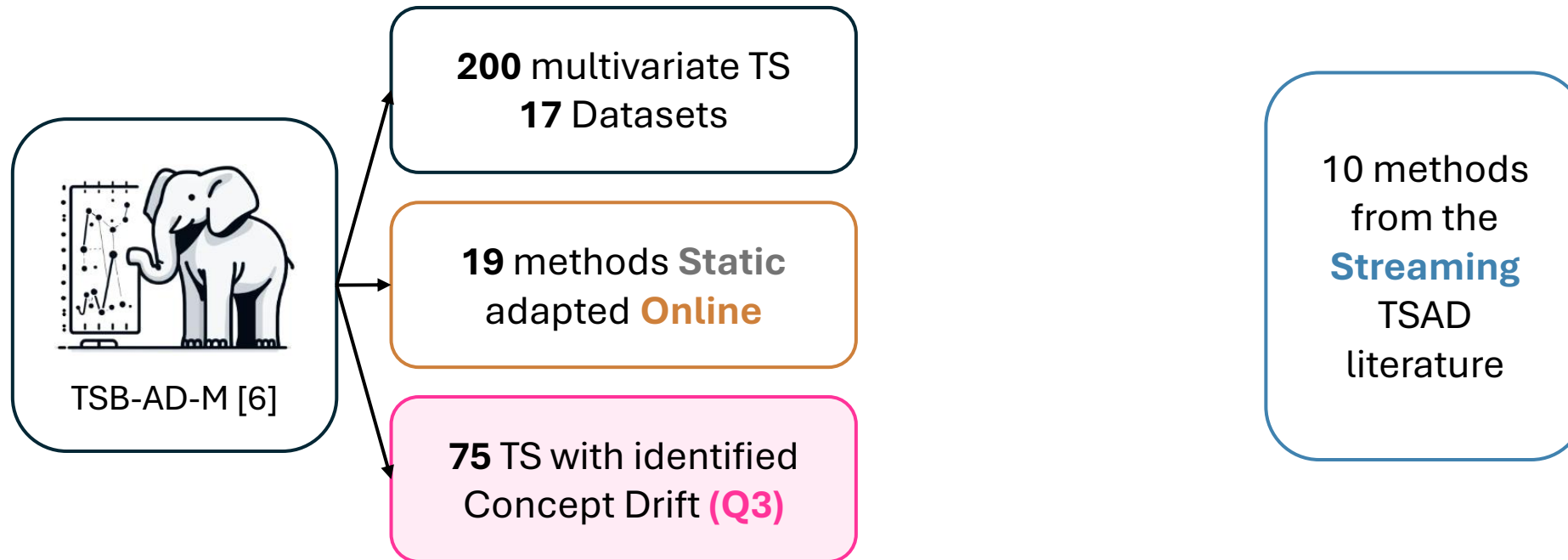
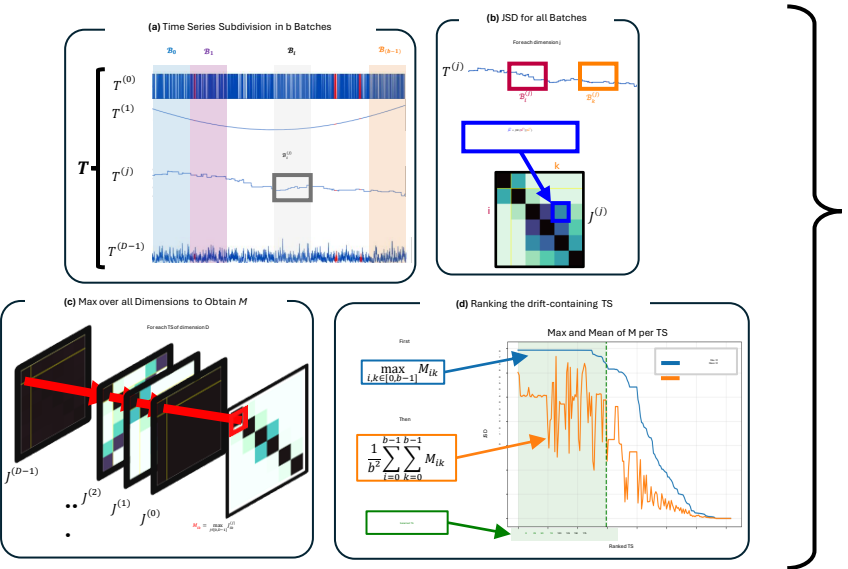
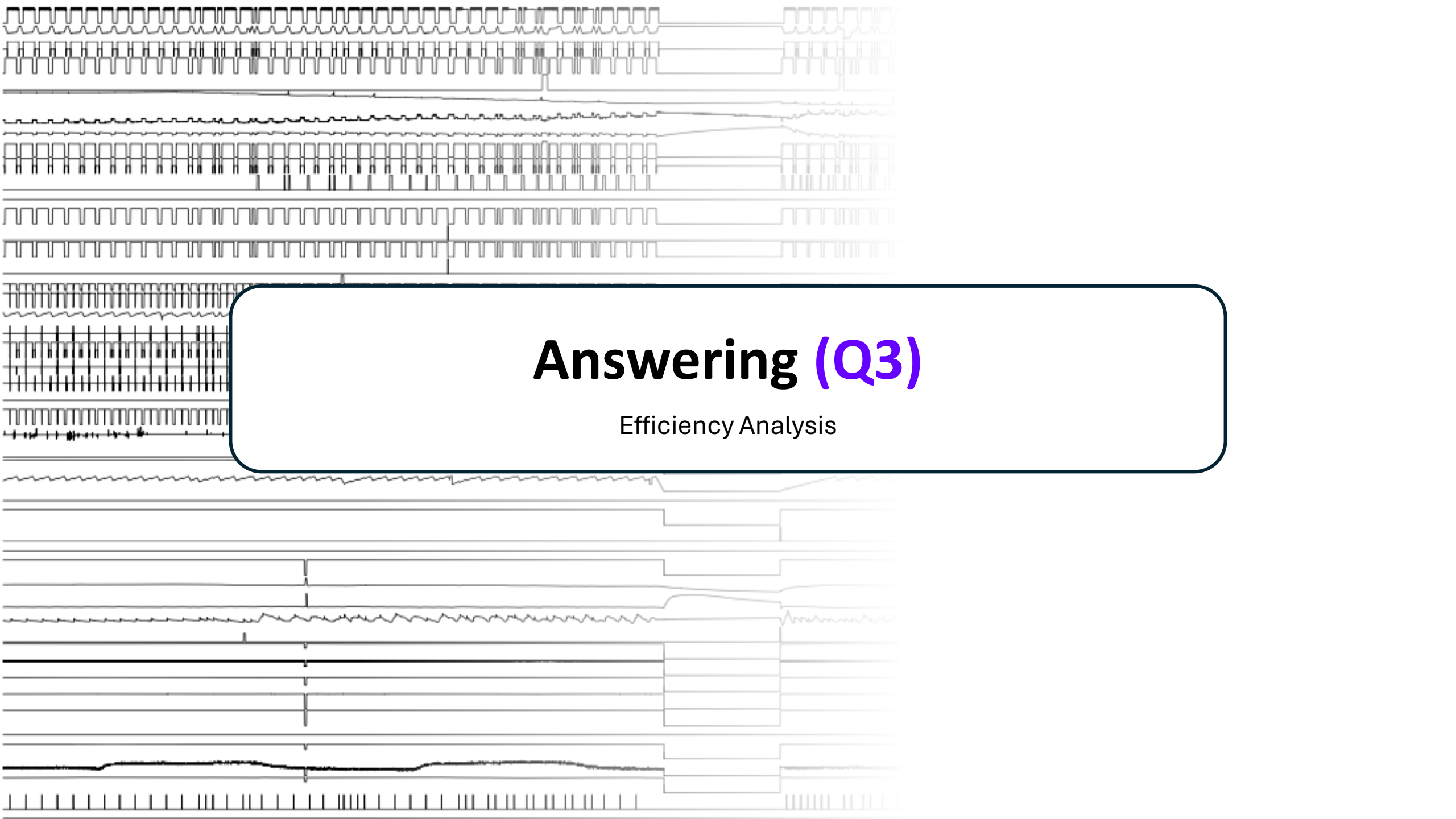


Figure 3: Construction process of TSB-drift

Appendix: TSB-drift drift characterization



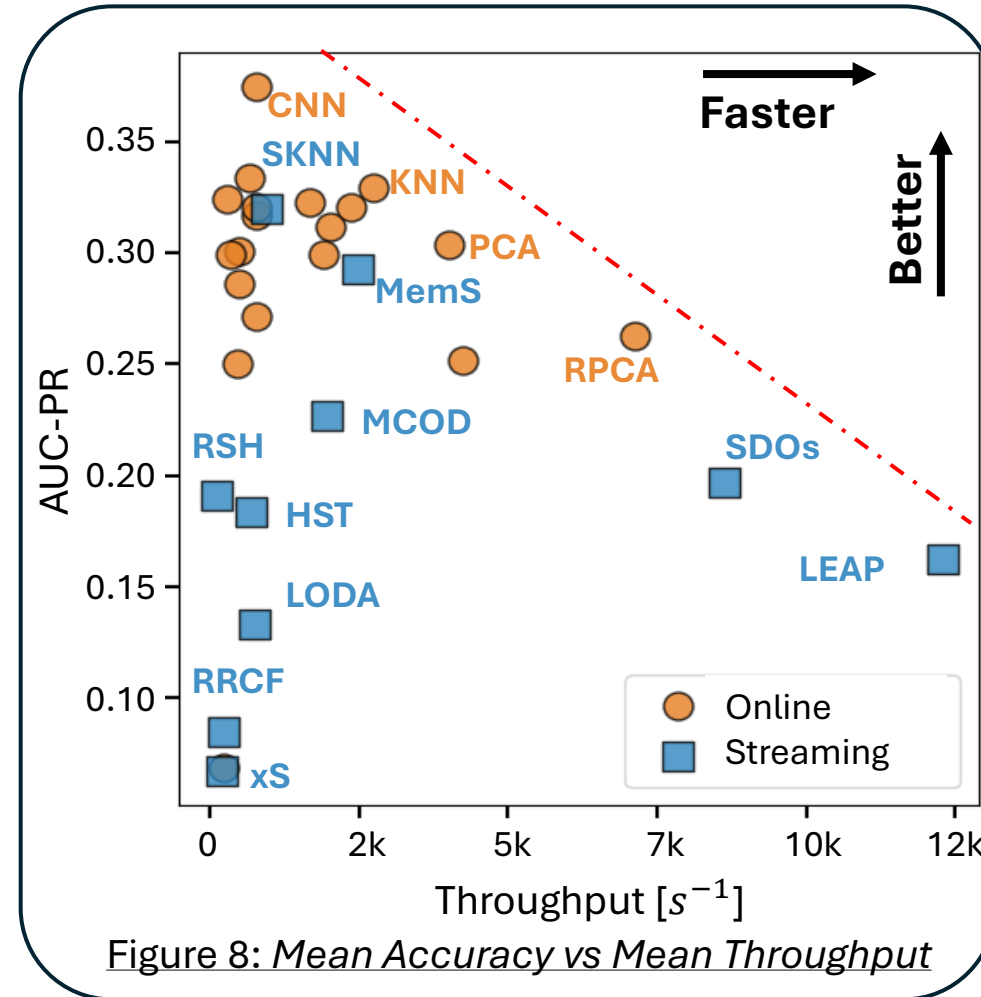
Matrix Pattern	Time Series	Concept Drift	Occurrence Examples
		Continuous (C)	Sensor Degradation
		Change Point (CP)	Dynamic Allocation
		Periodic drift (P)	Seasonality
		Random Walk (RW)	Physiological Variability
		No drift	Medical (ECG)

The background of the slide features several horizontal tracks of digital waveforms, resembling timing diagrams or signal traces. These tracks are rendered in a light gray, semi-transparent style, creating a technical and data-oriented aesthetic. The waveforms vary in frequency and complexity, with some showing regular square waves and others displaying more irregular, noisy patterns.

Answering (Q3)

Efficiency Analysis

Answering (Q3): Efficiency Analysis



(Q3) While **Online** methods are **more accurate**, **Streaming** methods **dominate the computational high efficiency** end of the spectrum, making them irreplaceable in high velocity domains.