DeepLLR-CUSUM: Sequential Change Detection with Learned Log-Likelihood Ratios for Site Reliability Engineering

Nadim Ahmed[1], Md. Mahmudul Hasan[1], Md. Ashraful Babu[1], Md. Mortuza Ahmmed[2], M. Mostafizur Rahman[2], and Mufti Mahmud[3]

[1]Independent University, Bangladesh, Dhaka-1229, Bangladesh, [2]American International University - Bangladesh, Dhaka-1229, Bangladesh, [3]King Fahd University of Petroleum and Minerals, Dhahran, KSA.

Submission Number: 46

Highlights

- Lightweight MLP learns a log-likelihood ratio from whitened features and integrates directly into the classical CUSUM framework.
- Provides principled ARL-controlled thresholds with matched-ARL evaluation on CESNET and SPY+VIX datasets.
- Produces interpretable per-sample log-evidence scores while remaining computationally efficient for real-time SRE deployment.

Problem Setup & Background

Sequential change-point detection concerns a data stream $\{x_t\}$ that follows p_0 until an unknown change time ν, after which it follows p_1. The objective is to raise an alarm quickly after ν while maintaining a large in-control Average Run Length (ARL), which quantifies the expected time to a false alarm under p_0.

Classical CUSUM [3] uses the log-likelihood ratio (LLR)

$$\ell(x_t) = \log \frac{p_1(x_t)}{p_0(x_t)},$$

updated recursively with an ARL-calibrated threshold. However, Gaussian and other simple parametric assumptions often break down in high-dimensional, non-Gaussian, and heavy-tailed telemetry streams [1].

DeepLLR–CUSUM addresses this limitation by learning a discriminative approximation to the LLR via a lightweight neural model and inserting it into the classical CUSUM recursion. This preserves likelihood-ratio optimality, maintains explicit ARL control, and scales effectively to nonlinear real-world data.

Methodology

DeepLLR trains a small MLP to estimate log-likelihood ratios from whitened features. Pre-change DeepLLR scores determine the drift correction and the bootstrap-based ARL calibration through an empirical MGF-root. The calibrated increments feed a classical CUSUM recursion, producing a fast and interpretable sequential detector. The complete pipeline is summarized in Fig. 1.

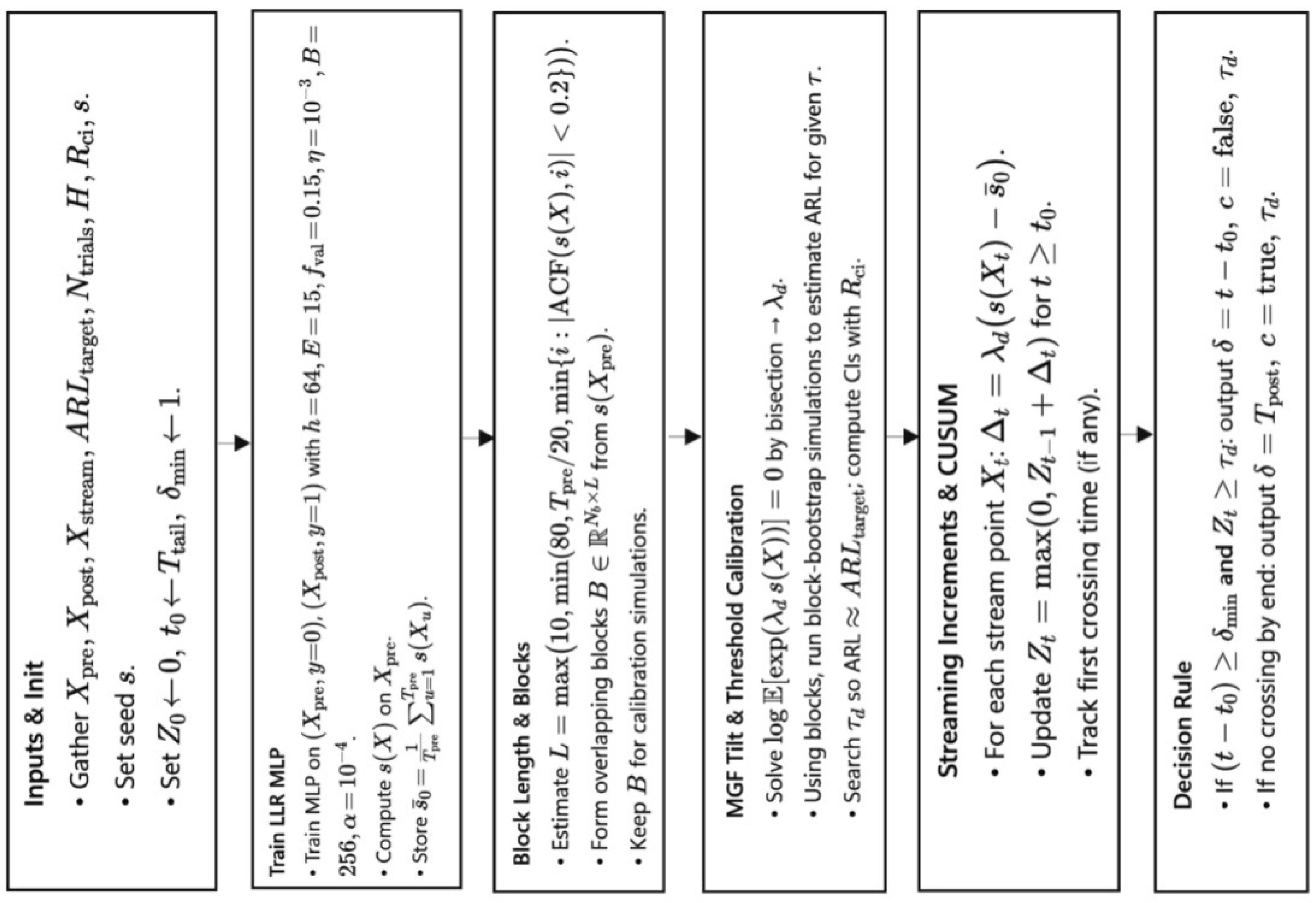

Fig. 1: DeepLLR-CUSUM pipeline: MLP-based LLR learning, drift adjustment, ARL calibration, and CUSUM recursion.

DeepLLR-CUSUM Workflow & Calibration

Offline: Train & Calibrate

- Build whitened features from X_{pre} and X_{post}.
- Train MLP on $(y = 0)$ vs. $(y = 1)$ to obtain $s(x)$.
- Estimate scale λ via empirical MGF-root using X_{pre}.
- Block-bootstrap pre-change scores to choose τ s.t. ARL $\approx$ target.

Online: Sequential Monitoring

- Initialize $Z_0 = 0$.
- For each new x_t:
  - Compute LLR $s(x_t)$. Update $Z_t = \max(0, Z_{t-1} + \lambda s(x_t))$.
  - If $Z_t \geq \tau$: raise alarm.

Abstract

DeepLLR-CUSUM is a sequential change detection method that uses a compact MLP to directly learn a discriminative log-likelihood ratio and plug it into a CUSUM-style detector. On CESNET telemetry and SPY+VIX financial data, DeepLLR-CUSUM achieves lower detection delay than Gaussian and LSTM-based baselines at matched ARL, while remaining computationally efficient and interpretable for online monitoring in site reliability engineering.

Results and Discussion

Experimental Setup. DeepLLR–CUSUM is evaluated on the **CESNET network telemetry dataset** [2], using controlled higher–order (shape and dependence) post-change injections. Baselines include Gaussian–CUSUM and LSTM–CUSUM, all calibrated to the same in-control ARL using block-bootstrap methods.

Key Findings.

- **Fastest detection:** DeepLLR reaches the CUSUM threshold within 1–2 samples, ahead of Gaussian and significantly faster than LSTM (see Fig. 2).
- **Robustness:** Learned LLR increments remain stable under heavy-tailed distortions where Gaussian modeling deteriorates.
- **Matched-ARL fairness:** Delay improvements persist under equal false-alarm budgets, as shown by RMST/EDD summaries in Fig. 3.

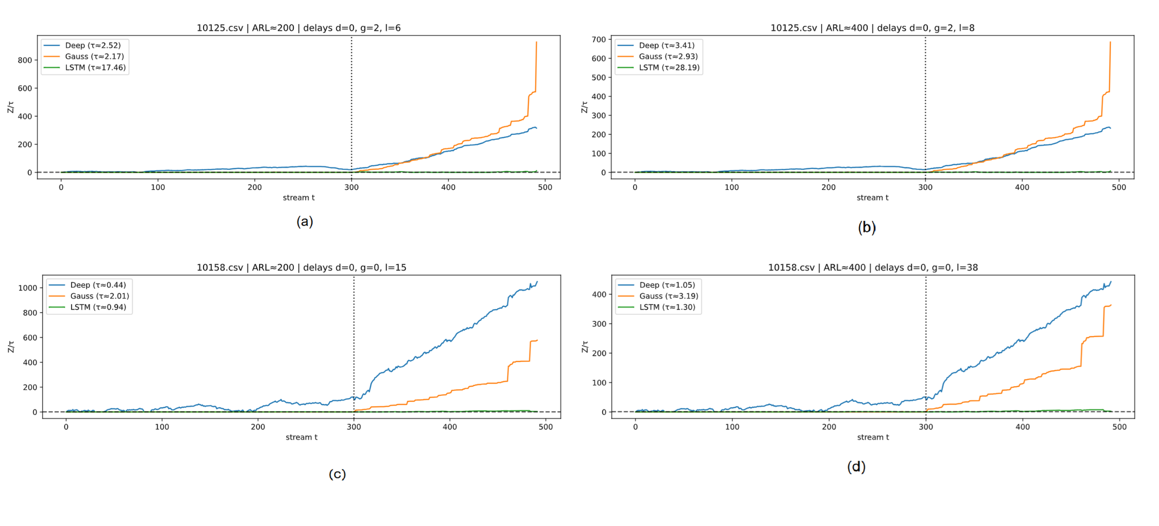

Fig. 2: Normalized CUSUM traces (Z/τ) showing DeepLLR's rapid threshold crossing relative to Gaussian and LSTM.

Interpretability.

- Each score $s(x_t)$ acts as a **per-sample log-evidence ratio**.
- Attribution analysis indicates variance and skewness features dominate near true changes.

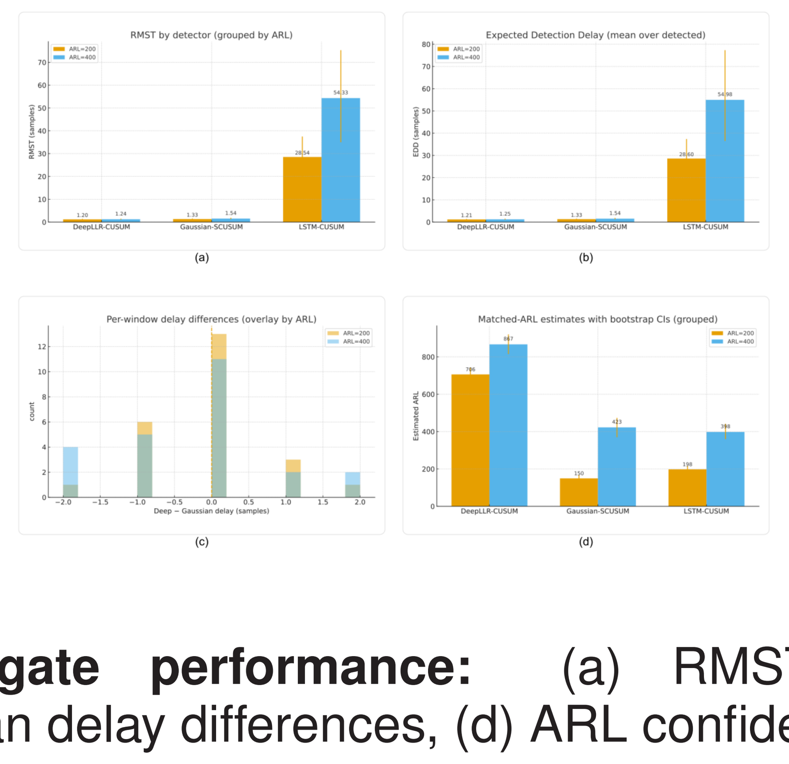

Fig. 3: Aggregate performance: (a) RMST, (b) EDD, (c) DeepLLR–Gaussian delay differences, (d) ARL confidence intervals.

Practical Implication. DeepLLR–CUSUM jointly provides:

- near-instant detection,
- conservative false-alarm behavior, and
- interpretable evidence scoring,

making it suitable for real-time SRE and anomaly detection pipelines.

References

[1] S. Aminikhanghahi and D. J. Cook. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51(2):339–367, 2017.

[2] CESNET-CZ. Timeseries24: Aggregated network traffic statistics, 2024. Network telemetry dataset used for change-point experiments.

[3] E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.