УДК DOI: https://doi.org/10.15276/eltecs.46.122.2026.10

Reliability-Constrained Multimodal Anomaly Detection for Streaming Non-Stationary Time Series Data Analysis

Illia Uzun [1], *Postgraduate Student*
ORCID: http://orcid.org/0000-0001-6619-4862; e-mail: uzun.i.s@op.edu.ua;
Scopus Author ID: 57223316393
Mykhaylo Lobachev [1], *PhD, Professor*
ORCID: http://orcid.org/0000-0002-4859-304X; e-mail: lobachev@op.edu.ua;
Scopus Author ID: 36845971100
[1] *Odesa Polytechnic National University*

Abstract. *Streaming decision support systems frequently face two coupled operational constraints: multimodal input streams are non-stationary and periodically exhibit modality degradation (missing values, additive noise, gain or scale shifts), while alarm frequency must be limited by an explicit false alarm rate (FAR) budget. Naive multimodal residual scoring then inflates anomaly scores during degraded segments, pushing FAR-constrained thresholds upward and reducing detector sensitivity to genuine anomaly events.*

The aim is to design and empirically validate a residual scoring policy that retains anomaly sensitivity under alternating modality degradation at a fixed FAR budget. The research tasks are to formalise the streaming multimodal anomaly detection problem under a FAR budget in a causal setting, to design an online modality reliability estimator from cheap causal degradation cues (missingness and feature-energy inflation), to construct a reliability-gated rule for selecting the source of the residual evidence, and to validate the approach on a controlled stream and on the UCI Air Quality dataset.

The paper proposes RC-AD, a reliability-constrained residual scoring policy that combines standardised residual scores from modality-specific and early-fusion multimodal predictors, online modality reliability weights estimated from short causal windows using the missingness and feature-energy inflation, and a winner-takes-all rule: in clean windows the early-fusion residual score is used, otherwise the residual score of the currently more reliable modality. RC-AD is a control policy on top of an arbitrary forecasting backbone, which makes it lightweight, auditable, and easy to integrate into existing monitoring pipelines.

The empirical study uses a causal protocol with ten fixed seeds and 95% confidence intervals: on a controlled benchmark with alternating modality degradation and injected event anomalies, RC-AD improves Recall@FAR from 0.103 to 0.335 at FAR 0.05 and from 0.182 to 0.426 at FAR 0.10, outperforming naive multimodal, equal-weight fusion and single-modality baselines; a demonstration on UCI Air Quality at FAR 0.01 confirms the same trend against the multimodal baselines. The scientific novelty consists in formulating a multimodal anomaly detection policy with an explicit online reliability constraint that, for the first time, combines gating of early-fusion evidence by causal degradation cues with selection of the most reliable single modality at a prescribed FAR budget. The practical significance is supported by reproducible Recall@FAR improvements on both controlled and real sensor data and by the lightness and interpretability of the method, which suits auditable decision support systems

Key words: *machine learning, data analysis, information systems, decision support systems, multimodal time series, non-stationary time series, streaming anomaly detection, data quality degradation, modality reliability estimation, false alarm rate budget.*

Article citation: Uzun I. S. and Lobachev M. V. (2026). Reliability-Constrained Multimodal Anomaly Detection for Streaming Non-Stationary Time Series Data Analysis. *Electrotechnic and Computer Systems,* 46(122), pp.107-119. doi:https://doi.org/10.15276/eltecs.46.122.2026.10

Introduction

In online monitoring and decision support, false alarms are not a minor inconvenience but a measurable operational cost: frequent false triggers can desensitize operators and destabilize downstream responses. Therefore, anomaly detection in deployed systems is often constrained by an explicit false alarm rate (FAR) budget, e.g., β=1% or β=5%, which limits the fraction of «normal» observations that may be mistakenly flagged. At the same time, modern decision support pipelines increasingly rely on multiple data sources or sensors, which creates a multimodal stream. In multimodal settings, the practical meaning of «abnormal» becomes ambiguous because temporary degradation of a modality (sensor outages, missing values, elevated noise, scale shifts) may produce large residuals even when the underlying process is not anomalous.

This paper addresses reliability-constrained anomaly detection in streaming multimodal time series under a fixed FAR budget. The key idea is to treat modality reliability as an online control signal that selects which residual evidence should be trusted at each time step. The resulting procedure, RC-AD, is intentionally lightweight and auditable: it uses residual scoring and a simple reliability-gated fusion policy that falls back to a multimodal score only when both modalities appear clean.

Two design principles guide the proposed approach. First, the detector must expose an explicit control parameter that corresponds to operational constraints; in this paper, that parameter is the FAR budget β. Second, the detector must be robust to a common failure mode in multimodal streams: a short-lived data quality drop in one source that manifests as large residuals and produces spurious alerts if multimodal evidence is used indiscriminately. These principles are implemented in RC-AD by using online reliability estimates to gate residual evidence: early fusion is used only in clean regimes, while degraded regimes fall back to the currently more reliable modality.

The contributions of this paper are as follows: (i) a reliability-gated residual scoring policy that mitigates false alarms caused by temporary modality degradation under a fixed FAR budget; (ii) a controlled benchmark with alternating degradation that isolates the interaction between data quality shifts and FAR constraints; and (iii) empirical results on a controlled stream and a real-data demonstration on UCI Air Quality.

1 Literature review

Anomaly detection has been studied broadly as a set of methods for identifying patterns that deviate from an expected notion of normality, with common families including statistical, distance-based, and model-based approaches. The survey in [1] emphasizes that the practical meaning of «anomaly» is application-dependent and that evaluation must reflect operational costs. For time-series data, multiple reviews summarize classical and modern approaches and highlight the role of temporal dependence and non-stationarity in evaluation [2-4]. Deep-learning-based anomaly detection is discussed in broader surveys [3, 5] and in time-series-specific surveys [4]. Representative deep-learning multivariate time-series detectors include LSTM encoder–decoder residuals [6], nonparametric dynamic thresholding [7], stochastic recurrent models [8], and deep auto-encoding Gaussian mixtures [9]. For streaming data, evaluation protocols must additionally respect causality, i.e., decisions at time t may depend only on information available up to t. The prequential (test-then-train) perspective proposed in [10] provides a principled framework for such causal evaluation and is widely used in data-stream learning and benchmarking [11-14]. In non-stationary streams, changes of the generating mechanism are often described as concept drift, and practical systems may need both drift-aware learning and monitoring. Surveys in the data-stream community summarize drift types, adaptation strategies, and the gap between batch and streaming evaluation [15, 16]. Common drift detectors (e.g., adaptive windowing) are frequently used as lightweight monitors of distributional changes in streaming features and can be combined with anomaly modules in a pipeline [17]. Classical drift and change detection methods (e.g., CUSUM-style schemes and general abrupt change detection frameworks) remain relevant when the goal is to detect distributional shifts in a principled and interpretable way [18, 19]. RC-AD is complementary to drift detectors: it does not attempt to identify drift explicitly, but it aims to keep anomaly decisions stable under non-stationarity by controlling the operating point through an explicit FAR budget.

Receiver operating characteristic (ROC) analysis clarifies the relationship between score thresholding and trade-offs between false alarms and missed detections [20]. In operational monitoring, this trade-off is often fixed by external requirements: for example, alarms must occur at most β fraction of normal observations. This motivates constrained metrics such as «Recall@FAR» , which report sensi tivity at a fixed FAR budget rather than optimizing a global ranking metric alone. RC-AD follows this constraint-first view by evaluating and configuring operation at a prescribed FAR budget β; in streaming deployments, the same contract can be approximated by online quantile calibration of thresholds on recent scores. Since anomaly datasets are often imbalanced, precision-recall analysis is also important; the relationship between ROC and precision-recall curves and recommended reporting practices are discussed in [21, 22].

Multimodal machine learning provides mechanisms to combine heterogeneous sources, but practical robustness depends on how fusion reacts to missing or degraded modalities. The taxonomy in [23] and the overview in [24] describe common fusion strategies and stress that real-world multimodal systems must address imperfect data sources. Additional surveys discuss multimodal fusion for multimedia analysis and multisensor data fusion as broader contexts for robust fusion design [25, 26]. In streaming decision support, temporary modality degradation can masquerade as a regime change and inflate residuals; therefore, anomaly scoring should incorporate an explicit notion of source reliability to avoid systematic false alarms during degraded segments. RC-AD operationalizes this idea by using online reliability estimates as a control signal for residual evidence selection and fusion under degradation.

2 Methodology

Consider a time-indexed multimodal stream with two modalities $\boldsymbol{x}_t^{(1)} \in \mathbb{R}^{p_1}$, $\boldsymbol{x}_t^{(2)} \in \mathbb{R}^{p_2}$ and a target y_t (e.g., the next-step value of a monitored variable). Let $\hat{y}_t^{(m)}$ be an online predictor that uses only modality m, and let $e_t^{(m)} = y_t - \hat{y}_t^{(m)}$ denote the corresponding one-step residual. Residual magnitudes $|e_t^{(m)}|$ can be used as anomaly evidence, but their interpretation becomes unreliable when modality m is degraded. To keep the anomaly module independent of the forecasting backbone, the paper treats the predictors $\hat{y}_t^{(m)}$ as black boxes and focuses on how residual evidence is aggregated and thresholded under operational constraints. This design is aligned with decision support practice, where anomaly detection is frequently an add-on monitoring component that must remain computationally lightweight and interoperable.

In addition to modality-specific predictors, a simple early-fusion multimodal predictor $\hat{y}_t^{(12)}$ can be formed by concatenating modality features and fitting a single model. In this paper, all residual magnitudes are standardized using statistics estimated on a clean training prefix to obtain comparable anomaly scores across modalities.

Let $r_t^{(m)} \in [0,1]$ denote an online reliability estimate for modality m. In the context of this paper, reliability is interpreted as a probability-like confidence that the modality is in a non-degraded state (based only on past and current observations). Reliability can be produced by a separate lightweight model that reacts to missingness and distributional changes in the modality features; however, RC-AD does not require a specific estimator and uses reliability only through its role as a control signal. In particular, reliability is used to decide whether multimodal evidence should be trusted (clean regime) and, if not, which modality should dominate the residual evidence (degraded regime). Since missingness is one of the most common degradation modes, standard missing-data literature provides useful context for interpreting reliability gates based on missingness rates [27].

Let $\beta \in (0,0.5]$ be a predefined FAR budget. Given a score stream $\{s_t\}$, the operating point is defined by a threshold $\tau(\beta)$ chosen so that the fraction of false alarms on normal observations does not exceed β. To evaluate sensitivity under this constraint, the paper uses Recall@FAR, defined as recall (true positive rate) achieved when the detector operates at the prescribed FAR budget β. In the experiments below, Recall@FAR is computed using the standard oracle threshold selection on the evaluation split (equivalently, a $(1-\beta)-$quantile of scores on normal points), which isolates score quality under the FAR constraint. Section 7 discusses streaming-friendly online threshold calibration as a deployment option.

Let $\hat{y}_t^{(m)}$ denote modality-specific predictors and $e_t^{(m)} = y_t - \hat{y}_t^{(m)}$ their one-step residuals. To obtain comparable anomaly evidence across modalities, RC-AD uses standardized residual magnitudes:

$$s_t^{(m)} = (|e_t^{(m)}| - \mu_m)/\sigma_m\,, \qquad (1)$$

where μ_m and σ_m are the mean and standard deviation of $|e_t^{(m)}|$ estimated on a clean training prefix. In addition, a simple early-fusion multimodal predictor $\hat{y}_t^{(12)}$ can be constructed using concatenated modality features, yielding a standardized multimodal residual score $s_t^{(12)}$. For comparison, an equal-weight late-fusion baseline can be defined at the prediction level as $\hat{y}_t^{(\mathrm{eq})} = 0.5\hat{y}_t^{(1)} + 0.5\hat{y}_t^{(2)}$, with the corresponding standardized residual score $s_t^{(\mathrm{eq})}$.

RC-AD estimates modality reliability using cheap causal cues computed over a short window:

missingness and inflation of feature energy relative to the clean prefix. Let $\text{miss}_t^{(m)} \in [0,1]$ be the fraction of missing values in modality m in the current window, and let $E_t^{(m)}$ be the NaN-safe mean squared feature magnitude in that window. With baseline energies $E_0^{(m)}$ estimated on the clean prefix, define an energy-inflation penalty $\text{pen}_t^{(m)} = \max(0, E_t^{(m)}/(E_0^{(m)} + \varepsilon) - 1)$. A simple reliability proxy is:

$$\tilde{r}_t^{(m)} = \pi_m / \left((1 + k_{\text{miss}} \text{miss}_t^{(m)})(1 + k_{\text{var}} \text{pen}_t^{(m)}) \right), \quad (2)$$

where π_m is an optional static prior (e.g., reflecting modality predictive strength on the clean prefix) and $k_{\text{miss}}, k_{\text{var}} > 0$ tune sensitivity. Reliabilities are normalized into an instantaneous weight $w_t^* \in [0,1]$ for modality 1:

$$w_t^* = \tilde{r}_t^{(1)} / (\tilde{r}_t^{(1)} + \tilde{r}_t^{(2)} + \varepsilon) \quad (3)$$

and smoothed to avoid jitter:

$$w_t = \alpha w_t^* + (1 - \alpha) w_{t-1} \quad (4)$$

with $\alpha \in (0,1]$.

RC-AD uses reliability to decide whether multimodal evidence should be trusted (clean regime) and, if not, which modality should dominate the score (degraded regime). Define a clean-window indicator:

$$[\text{miss}_t^{(1)} \le \delta \wedge \text{miss}_t^{(2)} \le \delta \wedge \text{pen}_t^{(1)} \le \gamma \wedge \text{pen}_t^{(2)} \le \gamma], \quad (5)$$

where $\delta \in (0,1)$ and $\gamma \ge 0$ are clean gates.

The RC-AD score is then selected as:

$$s_t = \begin{cases} s_t^{(12)}, & c_t = 1, \\ s_t^{(1)}, & c_t = 0 \wedge w_t \ge 0.5, \\ s_t^{(2)}, & c_t = 0 \wedge w_t < 0.5. \end{cases} \quad (6)$$

This policy preserves sensitivity in clean regimes via early fusion, while reducing score inflation when a modality is degraded by falling back to the currently more reliable single-modality residual.

Given a score stream $\{s_t\}$, the detector operates under a FAR budget β by choosing a threshold $\tau(\beta)$ so that the fraction of false alarms on normal points does not exceed β. In evaluation, Recall@FAR is computed using the oracle threshold that satisfies the FAR budget on the evaluation split; Section 7 discusses streaming-friendly online quantile calibration as a deployment option.

Figure 1 summarizes the RC-AD pipeline and highlights the separation between (i) reliability-gated scoring and (ii) FAR-budget threshold selection.

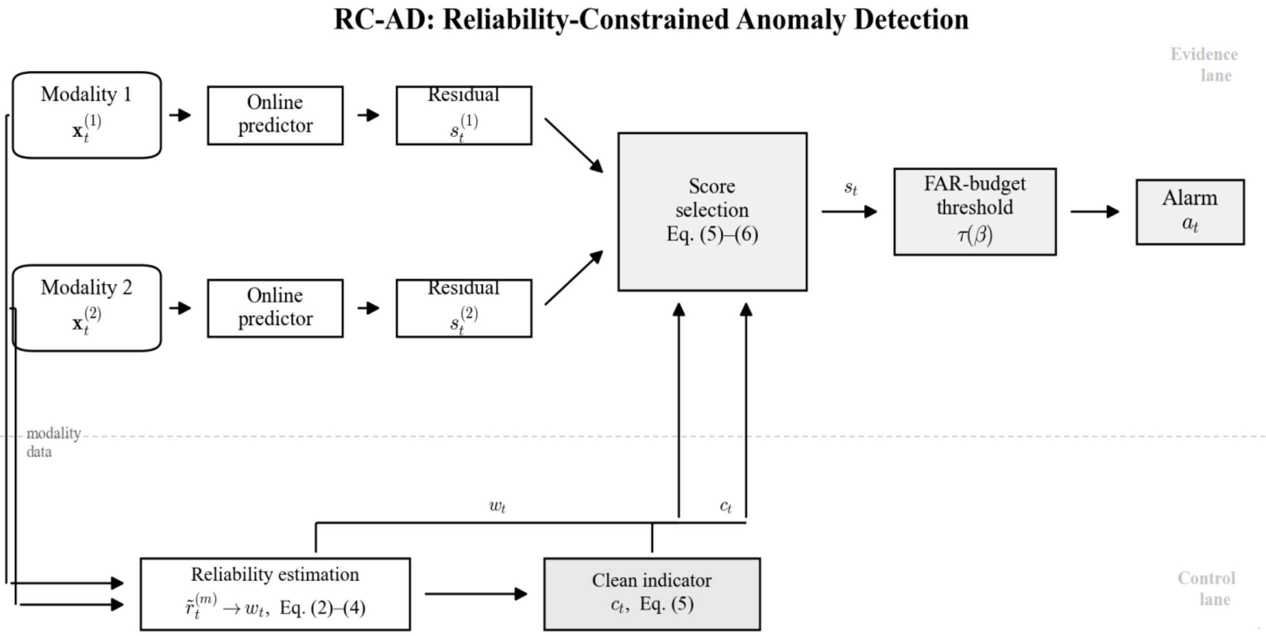

Fig. 1 - RC-AD schematic pipeline

The controlled benchmark uses a stream of length $T = 5000$ with a clean training prefix of length 1000. Anomalies are injected only into the test part and are generated as *events* (not isolated points), with an overall anomaly rate of 0.05 and an average event duration of 12 steps. Three event types are used: (i) spikes, (ii) level shifts, and (iii) variance bursts. To assess robustness, modality degradation alternates in the test part with segment length $\ell = 400$: modality 1 is degraded in even segments, modality 2 in odd segments. Degradation combines missing values (rate 0.2), additive noise (level 2.0), and a multiplicative scale factor (2.0). Performance is evaluated for FAR budgets $\beta \in \{0.01, 0.05, 0.10\}$ using a causal train/test split with 10 fixed random seeds. The use of event anomalies (rather than point outliers) is important for decision support, where anomalous situations often persist over a short interval and a detector should react within the event window. Alternating degradation further ensures that a robust multimodal detector cannot rely on a single permanently clean source.

The real-data demonstration uses the UCI Air Quality dataset with a chronological split $70\%/15\%/15\%$ (train/validation/test) and z-score normalization based on the training part. To construct a meaningful multimodal setting, modalities are formed from complementary sensor feature subsets with shared time context, and modality 2 additionally includes meteorological/context features. To enable controlled evaluation under degradation, anomaly events are injected into the target variable in the test part, and missingness degradation is applied in alternating segments. Results are reported as Recall@FAR at $\beta = 0.01$ on degraded segments (10 fixed seeds), along with PR-AUC as a complementary ranking metric. The degraded-segment focus is consistent with the paper's goal: to prevent modality degradation from dominating the FAR budget and to preserve sensitivity where the system is expected to operate despite partial data quality drops. The dataset is obtained from the UCI Machine Learning Repository and originates from an urban pollution monitoring scenario, which motivates the presence of missing values and sensor-related artifacts [28, 29].

To keep the method reproducible and easy to integrate into decision-support pipelines, the evaluation uses a compact, explicitly stated configuration. Table 1 summarizes the key parameters of the controlled benchmark and the online RC-AD module. This summary is intended to make the paper self-contained for later conversion to DOCX and for practical reimplementation.

Table 1 - Key configuration used in the controlled benchmark and RC-AD (streaming evaluation)

Component	**Setting**
FAR budgets (β)	$\{0.01,\ 0.05,\ 0.10\}$
Evaluation protocol	causal train/test stream split (clean prefix + test stream), 10 fixed seeds
Controlled stream length	$T = 5000$, clean training prefix $= 1000$
Anomaly injection (test only)	event anomalies; rate 0.05; mean event length 12
Event types (target anomalies)	spike, level shift, variance burst
Degradation regime (test)	alternating modality degradation, segment length $\ell = 400$
Degradation mix	missing rate 0.2, additive noise level 2.0, scale factor 2.0
Predictors (all methods)	ridge regression with mean-imputation for missing values
Residual standardization	μ, σ of $\lvert e_t \rvert$ estimated on the clean prefix (per score)
Reliability window	causal window length 50
Reliability cues	missingness + energy inflation ($k_{\text{miss}} = 10$, $k_{\text{var}} = 8$); EWMA smoothing $\alpha = 0.25$
Clean fallback gate	$\text{miss}_t^{(m)} \le 0.02$ and $\text{pen}_t^{(m)} \le 0.50$ for both modalities
RC-AD selection policy	if clean: use $s_t^{(12)}$; else: use $s_t^{(1)}$ if $w_t \ge 0.5$, otherwise $s_t^{(2)}$
Thresholding for Recall@FAR	oracle threshold per β on normal points in the evaluation split

3 Results

Across the full corpus, the majority of headlines were classified as *neutral*, reflecting the model's conservative design and the heterogeneous character of news coverage. Bullish and bearish labels were assigned when entailment probabilities surpassed the 0.5 confidence threshold, yielding subsets of articles with strong directional signals. The average sentiment scores were centered near zero, while tails contained news items with values close to ±1, corresponding to highly directional headlines.

Table 2 reports Recall@FAR for three multimodal scoring strategies: Naive Multimodal (early-fusion residual scoring on concatenated modality features), Equal-Weight Fusion (late fusion of modality-specific predictions), and RC-AD (reliability-gated score selection; Section 3). RC-AD yields higher recall at the same FAR budgets, indicating improved sensitivity under degradation while keeping the operational constraint explicit.

Table 2 - Recall@FAR on the controlled stream under alternating modality degradation

Method	**FAR budget β**	Recall@FAR **(mean)**	**95% CI**
Naive Multimodal	0.01	0.034	0.021-0.047
Equal-Weight Fusion	0.01	0.093	0.065-0.122
RC-AD (proposed)	0.01	0.121	0.091-0.153
Naive Multimodal	0.05	0.103	0.069-0.140
Equal-Weight Fusion	0.05	0.186	0.148-0.222
RC-AD (proposed)	0.05	0.335	0.286-0.389
Naive Multimodal	0.10	0.182	0.137-0.232
Equal-Weight Fusion	0.10	0.273	0.223-0.322
RC-AD (proposed)	0.10	0.426	0.379-0.473

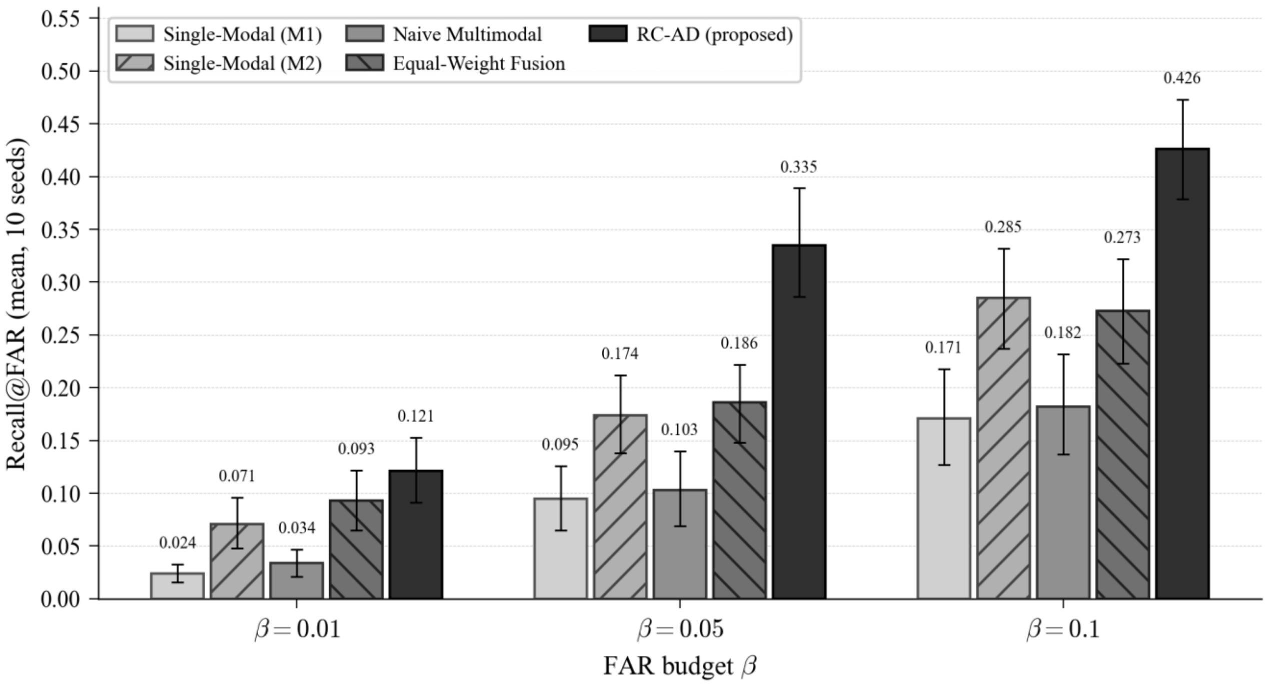

Fig. 2 - Recall@FAR comparison across budgets

To provide a broader view, Table 3 reports the same metric for additional baselines, including single-modality detectors. These results highlight two complementary aspects: (i) in degraded multimodal streams, naive multimodal scoring can underperform even simple equal-weight fusion, and (ii) reliability-constrained scoring can improve recall at fixed FAR budgets relative to both multimodal and single-modality baselines.

These results support two practical conclusions. First, when modality degradation alternates over time, naive multimodal scoring becomes unstable because the degraded modality periodically injects large residual magnitudes into the fused score. As a result, enforcing a FAR budget forces the threshold to rise, which reduces sensitivity to true anomaly events. Second, RC-AD mitigates this effect by reliability-gated score selection: in clean windows it uses the early-fusion multimodal residual score, while in degraded windows it falls back to the residual score of the currently more reliable modality. This reduces the influence of degraded modality features on the score distribution and yields higher Recall@FAR across budgets. For example, at $\beta = 0.05$, the mean Recall@FAR increases from 0.103 (naive multimodal) to 0.335 (RC-AD), while equal-weight fusion reaches 0.186. This pattern is consistent across $\beta \in \{0.01, 0.10\}$ and indicates that reliability-based gating is beneficial specifically in challenging regimes where data quality changes over time.

Table 4 summarizes results on UCI Air Quality. The evaluation isolates degraded segments to test whether the method reduces false alarms caused by degradation while retaining anomaly sensitivity under the prescribed FAR budget.

Table 3 - Controlled stream: extended comparison including single-modality baselines.

Method	**FAR budget β**	Recall@FAR **(mean)**	**95% CI**
Single-Modal AD (M1)	0.01	0.024	0.016-0.033
Single-Modal AD (M2)	0.01	0.071	0.048-0.096
Naive Multimodal	0.01	0.034	0.021-0.047
Equal-Weight Fusion	0.01	0.093	0.065-0.122
C-AD (proposed)	0.01	0.121	0.091-0.153
Single-Modal AD (M1)	0.05	0.095	0.065-0.126
Single-Modal AD (M2)	0.05	0.174	0.138-0.212
Naive Multimodal	0.05	0.103	0.069-0.140
Equal-Weight Fusion	0.05	0.186	0.148-0.222
RC-AD (proposed)	0.05	0.335	0.286-0.389
Single-Modal AD (M1)	0.10	0.171	0.127-0.218
Single-Modal AD (M2)	0.10	0.285	0.237-0.332
Naive Multimodal	0.10	0.182	0.137-0.232
Equal-Weight Fusion	0.10	0.273	0.223-0.322
RC-AD (proposed)	0.10	0.426	0.379-0.473

Table 4 - UCI Air Quality demonstration: evaluation on degraded segment.

Method	Recall@FAR **(mean)**	**95% CI**	**PR-AUC (mean)**	**95% CI**
Naive Multimodal	0.064	0.040-0.096	0.216	0.166-0.272
Equal-Weight Fusion	0.070	0.043-0.100	0.277	0.204-0.360
RC-AD (proposed)	0.097	0.059-0.135	0.290	0.215-0.368
Single-Modal AD (M1)	0.062	0.035-0.092	0.251	0.182-0.328
Single-Modal AD (M2)	0.105	0.070-0.141	0.302	0.229-0.383

On UCI Air Quality, RC-AD improves Recall@FAR relative to the two multimodal baselines on degraded segments, while the best single-modality baseline (M2) remains competitive in this setting. This suggests that when one modality is consistently more stable, single-modality monitoring can be a strong baseline; nevertheless, RC-AD provides a single operational score that adapts to alternating degradation patterns and can be paired with standard FAR-budget thresholding policies (offline calibration or online quantile calibration). Figure 3 provides an illustrative diagnostic visualization for interpretability; it mirrors the evaluation regime but uses a synthetic segment.

4 Discussion

The reported gains are obtained in a setting where anomaly events are injected into the target and modality degradation is controlled, which enables causal interpretation of why naive multimodal scoring loses sensitivity at a fixed FAR budget. RC-AD mitigates this failure mode by using online reliability cues to switch between score sources. However, the approach has limitations. If both modalities are degraded simultaneously, or if reliability cues are miscalibrated, RC-AD may not prevent score inflation and can lose sensitivity. The clean-fallback gates (δ, γ) and reliability parameters (window length, smoothing α, and penalty scales k_miss, k_var) require domain-appropriate tuning; overly permissive gates may let degraded data influence the multimodal score, while overly strict gates may overuse single-modality scoring and reduce the benefit of multimodal evidence.

From a validity perspective, the controlled benchmark is designed to isolate causal effects of alternating degradation and event anomalies, but it remains a simplified model of real-world complexity. External validity is supported by the UCI Air Quality demonstration, yet additional domains (e.g., industrial sensors, network telemetry) should be studied to assess generalizability.

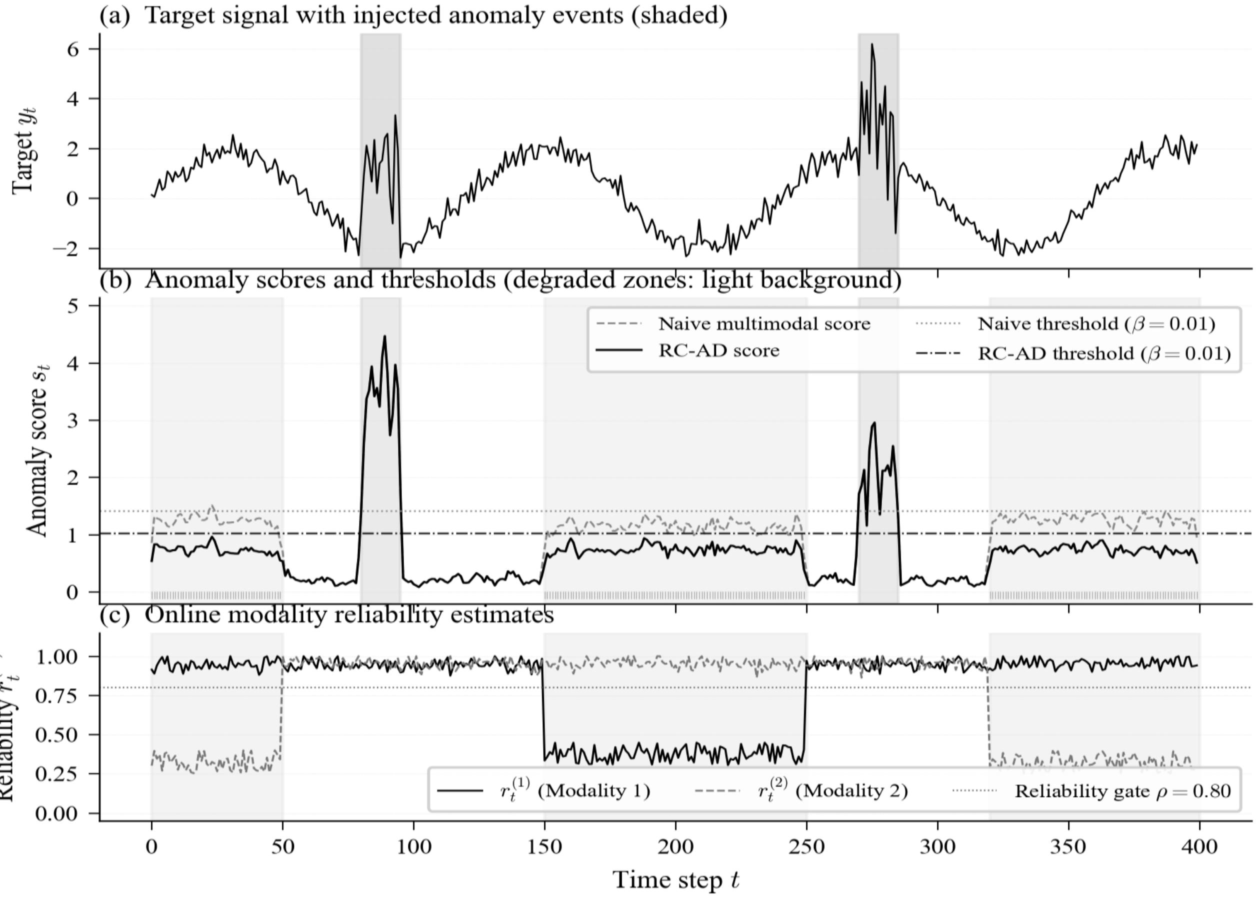

Fig. 3 - Illustrative diagnostic plot

Construct validity depends on whether injected anomalies in the target are representative of operational anomaly semantics; this work treats target-level deviations as a proxy for anomalous events that disrupt prediction, which is common in residual-based monitoring. Finally, conclusion validity may be influenced by the finite set of random seeds and by the chosen stress parameters; the paper mitigates this by reporting confidence intervals and using a fixed, disclosed seed set.

Two additional practical considerations are worth emphasizing. First, RC-AD is intentionally a control policy rather than a new forecasting backbone: it can be placed on top of different predictors, provided that modality-specific and multimodal residual scores and reliability cues can be computed causally. This modularity is useful in decision support settings where the forecasting model is fixed by prior engineering constraints, while the monitoring module must be auditable and tunable. Second, the method assumes that at least one modality remains informative often enough for reliability gating to have a meaningful choice; if data quality is chronically poor across all sources, additional data-quality remediation or a revised operational contract may be required.

In deployments, the FAR budget β is typically chosen by the operator or by downstream cost constraints, and RC-AD treats β as a first-class parameter that must remain stable across environments. In contrast, RC-AD-specific parameters are system-specific: the reliability window length, smoothing coefficient α, penalty scales k_{miss}, k_{var}, and the clean-fallback gates (δ, γ) used in (5). As a practical rule, (δ, γ) should be chosen so that the stream is classified as clean under normal operation (so early fusion is used when appropriate), while still triggering fallback under expected degradation patterns; logging c_t and w_t over time is recommended for auditing. Alarm thresholds can then be calibrated to satisfy the FAR budget using standard offline calibration on representative normal data, or online quantile calibration on recent scores when drift is expected. Table 5 summarizes qualitative properties of the compared scoring strategies, emphasizing which design elements explicitly address modality degradation.

RC-AD does not prescribe a specific reliability estimator, but practical use requires that the reliability signal is causal, stable, and aligned with the notion of «trust» used by operators.

Table 5 - Qualitative properties of the compared scoring strategies under modality degradation

Method	Multi-modal	Early fusion	Reliability cues	Degraded fallback
Single-Modal AD	-	-	-	-
Naive Multimodal	✓	✓	-	-
Equal-Weight Fusion	✓	-	-	-
RC-AD (proposed)	✓	✓	✓	✓

Since reliability affects both the score-selection switch and the clean-fallback decision, miscalibrated reliability can lead to systematic preference for the wrong modality or to overly frequent fallback to single-modality scoring. Therefore, reliability statistics should be logged and monitored over time (e.g., distribution shifts of $r_t^{(m)}$, correlation with missingness, and sensitivity to known maintenance events). If the reliability estimator is updated online, its own drift can become a failure mode; in such cases, conservative gates and periodic audits are preferable to aggressive adaptation.

RC-AD is designed for low-latency monitoring. Per time step, it requires computing residual scores (modality-specific and multimodal), computing missingness and energy penalties on a short window, updating the smoothed weight w_t, and selecting the score source according to (5)-(6). If online FAR calibration is desired in deployment, an empirical $(1-\beta)$ −quantile can be maintained on a bounded sliding buffer of recent scores; approximate quantile sketches are also applicable when memory is constrained. Table 6 summarizes a compact checklist that can be used to document the operational configuration and support auditing.

While the experiments use two modalities for clarity, reliability estimation and score fusion generalize directly to $M > 2$ modalities. Let $s_t^{(m)}$ be standardized residual scores and let $\tilde{r}_t^{(m)}$ be reliability proxies computed from degradation cues.

Table 6 - Deployment checklist for RC-AD

Item	Recommendation
Choose FAR budget β	Fix β from operational cost constraints; do not tune it post-hoc on the test stream.
Choose reliability and smoothing parameters	Set window length and $(k_{\text{miss}}, k_{\text{var}}, \alpha)$ from expected degradation dynamics; log w_t distributions.
Choose clean-fallback gates (δ, γ)	Ensure early fusion is used in normal operation but fallback triggers under expected degradation; log how often $c_t = 1$.
Threshold calibration policy	Calibrate thresholds offline on representative normal data, or maintain an online $(1-\beta)$ −quantile on recent scores (choose a bounded buffer length if used).
Reliability monitoring	Track distributions of $r_t^{(m)}$ and their relation to missingness and known degradation events.
Auditing alarms	Store (s_t, w_t, c_t, τ) to enable post-incident analysis under the FAR contract.

Normalized weights can be defined as $w_{t,m} = \tilde{r}_t^{(m)} / (\sum_{j=1}^{M} \tilde{r}_t^{(j)} + \varepsilon)$. A soft-fusion variant is $s_t = \sum_{m=1}^{M} w_{t,m}\, s_t^{(m)}$, while the winner-takes-all variant used in this paper corresponds to selecting $s_t = s_t^{(m^*)}$ for $m^* = \text{argmax}_m w_{t,m}$, optionally combined with a clean-regime multimodal score when all modalities satisfy the clean gates.

The practical behavior of RC-AD depends on how often the stream is classified as clean and how sharply the reliability proxy separates degraded from non-degraded regimes. If clean-fallback gates (δ, γ) are too strict, early fusion is rarely used and the method reduces to single-modality scoring more often than necessary. If they are too permissive, degraded data may still influence the multimodal score and degrade FAR-constrained sensitivity. An important extension is therefore a systematic sensitivity analysis of (δ, γ), reliability parameters $(k_{\text{miss}}, k_{\text{var}}, \alpha)$, and (if used in deployment) the online quantile calibration window, reporting Recall@FAR and achieved FAR as functions of these operational choices.

In this paper, reliability is treated abstractly as a probability-like confidence that a modality is currently trustworthy. Beyond missingness rates, useful reliability cues may include abrupt changes in feature distributions, non-physical sensor values, increased measurement noise, or sustained disagreement between modality-specific predictors. Another promising direction is to combine multiple lightweight cues into a single reliability score that is explicitly calibrated for decision support: the objective would be to minimize false alarms caused by data quality drops while maintaining timely detection of genuine abnormal events. Importantly, such reliability models should remain causal and auditable, since the reliability signal becomes part of the operational contract that governs which observations can trigger alarms.

Many operational settings interpret anomalies as short events rather than isolated points. While RC-AD focuses on score formation and operating-point control, a complete decision support module may also need an alert policy (cool-down windows, grouping of consecutive alarms into events, and escalation rules). Integrating simple, transparent alert policies with the FAR budget (e.g., budgeting alarms per unit time or per shift) is a natural extension that would further align the method with operator workload constraints.

5 Conclusions

This paper presented RC-AD, a reliability-constrained anomaly detection procedure for streaming multimodal time series with an explicit FAR budget. The method combines (i) standardized residual scoring (single-modality and multimodal) with (ii) online reliability estimation from degradation cues and (iii) a reliability-gated score selection policy that uses early fusion in clean regimes and falls back to the currently more reliable modality in degraded regimes. In controlled streaming benchmarks with alternating modality degradation, RC-AD improved «Recall@FAR» across FAR budgets compared to naive multimodal scoring and equal-weight fusion. A real-data demonstration on UCI Air Quality further supported that the approach can improve sensitivity under degradation when evaluation is performed on degraded segments and FAR is fixed.

Conflict of interest

The authors of this article has no conflicts of interest in writing it. There are no financial, personal, or other issues that could affect the fairness of this article. The article is entirely the researcher's own work, has not been copied from any other sources, and has never been sent to any other publication for con-sideration.

Financing

This study is completely funded by the authors' own money and there is no financial help from any organization or group. The money comes from the people who wrote the research.

Data availability

All data are available in numerical or graphical form in the main text of the research. The UCI Air Quality dataset used for the real-data demonstration is publicly available from the UCI Machine Learning Repository.

Use of artificial intelligence

The authors confirm that they used artificial intelligence technologies to search and review publicly available information and for language editing. At the same time, the finalized analysis of scientific research, conclusions and writing of the text, creation of figures and tables of this article are the exclusive results of the authors' work.

Список використаної літератури (References)

1. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey. ACM Computing Surveys. 2009. Vol. 41(3). Art. 15. DOI: 10.1145/1541880.1541882.
2. Blázquez-García A., Conde A., Mori U., Lozano J. A. A Review on Outlier/Anomaly Detection in Time Series Data. ACM Computing Surveys. 2021. Vol. 54(3). Pp. 1-33. DOI: 10.1145/3444690.
3. Pang G., Shen C., Cao L., van den Hengel A. Deep Learning for Anomaly Detection: A Review. ACM Computing Surveys. 2021. Vol. 54(2). Pp. 1-38. DOI: 10.1145/3439950.
4. Darban Z. Z., Webb G. I., Pan S., Aggarwal C. C., Salehi M. Deep Learning for Time Series Anomaly Detection: A Survey. ACM Computing Surveys. 2024. Vol. 57(1). Pp. 1-42. DOI: 10.1145/3691338.
5. Chalapathy R., Chawla S. Deep Learning for Anomaly Detection: A Survey. arXiv:1901.03407, 2019. URL: https://arxiv.org/abs/1901.03407.

6. Malhotra P., Ramakrishnan A., Anand G., Vig L., Agarwal P., Shroff G. LSTM-based Encoder-Decoder for Multi-Sensor Anomaly Detection. arXiv:1608.06154, 2016. URL: https://arxiv.org/abs/1608.06154.

7. Hundman K., Constantinou V., Laporte C., Colwell I., Söderstrom T. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. In: KDD 2018. DOI: 10.1145/3219819.3219845.

8. Su Y., Zhao Y., Niu C., Liu R., Sun W., Pei D. Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network. In: KDD 2019. DOI: 10.1145/3292500.3330672.

9. Zong B., Song Q., Min M. R., Cheng W., Lumezanu C., Cho D., Chen H. Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection. In: ICLR 2018. URL: https://openreview.net/forum?id=BJJLHbb0-.

10. Dawid A. P. Statistical Theory: The Prequential Approach. Journal of the Royal Statistical Society. Series A (General). 1984. Vol. 147(2). Pp. 278-292. DOI: 10.2307/2981683.

11. Bifet A., Gavaldà R., Holmes G., Pfahringer B. Machine Learning for Data Streams: With Practical Examples in MOA. MIT Press, 2018. ISBN: 9780262037792.

12. Aggarwal C. C. (ed.). Data Streams: Models and Algorithms. Springer, 2007. DOI: 10.1007/978-0-387-47534-9.

13. Gama J. Knowledge Discovery from Data Streams. Chapman and Hall/CRC, 2010. ISBN: 9781439826119.

14. Bifet A., Holmes G., Kirkby R., Pfahringer B. MOA: Massive Online Analysis. Journal of Machine Learning Research. 2010. Vol. 11. Pp. 1601-1604. URL: https://jmlr.org/papers/v11/bifet10a.html.

15. Gama J., Žliobaitė I., Bifet A., Pechenizkiy M., Bouchachia A. A Survey on Concept Drift Adaptation. ACM Computing Surveys. 2014. Vol. 46(4). Art. 44. DOI: 10.1145/2523813.

16.Webb G. I., Hyde R., Cao H., Nguyen H. L., Petitjean F. Characterizing Concept Drift. Data Mining and Knowledge Discovery. 2016. Vol. 30(4). Pp. 964-994. DOI: 10.1007/s10618-015-0448-4.

17. Bifet A., Gavaldà R. Learning from Time-Changing Data with Adaptive Windowing. In: SDM 2007. DOI: 10.1137/1.9781611972771.42.

18. Page E. S. Continuous Inspection Schemes. Biometrika. 1954. Vol. 41(1-2). Pp. 100-115. DOI: 10.1093/biomet/41.1-2.100.

19. Basseville M., Nikiforov I. V. Detection of Abrupt Changes: Theory and Application. Prentice Hall, 1993. ISBN: 9780131267800.

20. Fawcett T. An Introduction to ROC Analysis. Pattern Recognition Letters. 2006. Vol. 27(8). Pp. 861-874. DOI: 10.1016/j.patrec.2005.10.010.

21. Davis J., Goadrich M. H. The Relationship Between Precision-Recall and ROC Curves. In: ICML 2006. DOI: 10.1145/1143844.1143874.

22. Saito T., Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. PLOS ONE. 2015. Vol. 10(3). e0118432. DOI: 10.1371/journal.pone.0118432.

23. Baltrušaitis T., Ahuja C., Morency L.-P. Multimodal Machine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019. Vol. 41(2). Pp. 423-443. DOI: 10.1109/TPAMI.2018.2798607.

24. Lahat D., Adali T., Jutten C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. Proceedings of the IEEE. 2015. Vol. 103(9). Pp. 1449-1477. DOI: 10.1109/JPROC.2015.2460697.

25. Atrey P. K., Hossain M. A., El Saddik A., Kankanhalli M. S. Multimodal Fusion for Multimedia Analysis: A Survey. Multimedia Systems. 2010. Vol. 16. Pp. 345-379. DOI: 10.1007/s00530-010-0182-0.

26. Hall D. L., Llinas J. An Introduction to Multisensor Data Fusion. Proceedings of the IEEE. 1997. Vol. 85(1). Pp. 6-23. URL: https://ieeexplore.ieee.org/document/554205/.

27. Little R. J. A., Rubin D. B. Statistical Analysis with Missing Data. 3rd ed. Wiley, 2019. DOI: 10.1002/9781119482260.

28. UCI Machine Learning Repository. Air Quality. 2008. DOI: 10.24432/C5060Z. URL: https://archive.ics.uci.edu/dataset/387/air+quality (accessed: 07.02.2026).

29. De Vito S., Massera E., Piga M., Martinotto L., Di Francia G. On Field Calibration of an Electronic Nose for Benzene Estimation in an Urban Pollution Monitoring Scenario. Sensors and Actuators B: Chemical. 2008. Vol. 129(2). Pp. 750-757. DOI: 10.1016/j.snb.2007.09.060.

Отримано (Received) 28.04.2026
Отримано після доопрацювання (Received after revision) 07.05.2026
Прийнято (Accepted) 08.05.2026
Опубліковано (Published) 31.05.2026

Мультимодальне виявлення аномалій з обмеженням надійності для потокового аналізу даних нестаціонарних часових рядів

І. С. Узун[1], *аспірант*
ORCID: http://orcid.org/0000-0001-6619-4862; e-mail: uzun.i.s@op.edu.ua;
Scopus Author ID: 57223316393
М. В. Лобачев[1], *кандидат технічних наук, професор*
ORCID: http://orcid.org/0000-0002-4859-304X; e-mail: lobachev@op.edu.ua;
Scopus Author ID: 36845971100
[1] *Національний університет «Одеська політехніка»*

Анотація. *Системи підтримки прийняття рішень у потокових режимах часто стикаються з двома пов'язаними операційними обмеженнями: мультимодальні вхідні потоки є нестаціонарними та періодично демонструють деградацію окремих модальностей (пропуски значень, адитивний шум, зміни масштабу або підсилення сигналу), а частота хибних тривог (false alarm rate, FAR) має бути обмежена явно заданим бюджетом. У таких умовах пряме мультимодальне скорингування залишків завищує оцінки під час деградованих сегментів, зміщуючи FAR-обмежені пороги вгору та знижуючи чутливість детектора до справжніх аномальних подій.*

Мета роботи - розробити та експериментально перевірити політику скорингування залишків з обмеженням надійності модальностей, яка зберігає чутливість виявлення аномалій в умовах почергової деградації джерел при фіксованому бюджеті FAR. Задачі дослідження: формалізувати задачу мультимодального виявлення аномалій з FAR-бюджетом у каузальному режимі, побудувати онлайн-оцінювач надійності модальностей за дешевими каузальними ознаками деградації (пропуски та інфляція енергії ознак), розробити правило вибору джерела залишкової оцінки та перевірити підхід на контрольованому потоці й на реальних даних UCI Air Quality.

Запропоновано метод RC-AD - політику скорингу залишків з обмеженням надійності, яка поєднує стандартизовані оцінки залишків від модальність-специфічних та мультимодальних предикторів раннього злиття, онлайн-оцінювання надійності модальностей за пропусками й інфляцією енергії ознак у короткому каузальному вікні та правило «переможець отримує все»: у чистому вікні використовується мультимодальна оцінка раннього злиття, інакше - оцінка поточно більш надійної модальності. Метод не залежить від прогнозного ядра, є легковаговим, інтерпретовним і придатним для інтеграції в наявні моніторингові конвеєри.

Експериментальне дослідження виконано за каузальним протоколом на десяти фіксованих сідах із 95 % довірчими інтервалами: на контрольованому бенчмарку з почерговою деградацією модальностей та інжектованими подійними аномаліями RC-AD підвищує Recall@FAR з 0,103 до 0,335 при FAR 0,05 та з 0,182 до 0,426 при FAR 0,10, перевершуючи наївну мультимодальну, рівноважну (equal-weight) та одномодальні базові лінії; демонстрація на UCI Air Quality при FAR 0,01 підтверджує цю тенденцію відносно мультимодальних базових ліній. Наукова новизна полягає у формулюванні політики мультимодального виявлення аномалій з явним онлайн-обмеженням надійності, яка вперше поєднує гейтинг ранньої фузії за каузальними ознаками деградації з відбором найбільш надійної модальності при заданому FAR-бюджеті. Практична значущість підтверджена відтворюваним покращенням Recall@FAR на контрольованих та реальних сенсорних даних, а також легковаговістю й інтерпретовністю методу, придатного для аудитованих систем підтримки прийняття рішень.

Ключові слова: *машинне навчання, аналіз даних, інформаційні системи, системи підтримки прийняття рішень, мультимодальні часові ряди, нестаціонарні часові ряди, потокове виявлення аномалій, деградація якості даних, оцінювання надійності модальностей, бюджет хибних тривог.*

Цитування статті: Узун І. С., Лобачев М. В. (2026). Мультимодальне виявлення аномалій з обмеженням надійності для потокового аналізу даних нестаціонарних часових рядів. *Електротехнічні та комп'ютерні системи*, 2026, 46(122), с.107-119. doi:https://doi.org/10.15276/eltecs.46.122.2026.10

About the authors (Про авторів)

Illia Uzun, Postgraduate Student of the Department of Artificial Intelligence and Data Analysis, Institute of Artificial Intelligence and Robotics, Senior Lecturer, Odesa Polytechnic National University; 1, Shevchenko Ave., Odesa, 65044, Ukraine.
E-mail: uzun.illia.main@gmail.com; ph.: +380679172040

Узун Ілля Святославович, аспірант кафедри Штучного Інтелекту та Аналізу Даних, Інститут Штучного Інтелекту та Робототехніки, старший викладач, Національний університет «Одеська політехніка»; просп. Шевченка, 1, Одеса, Україна
E-mail: uzun.illia.main@gmail.com, тел. +380679172040

ORCID: http://orcid.org/0000-0001-6619-4862 Scopus Author ID: 57223316393

Mykhaylo Lobachev, The Director of the Artificial Intelligence and Robotics Institute, Ph.D, Professor of the Department of Artificial Intelligence and Data Analysis, Odesa Polytechnic National University; Shevchenko av., 1, Odessa,
E-mail: lobachev@op.edu.ua, ph. +380952788602

Лобачев Михайло Вікторович, директор Інституту Штучного Інтелекту та Робототехніки, к.т.н., професор кафедри Штучного Інтелекту та Аналізу Даних, Національний університет «Одеська політехніка»; просп. Шевченка, 1, Одеса, Україна.
E-mail: lobachev@op.edu.ua, тел. +380952788602

ORCID: http://orcid.org/0000-0002-4859-304X Scopus Author ID: 36845971100