Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU32749

Diagnosing Long-tail Bottlenecks in Business Processes: An Integrated Framework of Process Mining and Multi-dimensional Anomaly Detection

Lin Shi

Faculty of Science and Technology, Beijing Normal-Hong Kong Baptist University, Zhuhai, China
t330026136@mail.bnbu.edu.cn

Abstract. Traditional process bottleneck analysis, relying on frequency-based metrics, often fails to capture "long-tail" anomalies—low-frequency deviations with high business impact. To address this, we propose a novel framework that integrates process mining with multi-dimensional anomaly detection. Our approach constructs a Petri net via heuristic mining to model the control-flow and quantifies anomalous behavioral patterns using Isolation Forest (IF). To ensure rigorous evaluation, we innovatively employ case-level GroupKFold cross-validation and feature ablation to mitigate data leakage. Validated on the BPI 2017 dataset (2,000 cases), our discovered process model achieves a high fitness of 0.9451 via token replay. Crucially, our ablation study reveals a 6% performance inflation caused by label leakage, which our rigorous methodology successfully strips away, yielding a robust predictive model with an F1-score of 0.9394 and a ROC-AUC of 0.9660 using only numerical features. This work expands bottleneck identification from a univariate statistical view to a multi-dimensional behavioral spectrum, providing a leakage-proof and highly generalizable benchmark for diagnosing complex operational risks in real-world scenarios.

Keywords: Process Mining, Anomaly Detection, Bottleneck Diagnosis, Isolation Forest, Data Leakage

1. Introduction

Enterprise digital transformation relies heavily on event logs to enhance operational efficiency. Traditional process mining focuses on control-flow analysis and frequency-based metrics, such as mean duration, to identify bottlenecks. However, these methods frequently overlook "long-tail" anomalies—low-frequency yet high-impact deviations that can incur significant costs due to rework, delays, or compliance issues. While anomaly detection algorithms such as Isolation Forest (IF) [1] and Local Outlier Factor (LOF) [2] are effective at identifying high-dimensional outliers, mapping these statistical outliers back to interpretable process structures (e.g., Petri nets) remains a major challenge. Furthermore, current research often struggles with this semantic association and lacks rigorous validation methods to prevent data leakage in subsequent predictive models, which in turn inflates performance estimates [3].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU32749

To bridge these gaps, this paper introduces an integrated framework that synergizes process discovery with anomaly detection. Our approach utilizes heuristic mining to construct a Petri net, extracts multi-dimensional features to quantify anomalies with IF, and employs case-level GroupKFold validation to ensure reliable and unbiased bottleneck prediction. By synthesizing the structural insights of process models with the sensitivity of anomaly detectors, this study establishes a robust, leakage-proof methodological benchmark for identifying and monitoring complex operational bottlenecks in real-world business scenarios.

2. Methodology

This paper proposes a bottleneck diagnosis and prediction framework that integrates process mining with anomaly detection. The framework consists of four core modules: multi-dimensional feature extraction, process discovery, anomaly quantification, and replay validation.

2.1. Event log and feature space

Definitions: Let $\mathscr{A}$ be a finite set of activities. An event $e_i = (a_i, t_i, c_i, r_i, \ldots)$ is a tuple containing activity ($a_i \in \mathscr{A}$) , timestamp, Case ID, and resource. A trace $\sigma = \langle e_1, e_2, \ldots, e_n \rangle$ is an ordered sequence of events belonging to the same case, and an event log $L = \{\sigma_1, \sigma_2, \ldots, \sigma_m\}$ is a set of these traces.

Feature Mapping Φ : To capture spatio-temporal anomalies, define a mapping function $\Phi : \mathscr{E} \to \mathbb{R}^d$. The feature space covers temporal dimensions (duration and intervals) and structural dimensions (case-level event counts and total timespan).

Robustness: Given the extreme long-tail distribution of temporal attributes, non-linear smoothing $\log(1 + |\cdot|)$ is applied to stabilize the feature variance.

2.2. Process discovery and petri nets

The Heuristics Miner algorithm [4,5] is employed to extract the underlying business logic and represent it as a Petri net $SN = (P, T, F)$, where:

P is a finite set of places;

T is a finite set of transitions, where visible transitions correspond to specific activities in the activity set $\mathscr{A}$, also including invisible transitions (τ -transitions) generated by the heuristic mining algorithm;

$F \subseteq (P \times T) \cup (T \times P)$ is the set of flow relations.

Additionally, a Directly-Follows Graph (DFG) is derived to provide an explicit path structure for subsequent bottleneck localization [5].

2.3. Anomaly scoring based on Isolation Forest (IF)

The Isolation Forest algorithm is introduced to calculate anomaly scores $s(x, n)$ for each observation [1]:

$$s\left(x, n\right) = 2^{-\frac{E(h(x))}{c(n)}} \tag{1}$$

DOI: 10.54254/2755-2721/2026.GU32749

where $h(x)$ is the path length in an isolation tree, $E(h(x))$ is the expected path length of the sample across the entire forest of isolation trees; and $c(n)$ is the average path length of an unsuccessful search in a Binary Search Tree (BST), which is used for normalization.

The score satisfies $s(x,n) \in [0,1]$. As $s \to 1$, the event is classified as an anomalous bottleneck candidate, allowing the framework to capture "low-frequency but high-impact" deviations.

2.4. Replay validation and model evaluation

Token Replay is employed to evaluate the alignment between Petri nets and logs. Fitness measures the model's coverage of observed behaviors, supplemented by Precision to prevent overgeneralization [5].

3. Experiments

3.1. Dataset and ground truth construction

The study utilizes the BPI Challenge 2017 event log [6]. To balance computational feasibility, the first 2,000 cases (76,888 event records) were sampled.

Proxy Ground Truth (GT): Since the log lacks explicit annotations, a heuristic GT was constructed by merging the Top-5 activities by frequency and the Top-5 by mean duration, resulting in 9 confirmed bottleneck labels.

Evaluation Intent: The evaluation verifies whether the framework can identify traditional bottlenecks while sensitively locating "long-tail" latent risks that are missed by statistical features.

3.2. Experimental setup and validation strategy

Feature Engineering: A 9-dimensional vector was constructed, covering temporal (5 dimensions), structural (2 dimensions), and categorical (Activity/Resource names) features.

Leakage Control: Two split strategies were used:

(1) Event-level Split: Randomly split into 80/20 ratios as a control group for quantifying performance inflation.

(2) Case-level GroupKFold ($K = 5$): Group by Case ID to ensure all events of a single case remain in the same fold to block information penetration.

Ablation Study: Dual feature groups were established: "Full Features" (including categorical names) and "Pure Numerical Features" (excluding names that might carry label attributes).

4. Results and discussion

4.1. Process model discovery and visualization analysis

The Heuristics Miner discovered a model with 25 activity nodes.

DOI: 10.54254/2755-2721/2026.GU32749

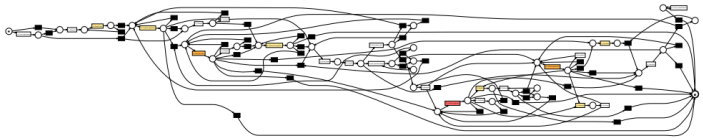


Figure 1. Bottleneck visualization of the process (red = most severe bottleneck, orange = secondary bottleneck, yellow = potential bottleneck)

Figure 1 presents the bottleneck-highlighted Petri net, where nodes are colored based on traditional mean duration and frequency-based metrics.

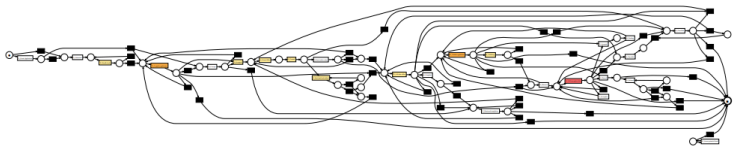


Figure 2. Visualization of anomalies detected by Isolation Forest

Figure 2 overlays the results from the Isolation Forest (IF) detection, mapping nodes with high anomaly occurrences to the graph using gradient colors.

Comparison reveals that the primary risk areas identified by anomaly detection do not completely overlap with traditional high-frequency bottlenecks. Traditional metrics highlight W_Validate application as the most severe bottleneck. However, IF detection identifies W_Call incomplete files as the top anomaly risk.

This indicates that multi-dimensional anomaly scoring can reveal potential risks characterized by low execution frequency but unusual behavioral patterns, complementing traditional frequency-based bottleneck analysis [7].

4.2. Complementarity of IF and LOF

Table 1. Baseline bottleneck identification comparison (Top-k = 5)

Method	Precision	Recall	F1	Pred/GT
Frequency (Top-k)	1.0000	0.5556	0.7143	5/9
Mean-Duration Top-k	1.0000	0.5556	0.7143	5/9
LOF Top-k	0.8000	0.4444	0.5714	5/9
Isolation Forest (Top-k)	0.8000	0.4444	0.5714	5/9

As shown in Table 1, traditional frequency-based methods achieved extremely high precision (1.0000) in proxy ground truth evaluation. However, constrained by the Top-5 setting and the total of 9 pseudo-labels, their recall (0.5556) essentially reached the mathematical upper bound defined by this evaluation framework.

In contrast, while the recall of Isolation Forest (IF) and Local Outlier Factor (LOF) was only 0.4444, this precisely reflects the design intent of anomaly detection algorithms—they tend to locate outlier samples in a multi-dimensional feature space rather than simply fitting one-dimensional

DOI: 10.54254/2755-2721/2026.GU32749

high-frequency delays, thereby providing a complementary structural perspective for process diagnosis [3].

While Isolation Forest (IF) and Local Outlier Factor (LOF) show similar macro-metrics (F1=0.571), their Jaccard similarity is only 0.011. This reveals they are almost entirely mutually exclusive at the micro-event level, with IF capturing global outliers and LOF focusing on local density shifts. This phenomenon of "macro-metric equivalence combined with micro-level near-total mutual exclusivity" profoundly reveals the strong complementarity of the two algorithms in capturing anomaly patterns within the multi-dimensional process feature space. The underlying reason lies in their distinct operating principles. IF isolates anomalies by randomly partitioning the feature space, making it exceptionally effective at detecting global outliers—points that are aberrant relative to the entire dataset. In contrast, LOF assesses the local density deviation of a point, making it sensitive to local outliers—points that are anomalous within their immediate neighborhood but might not be extreme globally. This fundamental difference explains the Jaccard similarity of only 0.011; they are capturing two orthogonal yet equally important dimensions of process anomalies, thereby providing a holistic view of process health when used in tandem.

4.3. Case study of typical abnormal traces

To concretize the advantages of multi-dimensional anomaly detection, this study extracted two typical latent high-risk traces for in-depth analysis:

Case 1 (Application 1002485344) (Car loan, ¥16,000): Lasted 95.4 days with a total of 80 events. The activity W_Call incomplete files (request for missing documents) was executed 35 times, and the case was eventually canceled due to incomplete documentation. The case timespan was z = +4.76, and the event count was z = +2.29.

Case 2 (Application 2037628374) (Home residual debt loan, ¥18,000): Lasted 133.7 days with 180 events (the highest event count in the dataset). W_Call incomplete files was executed 97 times, and W_Validate application was executed 30 times, together accounting for 71% of the total events, with the case eventually ending in multiple canceled offers. The case timespan was z = +7.25, and the event count was z = +8.00.

Traditional frequency-based metrics fail to distinguish between standard business interactions and extreme loops—such as the 35 to 97 repetitions observed in these cases—because they average durations across all 10,916 events. In contrast, the Isolation Forest utilizes multi-dimensional joint analysis of $\Phi_{IF}(e_i)$ to sensitively identify these long-tail deviations as top-priority bottlenecks. By capturing atypical events that are rare and behaviorally distinct in the multi-dimensional feature space, this framework complements unidimensional statistics and helps uncover high-risk patterns for complex process diagnosis [1].

4.4. Predictive performance comparison

The table 2 below presents a consolidated comparison of the performance of full-featured models (Experimental Group 1) versus purely numerical models (Experimental Group 2) under different strategies.

DOI: 10.54254/2755-2721/2026.GU32749

Table 2. Supervised bottleneck prediction comparison

Experiment Group	Split Strategy	Optimal Model	P(+)	R(+)	F1(+)	ROC-AUC
Group 1: Full Features (Including Activity Names)	Event-level	Random Forest	0.9997	1.0000	0.9999	1.0000
Group 1: Full Features (Including Activity Names)	GroupKFold	Random Forest	0.9996	1.0000	0.9998	1.0000
Group 2: Pure Numerical (Spatio-temporal Only)	Event-level	Random Forest	0.9259	0.9618	0.9435	0.9684
Group 2: Pure Numerical (Spatio-temporal Only)	GroupKFold	Random Forest	0.9232	0.9562	0.9394	0.9660

Information Leakage: Group 1 exhibits near-perfect scores because activity names directly map to the label generation rules. The 6% F1 gap between Group 1 and Group 2 quantifies the performance "bubble" caused by label leakage.

Generalization: Under the rigorous GroupKFold with pure numerical features, Random Forest achieves an F1 of 0.9394, proving the framework can predict hidden bottlenecks by relying solely on objective temporal structures.

4.5. Model quality replay validation

To ensure all inferences are based on valid process logic, this study performed token-based replay conformance checking [8].

Table 3. Replay validation results

Metric	Value
Log Fitness	0.9451
Average Trace Fitness	0.9367
Precision	0.7441

Replay validation results(table 3) show that the model possesses high log fitting capability ($\text{Log Fitness} = 0.9451, \text{Precision} = 0.7441$). Compared to smaller samples (e.g., only 100 cases, $\text{Fitness} =$ 0.9472, Precision = 0.7169), the fitness of the 2,000-case subset decreased slightly while precision steadily improved, proving that the algorithm successfully pruned over-generalized false paths while retaining true variants.

4.6. Comprehensive discussion

In summary, by integrating multi-dimensional anomalous features with Petri net topology, this framework achieves a leap in process optimization from "ex-post statistics" to "pattern recognition":

Unsupervised Diagnosis Constructs a "Global-Local" Management Matrix:

The complementarity between IF and LOF precisely maps to two distinct management pain points in real-world scenarios. It empowers managers to distinguish between "systemic mechanism defects (requiring top-down restructuring)" and "operational-level local friction (requiring bottom-up intervention)," thereby avoiding mismatched management interventions.

Supervised Prediction Establishes a Real-time Monitoring Paradigm:

In a numerical feature space strictly avoiding information leakage, the robust generalization of Random Forest ($\text{F1} = 0.9394$) proves that efficient prediction of hidden bottlenecks can be achieved by relying solely on objective temporal and aggregated structures. This result provides a reliable methodological foundation for enterprises to deploy adaptive process risk control systems.

DOI: 10.54254/2755-2721/2026.GU32749

5. Conclusion

This study introduces a novel integrated framework that combines process mining with anomaly detection to effectively diagnose "long-tail" bottlenecks overlooked by traditional statistical methods. Through an empirical analysis of 2,000 cases (76,888 events) from the BPI 2017 dataset, we demonstrate the framework's efficacy. Key findings reveal a significant complementarity between Isolation Forest (IF) and Local Outlier Factor (LOF), which exhibit extreme micro-level orthogonality (Jaccard similarity of 0.011), capturing distinct and valuable facets of process anomalies. Our approach is built on a solid foundation, with the underlying Petri net achieving a Log Fitness of 0.9451 and a Precision of 0.7441. Crucially, by employing rigorous case-level GroupKFold validation and feature ablation, we quantified and eliminated a 6% performance "bubble" caused by data leakage, yielding a robust predictive model (Random Forest) with an F1-score of 0.9394 using only numerical features. This work not only extends the definition of process bottlenecks into a multi-dimensional behavioral spectrum but also establishes a rigorous, leakage-proof benchmark for future research. Subsequent work will focus on integrating Graph Neural Networks (GNNs) for non-linear representation learning on process graphs, implementing real-time stream mining to address concept drift, and enhancing business transparency through Explainable AI (XAI), and incorporating expert priors with external KPI data to validate the framework's economic value in real-world decision-making.

References

[1] Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2012). Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data, 6(1), Article 3.
[2] Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 93–104.
[3] Peeperkorn, J., vanden Broucke, S., & De Weerdt, J. (2024). Validation set sampling strategies for predictive process monitoring. Information Systems, 121, 102330.
[4] Weijters, A. J. M. M., van der Aalst, W. M. P., & Alves de Medeiros, A. K. (2006). Process mining with the HeuristicsMiner algorithm (BETA Working Paper Series No. 166). Eindhoven University of Technology.
[5] Leemans, S. J. J., Poppe, E., & Wynn, M. T. (2019). Directly follows-based process mining: Exploration & a case study. In 2019 International Conference on Process Mining (ICPM) (pp. 25–32). IEEE.
[6] van Dongen, B. (2017). BPI Challenge 2017 [Data set]. Eindhoven University of Technology.
[7] Ko, J., & Comuzzi, M. (2023). A systematic review of anomaly detection for business process event logs. Business & Information Systems Engineering, 65(4), 441–462.
[8] Rozinat, A., & van der Aalst, W. M. P. (2008). Conformance checking of processes based on monitoring real behavior. Information Systems, 33(1), 64–95.