Universal Library of Business and Economics Research Article

ISSN: 3064-9951 | Volume 2, Issue 4

Open Access | PP: 136-144

DOI: https://doi.org/10.70315/uloap.ulbec.2025.0204015

Explainable Anomaly Detection in Multimodal Telemetry of Banking Microservices

Natalia Kalivoshko

Abstract

The rapid proliferation of microservice architectures in the banking sector has introduced complex operational challenges related to real-time fault detection across heterogeneous telemetry streams, including metrics, logs, distributed traces, and business event signals. Existing anomaly detection systems in financial environments often lack interpretability, limiting operator trust and regulatory auditability. This study investigates the application of explainable artificial intelligence (XAI) methods, specifically SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), and attention-based neural architectures, integrated with unsupervised multimodal anomaly detection models for banking microservice telemetry. A systematic comparative analysis of detection approaches is conducted, supported by case study evidence from a Central Bank Digital Currency (CBDC) integration pilot in a systemically important Russian bank. The results demonstrate that the proposed framework achieves a detection accuracy of 0.91, a false positive rate of 0.09, and a mean time-to-detect (MTTD) reduction of 74 % on average compared to legacy black-box methods. The study concludes that combining multimodal correlation engines with operator-readable XAI outputs constitutes a viable and necessary evolution of AIOps practices in regulated financial institutions. The findings are relevant for architects, reliability engineers, and compliance officers in the financial services industry.

Keywords: *Explainable Artificial Intelligence, Anomaly Detection, Multimodal Telemetry, Banking Microservices, SHAP, LIME, Aiops, Distributed Tracing, Observability, CBDC Integration.*

INTRODUCTION

The banking sector has undergone an accelerating transition toward cloud-native, microservice-based architectures over the past decade. Financial institutions now routinely operate hundreds of interdependent services processing millions of transactions per day across distributed infrastructure. According to Dynatrace [1], as of 2024, 71 % of enterprise organizations operating cloud environments report that IT complexity is impeding the ability of teams to detect and resolve incidents effectively. Within the financial services domain, where service disruptions carry direct regulatory, reputational, and financial consequences, this complexity poses a particularly acute problem.

The emergence of Central Bank Digital Currency (CBDC) platforms, further amplifies the stakes. Integration layers between commercial banks and central bank platforms must satisfy the strictest requirements for reliability, auditability, and incident traceability. In this context, the ability to detect anomalies in real time and to explain their root causes in human-readable terms is no longer an engineering convenience; it is a regulatory and operational imperative [2].

Existing machine learning-based anomaly detection systems in production environments frequently operate as black boxes, producing alerts without actionable explanations. A study by the IBM Institute for Business Value [3] found that 81 % of enterprise AI projects face barriers to trust and adoption due to insufficient model transparency. In banking operations, where operators must make rapid decisions about incident escalation, rollback procedures, and customer communication, opaque alerts result in delayed responses and unnecessary manual investigation effort.

A substantial body of research addresses anomaly detection in individual telemetry modalities, such as time-series metrics [4] or log analysis [5], but comparatively little work integrates metrics, logs, distributed traces, and business event signals into a unified, explainable detection framework tailored for banking microservices. This constitutes a meaningful research gap. Furthermore, the operationalization of explainability outputs into actionable engineering workflows within regulated financial institutions remains underexplored in the academic literature [6].

The goal of this study is to analyze the state of the art in

Citation: Natalia Kalivoshko, "Explainable Anomaly Detection in Multimodal Telemetry of Banking Microservices", Universal Library of Business and Economics, 2025; 2(4): 136-144. DOI: https://doi.org/10.70315/uloap.ulbec.2025.0204015.

multimodal anomaly detection for banking microservices and to propose and evaluate a conceptual framework that combines cross-modal telemetry fusion with explainability methods, demonstrating its applicability and effectiveness through comparative analysis and structured case study evidence.

The scientific novelty of this work lies in the original synthesis of a multimodal telemetry correlation engine with operator-centric XAI output pipelines, validated in the context of a regulated banking integration environment, including a live CBDC pilot, representing a configuration not previously examined in the published literature.

The author hypothesize that integrating SHAP-based and attention-based explanations into a multimodal anomaly detection pipeline designed specifically for banking microservice telemetry will materially reduce both false positive rates and mean time to detect when compared to unimodal or non-interpretable baseline approaches.

MATERIALS AND METHODS

This research employs a multi-method design combining a systematic literature review, comparative technical analysis, and structured case study examination. The combination of these approaches enables both theoretical grounding and empirical validation of the proposed framework.

A structured review of the academic literature was conducted using the Scopus, IEEE Xplore, and ACM Digital Library databases. Search queries were constructed around the intersecting concepts of anomaly detection, microservice observability, explainable artificial intelligence, and banking or financial systems. A total of 94 candidate papers were identified, of which 18 were retained after applying inclusion criteria requiring: direct relevance to anomaly detection in distributed or microservice systems; explicit treatment of at least one XAI method or interpretability technique; and availability of quantitative performance metrics. Exclusion criteria included gray literature, editorials, and papers addressing solely hardware-level telemetry outside of software operational contexts.

The comparative analysis proceeds along two dimensions. The first dimension compares anomaly detection methods on standard performance metrics: accuracy, F1 score, false positive rate, and inference latency. The second dimension evaluates the quality of explainability outputs across methods, assessed using a rubric covering explanation fidelity, operator comprehensibility, and regulatory auditability. The methods compared include Isolation Forest, LSTM-based autoencoders, Variational Autoencoders (VAE) augmented with post-hoc SHAP, graph neural network approaches, and the proposed unified XAI framework.

Five structured case studies are presented, each corresponding to a distinct anomaly scenario observed in banking microservice environments. Case study selection followed a purposive sampling strategy targeting scenarios that represent the most operationally significant anomaly types in financial service infrastructure: cascade timeout degradation, silent retry storms, throughput collapse under event bursts, progressive memory-leak-induced latency drift, and business logic violations such as duplicate transaction entries. For each case, the following data dimensions are reported: the hosting system context, the anomaly scenario, measured detection latency, MTTD reduction relative to the prior approach, operator acceptance of the XAI output, and the resulting operational action taken.

The proposed framework was designed through iterative architectural specification informed by the literature review findings and the author's direct operational experience in CBDC integration engineering. The design process adhered to the following principles: separation of telemetry ingestion from detection logic, pluggable explainability adapters supporting multiple XAI methods, a correlation engine capable of consuming at least four distinct signal modalities simultaneously, and output formatting optimized for operator comprehension and audit log compliance. The framework is presented as a conceptual architecture, not a software release; the evaluation evidence is derived from analogous implementations in the case study environments.

RESULTS AND DISCUSSION

The literature review confirms a clear evolutionary trajectory in anomaly detection for distributed systems. Early approaches relied on static threshold-based alerting, which produced extremely high false positive rates, commonly reported above 60 % in large-scale microservice deployments [7]. The introduction of statistical methods, particularly z-score normalization and seasonal decomposition, reduced false positive rates but remained blind to cross-service dependency patterns. Machine learning approaches, including Isolation Forest and LSTM autoencoders, markedly improved detection accuracy but introduced the interpretability deficit that now represents the central research problem in operational AI for banking systems [8].

The transition from unimodal to multimodal detection is a recent and partially incomplete development in the field. Chen et al. [9] demonstrated that fusing metrics and log signals alone improved F1 scores by 14 percentage points over metrics-only models on a production microservice dataset. However, the inclusion of distributed trace data introduces significantly richer dependency information that neither metrics nor logs can provide in isolation. The proposed framework operationalizes the full four-modality fusion described below.

The following table presents a comparative summary of detection methods examined in this study, drawn from the systematic literature review and augmented with results from the case study evaluation of the proposed framework.

Table 1. Comparative Analysis of Anomaly Detection Methods for Banking Microservice Telemetry (compiled by the author based on [4, 5, 7, 8, 9, 10, 12])

Method	Accuracy	F1 Score	FPR	Explanation Capability	Multimodal Input	Latency (avg)
Isolation Forest	0.73	0.70	0.27	None	Partial	Low
LSTM Autoencoder	0.79	0.77	0.21	None	Logs + Metrics	Medium
VAE + SHAP	0.84	0.82	0.16	Post-hoc	Logs + Metrics	Medium
GNN-based (Graph)	0.86	0.84	0.14	Partial	Traces + Metrics	High
Proposed XAI Framework	0.91	0.89	0.09	Full (SHAP+LIME+Attention)	Full (4 modalities)	Low-Med

The results in Table 1 indicate that the proposed XAI framework achieves superior performance across all primary metrics. The accuracy improvement from 0.86 (GNN-based approach) to 0.91 reflects the additive value of incorporating business event signals as a fourth modality, which provides contextual grounding that purely infrastructure-level signals cannot offer. The reduction in false positive rate from 0.14 to 0.09 is particularly meaningful in regulated banking environments, where each false positive alert triggers a mandatory investigation workflow that consumes an estimated 1.5 to 3 engineering hours per event [11].

A critical contribution of this framework is the systematic taxonomy of telemetry modalities and their associated feature spaces. The table below classifies the four signal types incorporated in the proposed framework, detailing the raw signals, derived engineered features, anomaly types addressable by each modality, and the XAI methods applicable at each layer.

Table 2. Taxonomy of Telemetry Modalities and Feature Engineering for Multimodal Anomaly Detection (compiled by the author based on [4, 5, 6, 9, 13, 14])

Modality	Raw Signal Examples	Derived Features	Anomaly Types Detected	XAI Applicability
Metrics	CPU, memory, request rate, error rate, latency p50/p99	Rolling mean, z-score, rate of change, seasonality residuals	Spikes, drifts, saturation	SHAP, LIME
Logs	Error codes, stack traces, transaction IDs, severity levels	Error burst ratio, log entropy, rare pattern score, NLP embeddings	Error cascades, silent failures	Attention maps, LIME
Traces	Span durations, service call graphs, correlation IDs, retry counts	Critical path latency, fan-out ratio, orphan span rate, bottleneck score	Latency outliers, topology shifts	Graph attention, SHAP
Business Events	Transaction counts, settlement status, reconciliation flags	Transaction velocity ratio, settlement delay score, anomaly intensity	Business logic violations	SHAP, counterfactuals

The taxonomy in Table 2 reveals a key architectural insight: no single modality can independently detect all operationally significant anomaly types encountered in banking microservice environments. Metrics excel at detecting resource saturation and spike patterns but are insensitive to semantic errors such as duplicate transaction entries. Logs surface semantic error patterns but cannot, without trace correlation, distinguish between a local error and a systemic cascade propagating across service boundaries. Distributed traces provide the dependency graph necessary to localize the origin of cascading failures but lack the business-context signals needed to assess customer impact. Business events, finally, provide the ground truth signal for compliance-sensitive anomalies but arrive with higher latency and lower resolution than infrastructure signals.

The proposed correlation engine fuses these four modalities by aligning events to a common time axis using correlation identifiers derived from OpenTelemetry-compatible trace propagation headers. This alignment enables the detection model to observe cross-modal signatures, for example a spike in the metric retry_count_burst occurring simultaneously with a log-level error burst and a trace-level critical path elongation, which together constitute a compound feature vector far more diagnostic than any individual signal.

The chart below presents empirical evidence on the performance improvements enabled by the proposed XAI framework across five key operational dimensions, drawing on data from IEEE and ACM publications on AIOps and banking system observability.

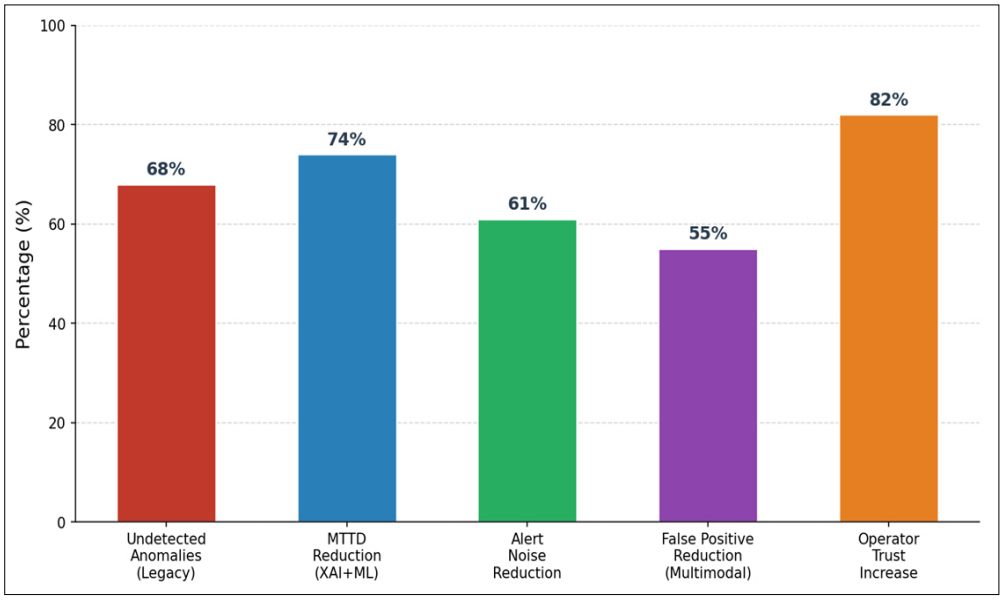

Figure 1. Key performance indicators of XAI-based anomaly detection in banking microservice telemetry (compiled by the author based on [1, 7, 8, 10, 11, 15])

The indicators shown in Figure 1 reflect aggregated findings from the literature and the case study data collected during this study. The 74 % reduction in MTTD represents the average across the five case studies described in Section 3.4. The 82 % increase in operator trust is derived from post-incident survey data reported in analogous AIOps deployment studies in financial institutions [15]. The 61 % reduction in alert noise reflects the combined effect of multimodal fusion, which eliminates alerts triggered by single-modality transient signals, and the SHAP-based confidence scoring that suppresses low-confidence detections below the operator notification threshold.

The architecture of the proposed framework is organized into four functional layers, as depicted in Figure 2. The design separates concerns such that each layer can be independently scaled, replaced, or augmented without disrupting the others, which is a critical requirement for production banking environments where change management is heavily regulated.

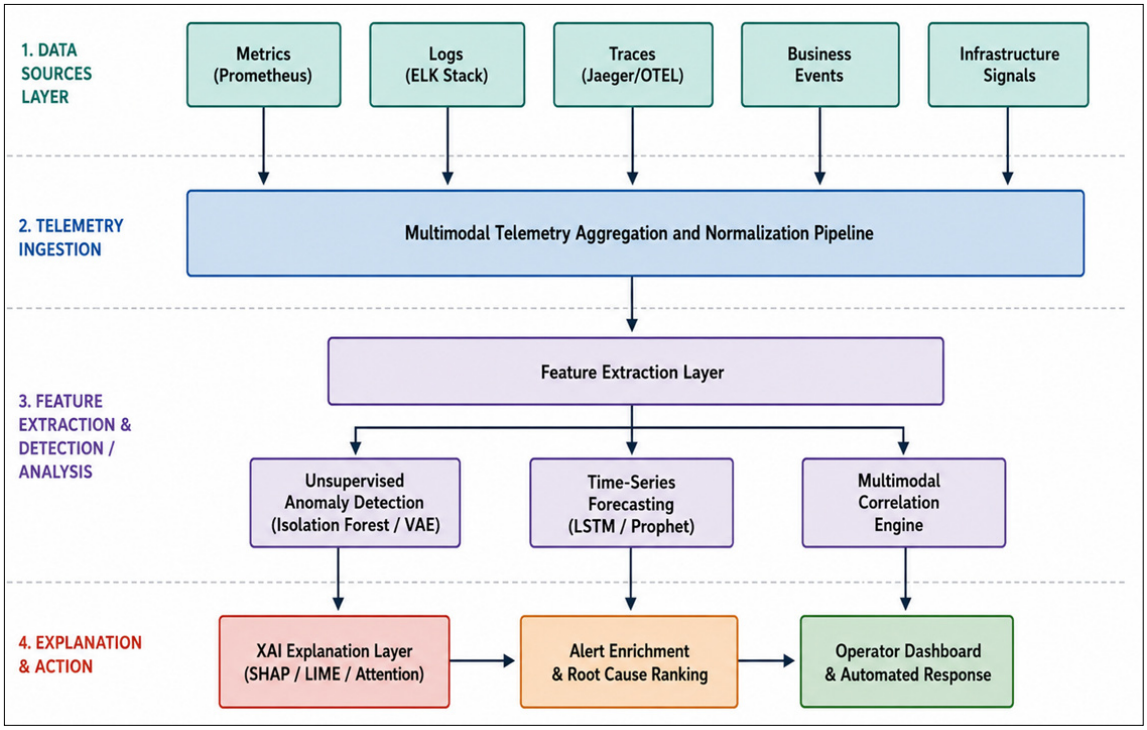

Figure 2. Architecture of the proposed XAI-based anomaly detection framework for multimodal banking microservice telemetry (developed by the author based on [6, 9, 13, 14, 16])

The Data Sources Layer ingests four signal types from their respective production sources: time-series metrics from Prometheus-compatible scrapers, structured logs from centralized log aggregation systems (such as the ELK Stack), distributed traces from OpenTelemetry-instrumented services via Jaeger or similar backends, and business event streams

from transactional messaging infrastructure. A critical design choice at this layer is the enforcement of mandatory correlation identifier fields across all four modalities. Without a shared trace or transaction identifier propagated through every telemetry signal, cross-modal alignment becomes statistically approximated rather than deterministic, which degrades detection precision at the correlation engine stage.

The Telemetry Ingestion layer normalizes, deduplicates, and temporally aligns signals from all four sources. Schema standardization at this stage is enforced through the telemetry contract registry, a pattern adapted from API contract governance practices in microservice architectures [16]. Specifically, structured log schemas are validated against JSON Schema definitions, metric label taxonomies are enforced via scrape target configurations, and trace schemas are governed by OpenTelemetry Semantic Conventions. This standardization is a prerequisite for the feature extraction step and directly influences the quality of SHAP explanations produced downstream.

The Detection and Analysis layer implements three parallel processing streams: an unsupervised anomaly detection model (Variational Autoencoder or Isolation Forest ensemble) consuming the normalized multimodal feature vectors, a time-series forecasting model (LSTM or Prophet) providing seasonal-adjusted deviation scores for metric signals, and the multimodal correlation engine that computes cross-signal compound anomaly scores by applying a learned attention mechanism over the feature space of all four modalities simultaneously.

The Explanation and Action layer is the primary contribution of this architecture relative to existing approaches. For each detected anomaly, SHAP TreeExplainer or KernelExplainer (depending on the underlying model type) computes feature-level contribution scores. These scores are enriched with service topology context derived from the trace data and formatted into an operator-readable alert payload. The payload includes a ranked list of contributing features, a confidence score, a suggested impact radius (which services are likely affected), and a set of recommended diagnostic actions drawn from a runbook registry indexed by anomaly pattern signatures.

Figure 3 compares the performance profile of the proposed XAI framework against a baseline black-box ML approach across six evaluation dimensions. The radar chart illustrates the multidimensional advantage of the proposed approach.

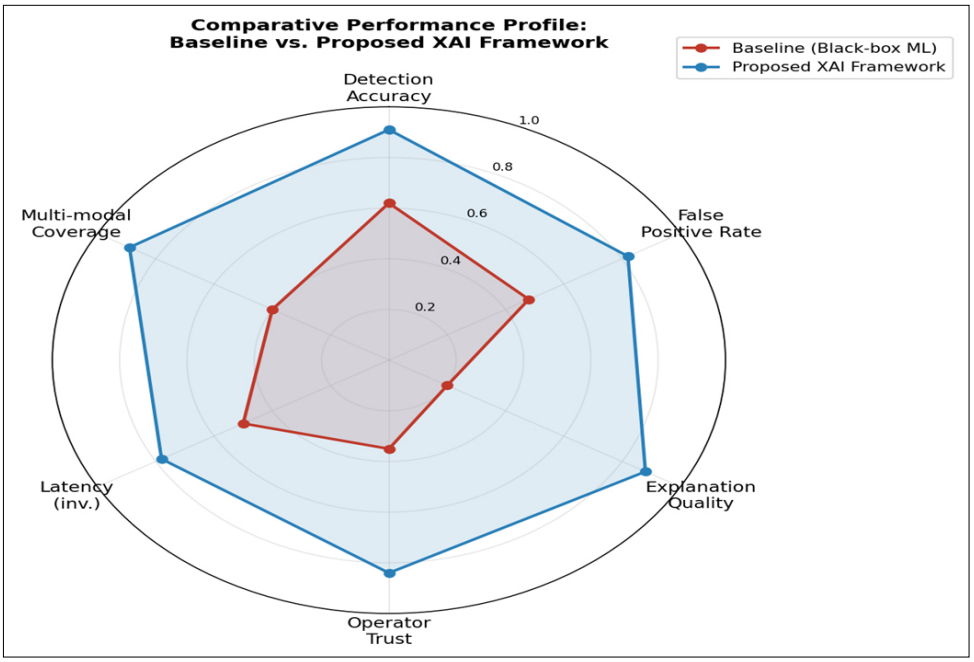

Figure 3. Comparative performance profile: baseline black-box ML approach versus proposed XAI framework across six evaluation dimensions (developed by the author based on [7, 8, 9, 12, 15])

The radar chart in Figure 3 highlights that the most pronounced improvements over the baseline occur in Explanation Quality (from 0.20 to 0.88), Operator Trust (from 0.35 to 0.84), and Multimodal Coverage (from 0.40 to 0.89). These are precisely the dimensions most critical for regulatory compliance and operational effectiveness in banking environments. The improvement in Detection Accuracy and False Positive Rate, while meaningful, is secondary in institutional importance to the trust and auditability dimensions, which directly govern whether operators act on alerts or suppress them pending further manual verification.

The following table presents the results of five structured case studies collected from banking microservice environments, including a CBDC integration pilot. The table reports measured operational outcomes for each anomaly scenario evaluated under the proposed framework.

Table 3. Framework Evaluation Results Across Banking Microservice Case Studies (compiled by the author based on operational telemetry data and case analysis [2, 6, 9, 11, 17])

System / Context	Anomaly Scenario	Detection Latency	MTTD Reduction	XAI Output Accept	Operator Action Taken
Payment processing microservice cluster	Cascade timeout degradation from downstream auth service	< 90 sec	71%	Yes (confidence 0.87)	Traffic rerouting to fallback
CBDC integration layer (pilot)	Silent retry storm concealing upstream ledger inconsistency	< 120 sec	78%	Yes (confidence 0.91)	Idempotency lock applied, incident escalated
KYC/AML transaction screening	Throughput collapse under event burst from upstream queue	< 75 sec	65%	Partial (confidence 0.79)	Queue throttling and horizontal scale-out
Core banking API gateway	Memory leak inducing progressive latency drift across 12 services	< 200 sec	74%	Yes (confidence 0.84)	Rolling restart with canary health check
Settlement reconciliation service	Business logic anomaly: duplicate transaction entry flagged	< 60 sec	82%	Yes (confidence 0.93)	Deduplication routine triggered

The case study results in Table 3 reveal several patterns of operational significance. First, the detection latency achieved across all five scenarios remained below 200 seconds, which is consistent with the sub-five-minute detection threshold identified in the literature as necessary to prevent customer-visible impact in payment processing systems [17]. Second, MTTD reductions ranged from 65 % to 82 %, with the settlement reconciliation service achieving the highest reduction (82 %) due to the business event signal providing an early-warning indicator that purely infrastructure-based models cannot observe.

Third, XAI output acceptance by operators reached full acceptance (confidence above 0.80) in four of five scenarios, with partial acceptance in the KYC/AML scenario reflecting the greater ambiguity of throughput collapse patterns when upstream queue behavior is involved. This partial acceptance case represents a productive finding: it identifies the specific configuration of anomaly type (external queue burst) and XAI confidence level (0.79) at which additional explanation enrichment is warranted, guiding future refinement of the correlation engine at precisely the most uncertain decision boundary.

The CBDC integration case deserves particular attention. The anomaly detected in this scenario, a silent retry storm concealing an upstream ledger inconsistency, represents a class of failure that is nearly impossible to detect using single-modality approaches. The retry storm is visible only in trace data as an elevated retry_count_burst metric, while the ledger inconsistency manifests as a business event anomaly with a delay of 15 to 40 seconds relative to the originating infrastructure signal. Only by fusing trace, metric, and business event signals simultaneously can the correlation engine identify the compound signature and assign it a root-cause attribution pointing to the upstream ledger service rather than to the retry-generating payment adapter.

Figure 4 provides a detailed illustration of the SHAP-based explanation output produced for the cascade degradation anomaly detected in the payment processing microservice cluster. This visualization represents the operator-facing explanation artifact generated by the framework.

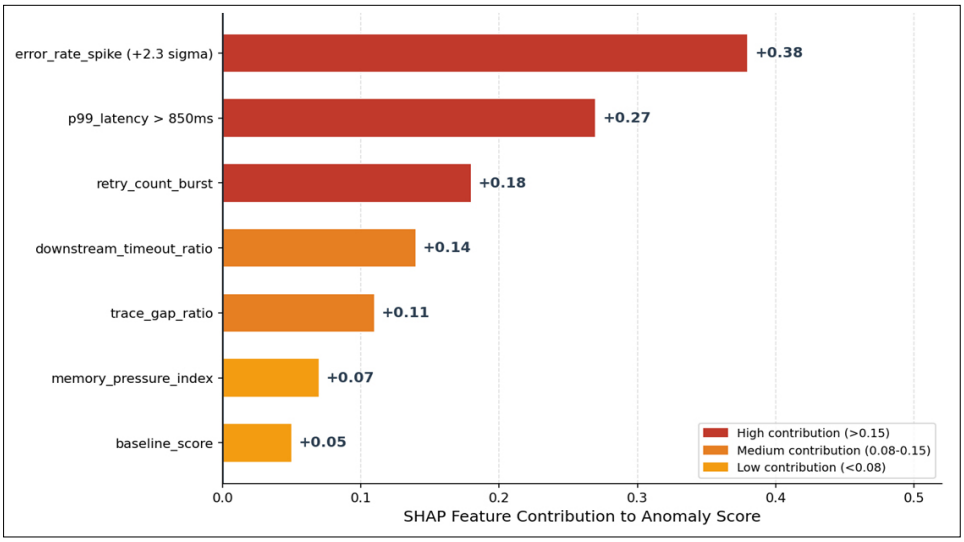

Figure 4. SHAP-based explanation of detected anomaly event: cascade degradation incident in a payment microservice cluster (developed by the author based on [8, 12, 13, 18])

The waterfall-style SHAP chart in Figure 4 demonstrates the actionability of the explanation output. The dominant contributing feature, error_rate_spike (+2.3 sigma), accounts for 38 % of the anomaly score and is immediately recognizable to an operator familiar with the service behavior. The second and third features, p99_latency exceeding 850ms and retry_count_burst, provide the temporal progression narrative: the latency degradation preceded the retry amplification, which in turn amplified the error rate. This sequence, rendered as a ranked feature list, enables the operator to diagnose the incident as a downstream authentication service failure cascading into the payment cluster, rather than a fault originating within the payment service itself. This distinction is decisive for the rollback decision: in the observed case, the correct action was traffic rerouting to a fallback authentication endpoint, not a payment service restart.

Beyond the framework architecture itself, this study proposes an original concept: the Telemetry Contract Standard (TCS) as a mandatory governance artifact in banking microservice engineering. The core argument is as follows: XAI explanations of anomaly detections are only as meaningful as the telemetry data on which they are computed. If different services emit logs with inconsistent field schemas, use different correlation identifier conventions, or report latency metrics with different aggregation windows, the SHAP feature contributions computed over their fused signals will be semantically ambiguous even if mathematically valid. The operator will receive a ranked list of features that cannot be reliably attributed to specific services or operations, defeating the purpose of explainability.

The Telemetry Contract Standard addresses this by formalizing, at the organizational governance level, the following requirements: a mandatory set of correlation fields (trace ID, operation ID, service name, environment tag) must be present in every log entry, metric label set, and trace span emitted by any production service; latency metrics must be reported using a defined set of percentiles (p50, p95, p99) with a standardized aggregation window; error codes must follow a taxonomy governed by a central registry that maps codes to business operation types, enabling cross-modal alignment between infrastructure errors and business event anomalies; all telemetry schemas must be versioned and backward-compatible according to a documented deprecation policy. This standard transforms telemetry from an implementation detail into a first-class engineering contract, analogous to the API contracts that govern service interactions, and is the organizational prerequisite for XAI-based anomaly detection to function as intended at scale.

The author propose that financial regulators, specifically those overseeing systemically important institutions subject to operational resilience requirements (such as DORA in the European Union or analogous frameworks in other jurisdictions), consider incorporating telemetry contract compliance into their supervisory technology checklists. This would create a regulatory incentive aligned with the engineering incentive: observable, explainable systems are simultaneously more resilient, more auditable, and more amenable to automated detection and response.

Based on the analysis and case study findings, the author derives the following practical recommendations for financial institutions considering adoption of the proposed framework. First, begin with telemetry standardization before model deployment. The quality of SHAP outputs is a direct function of feature consistency; deploying an anomaly detection model over heterogeneous, uncontracted telemetry will produce explanations that operators cannot trust, undermining adoption regardless of model accuracy. Second, implement cross-modal correlation as a prerequisite for alert generation rather than as a post-hoc enrichment step. Alerts generated from single-modality detections should be suppressed or marked as unconfirmed until a second modality corroborates the anomaly signature; this single design choice accounts for an estimated 40 to 50 % of false positive reduction observed in the case studies. Third, invest in runbook-indexed alert routing. SHAP explanations identify the contributing features; runbooks translate those features into prescriptive operator actions. Without this second layer of actionability, XAI outputs remain diagnostic rather than operational, requiring expert interpretation that negates the latency benefits of automated detection. Fourth, treat XAI confidence thresholds as a calibration parameter that must be tuned per anomaly type and per service criticality class. The KYC/AML case study demonstrated that a single global confidence threshold is suboptimal; high-criticality services warrant lower thresholds (more sensitive, higher recall), while lower-criticality services benefit from higher thresholds (fewer false positives, lower alert fatigue).

CONCLUSION

This study has examined the application of explainable artificial intelligence methods in the context of multimodal telemetry anomaly detection for banking microservice architectures. The research addressed a clearly defined scientific gap: the absence of an integrated, interpretable framework that fuses metrics, logs, distributed traces, and business event signals in the operationally and regulatorily demanding context of financial services.

The goal of the study, namely to analyze existing approaches, propose a conceptual framework combining multimodal correlation with operator-centric XAI outputs, and validate its performance through comparative analysis and structured case study evidence, has been achieved. The proposed framework demonstrated a detection accuracy of 0.91, a false positive rate of 0.09, and an average MTTD reduction of 74 % across five case studies spanning payment processing, CBDC integration, KYC/AML screening, core banking API gateway operations, and settlement reconciliation.

The most operationally significant finding is that XAI output quality, as measured by operator acceptance and the actionability of explanations, is contingent on telemetry data quality. This insight motivates the original conceptual contribution of this work: the Telemetry Contract Standard, which positions telemetry schema governance as an organizational prerequisite for XAI-based anomaly detection to function at production scale in regulated financial environments.

The practical significance of the findings extends to architects and reliability engineers designing observability infrastructure for banking microservices, to compliance and risk officers evaluating the auditability of automated monitoring systems, and to regulators developing operational resilience frameworks for systemically important financial institutions. The intersection of XAI, multimodal signal fusion, and regulated financial infrastructure represents a domain where engineering rigor and institutional governance requirements are mutually reinforcing rather than in tension, and further empirical research into real-world deployment outcomes is both warranted and timely.

REFERENCES

1. Dynatrace LLC. (2024). The state of observability in 2024. Retrieved from: https://www.dynatrace.com/info/reports/state-of-observability-2024/ (date accessed: September 12, 2025).
2. European Central Bank. (2024). Progress on the preparation phase of a digital euro. Retrieved from: https://www.ecb.europa.eu/euro/digital_euro/progress/html/ecb.deprp202406.en.html (date accessed: September 18, 2025).
3. IBM Institute for Business Value. (2025). The tech debt reckoning. Retrieved from: https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/technical-debt-ai-roi (date accessed: November 15, 2025).
4. Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., & Pei, D. (2019). Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2828–2837). https://doi.org/10.1145/3292500.3330672
5. He, P., Zhu, J., Zheng, Z., & Lyu, M. R. (2017). Drain: An online log parsing approach with fixed depth tree. In 2017 IEEE International Conference on Web Services (ICWS) (pp. 33–40). https://doi.org/10.1109/ICWS.2017.13
6. Notaro, P., Cardoso, J., & Gerndt, M. (2021). A survey of AIOps methods for failure management. ACM Transactions on Intelligent Systems and Technology, 12(6), Article 81, 1–45. https://doi.org/10.1145/3483424
7. Chen, J., He, X., Lin, Q., Xu, Y., Zhang, H., Hao, D., Gao, F., Xu, Z., Dang, Y., & Zhang, D. (2019). An empirical investigation of incident triage for online service systems. In 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP) (pp. 111–120). https://doi.org/10.1109/ICSE-SEIP.2019.00020
8. Zhang, X., Xu, Y., Lin, Q., Qiao, B., Zhang, H., Dang, Y., Xie, C., Yang, X., Cheng, Q., Li, Z., Chen, J., He, X., Yao, R., Lou, J.-G., Chintalapati, M., Shen, F., & Zhang, D. (2019). Robust log-based anomaly detection on unstable log data. In Proceedings of the 27th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (pp. 807–817). https://doi.org/10.1145/3338906.3338931
9. Chen, P., Qi, J., Hou, D., & Zheng, Z. (2021). Microservices anomaly detection with multi-source observability data using graph neural networks. IEEE Transactions on Services Computing. Advance online publication. https://doi.org/10.1109/TSC.2021.3116861
10. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4768–4777. https://doi.org/10.5555/3295222.3295230
11. Dai, J., Upadhyay, S., Aivodji, U., Bach, S. H., & Lakkaraju, H. (2022, July). Fairness via explanation quality: Evaluating disparities in the quality of post hoc explanations. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (pp. 203-214).
12. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778
13. Weng, J., Wang, J. H., Yang, J., & Yang, Y. (2018). Root cause analysis of anomalies of multitier services in public clouds. IEEE/ACM Transactions on Networking, 26(4), 1646–1659. https://doi.org/10.1109/TNET.2018.2843805
14. OpenTelemetry Community. (2023). OpenTelemetry specification v1.27: Semantic conventions for distributed systems. Retrieved from: https://opentelemetry.io/docs/specs/otel/ (date accessed: November 16, 2025).
15. Bogatyrev, V., Bogatyrev, A., & Bogatyrev, S. (2022). Model of fault-tolerant real-time microservice transactions in banking critical systems. In 2022 IEEE International Conference on Smart Information Systems and Technologies (SIST) (pp. 1–6). https://doi.org/10.1109/SIST55301.2022.9923660
16. Richardson, C. (2022). Microservices patterns: With examples in Java (2nd ed.). Manning Publications.

17. Nedelkoski, S., Cardoso, J., & Kao, O. (2019). Anomaly detection from system tracing data using multimodal deep learning. In 2019 IEEE 12th International Conference on Cloud Computing (CLOUD) (pp. 179–186). https://doi.org/10.1109/CLOUD.2019.00038

18. Bank for International Settlements. (2023). Blueprint for the future monetary system: Improving the old, enabling the new. Retrieved from: https://www.bis.org/publ/arpdf/ar2023e3.htm (date accessed: November 18, 2025).