Scholar Publishing Group
International Journal of Big Data Intelligent Technology
https://doi.org/10.38007/IJBDIT.2026.070205
ISSN 2790-0932 Vol. 7, Issue 2: 40-50

Research on AI Driven Big Data Platform Intelligent Operation and Anomaly Detection Technology

Chenghao Shi

Georgia Institue of Technology, Atlanta, GA, 30332, USA

Keywords: Big Data Platform; Intelligent Operation and Maintenance; Anomaly Detection; Cloud-Native; Time Series Analysis; Root Cause Localization

Abstract: Due to the combination of cloud-native architecture, distributed storage, real-time computing and multi-tenant resource scheduling, the operational attributes of big data platforms are high-dimensional, dynamic, strongly coupled and have a low labeling level. The existing operation and maintenance method is a static-threshold, manual-inspection approach that does not meet the requirements of low-latency, high-reliability and interpretable governance. This paper introduces an AI-based intelligent operation and maintenance framework for cloud-native big data platforms and focuses on the technical chain of "intelligent perception - anomaly identification - root cause localization - closed-loop handling". It unifies the modelling of multi-source data, including metrics, logs, traces, configurations and events, and integrates strong standardization, time-series predictive reconstruction, dynamic graph association, and adaptive thresholds to achieve anomaly detection. According to research, the core of intelligent operation and maintenance is not a high-accuracy single-point model, but rather the cooperation of data governance, real-time inference, alarm suppression and root cause analysis. This paper also proposes a multi-layered platform architecture and practical implementation plans to offer methodological support for transforming big data platforms from passive fault handling to proactive risk warning.

1 Introduction

Big data platforms have moved beyond the era of a single offline computing cluster and are now integrated infrastructures that support data lakes, streaming computation, interactive analysis, machine learning training, data services, and scheduling orchestration simultaneously. The platform has distributed file systems, resource managers, message queues, metadata services and computing engines, as well as support functions such as container orchestration, service mesh, monitoring and collection, log center and permission auditing. Due to the continuous superposition of factors such as business peaks and troughs, data skew, job dependencies, network jitter, node aging and version upgrades, platform anomalies are often not caused by a single indicator exceeding its limit but rather by the linkage offset of multiple subsystems in the time dimension, topology dimension and semantic dimension. Research in intelligent operation and maintenance has applied machine

learning and large language models to improve fault detection, root cause analysis and auxiliary handling functions. According to the above research, the new stage of intelligent operation and maintenance is moving from single-task detection to cross-data source understanding, automated reasoning and scripted repair [1].

Operation and maintenance of the old large data platform are mainly based on fixed thresholds, expert rules and human experience. The model has the advantage of a small-scale implementation and is relatively straightforward to understand for small systems with relatively stable loads. However, its deficiencies are exposed in the course of cloud-native development. First, container instances have a short life cycle and elastic scaling features; therefore, the baseline for the same service is inconsistent among different nodes and at different times. Second, log texts, call chains and performance indicators are in different formats; thus, a single display window cannot observe the whole process of an abnormality. Third, operations and maintenance teams are susceptible to alert fatigue due to a large number of alerts; without root cause aggregation and prioritization, even detection models with high recall rates may be practically useless.

Given the above reasons, this paper will conduct research on intelligent operation and maintenance, as well as anomaly detection technologies for big data platforms powered by artificial intelligence. Anomaly detection in this paper does not refer solely to the output of a classification model but rather encompasses all links in the entire process of platform operation and governance, including data access, feature modelling, online inference, anomaly interpretation, root cause localisation and strategy execution. The three goals of this study are: first, to examine the present situation of technology and major problems in intelligent operation and maintenance of cloud-native big data platforms; second, to build a mathematical expression and platform architecture for anomaly detection that supports multiple sources of observable data; and third, to put forward strategies for model updates, threshold calibration, alarm convergence and closed-loop control in engineering applications. The general structure of this paper is as follows: Introduction, analysis of the current situation in research on this topic, problem statement, solutions and strategies, and conclusion.

2. Current Status Analysis of the Research Topic

2.1 Multi-source observable data forms the foundation of intelligent operation and maintenance.

The first level of capability for intelligent operation and maintenance (O&M) is full knowledge of the platform's working conditions. Observable data provided by the big data platform generally includes resource statistics, system logs, task scheduling information, call chains, configuration modifications, container events and user access behaviour. Compared with traditional single-machine monitoring, this type of data has three specific features: First, there is a huge volume of data, and metric sampling frequency has reached the second level; Second, the structure is fundamentally different; metrics are numerical sequences, logs are semi-structured text, and call chains are directional service graphs; Third, there is strong semantic coupling, and an increase in computational job latency can be caused by storage congestion, resource contention, network retransmissions, or delayed upstream data. If there is no single-data model, the different monitoring systems will not be interconnected and the scope of anomaly detection will be limited.

In unified modelling, the platform state at the same time step can be represented as a multimodal runtime vector:

$$\mathbf{x}_t = [\mathbf{m}_t; \mathbf{l}_t; \mathbf{r}_t; \mathbf{c}_t; \mathbf{e}_t] \in \mathbb{R}^d \tag{1}$$

t is the sampling time in Equation (1), x_t is the comprehensive state vector of the platform at that time, m_t is the resource and performance indicator vector, l_t is the log template or log semantic embedding, r_t is the call chain and service dependency features, c_t is the configuration, version and change information, e_t is the external events and alarm context, and d is the fused feature dimension. This expression puts the numerical, text and topological information into the same state space to provide a unified input for the following time-series prediction, anomaly scoring and root-cause reasoning.

Table 1. Multiple sources of observable data and their operational values.

Data source	Main fields	Update granularity	Operational value	Typical noise
Metrics	CPU, memory, I/O, network, latency	1–60 s	Baseline learning and trend prediction	Missing points
Logs	Template ID, severity, message embedding	Event-driven	Semantic anomaly and event explanation	Format drift
Traces	Span, dependency edge, duration, status	Request-level	Propagation path and bottleneck analysis	Sampling bias
Configuration	Version, parameter, deployment record	Change-driven	Change impact and rollback decision	Incomplete record
Scheduler events	Job state, queue, resource request	Event-driven	Job delay and resource contention analysis	Duplicate events
User behavior	Query volume, data access, tenant ID	Minute-level	Demand shift and abnormal access detection	Privacy Masking

Table 1 shows the five foundations for intelligent operation and maintenance of a big data platform: data source, fields, granularity, operational value and noise type. It can be seen that metrics are suitable for describing continuous changes, logs are suitable for preserving semantic evidence, call chains are suitable for showing propagation relationships, and configuration and scheduling events can explain the circumstances under which an anomaly is triggered. The relationship among multi-source data is not a simple collection; time alignment, semantic cleaning and topological mapping must all be performed before the generation of reliable features that can be used in online inference.

2.2 The time series anomaly detection model shifts from single-point identification to comprehensive discrimination.

Time-series anomaly detection algorithms have evolved from traditional statistics to deep learning, contrastive learning, pre-trained models, and multimodal fusion at the algorithm level. Recent studies have shown that the main categories of deep time-series anomaly detection fall into reconstruction error, prediction error, density estimation, adversarial learning, attention mechanism and graph structure modelling. It is good at extracting complex nonlinear patterns, but it is also relatively difficult to interpret, and the evaluation criteria and generalization capacity are lacking [2]. For large-data platforms, abnormal conditions may take various forms, such as an abrupt rise or fall in the value of a single indicator, a regular periodic fluctuation, a loss of correlation structure, inconsistent semantics in log sequences, or partial disruption of job dependency chains. Therefore, relying solely on whether a certain indicator exceeds a fixed threshold for the basis of an alarm is no longer suitable for the demands of platform-level intelligent operation and maintenance.

To reduce the effect of different dimensions and extreme values in the indicators, a robust normalised expression based on a sliding window is used here:

$$z_{t,j} = \frac{x_{t,j} - \text{median}(W_{t,j})}{\text{MAD}(W_{t,j}) + \varepsilon} \quad (2)$$

$x_{t,j}$ is the original value of the j-th feature at time t; $W_{t,j}$ is the sliding window sample ending at t; median is the median value in the window; MAD is the median absolute deviation; ε is a small constant to avoid division by zero; and $z_{t,j}$ is the feature value after robust standardization. Mean-Variance Normalization is relatively susceptible to sudden spikes, and therefore, it may not be suitable for online data streams with local anomalies before and after an alarm.

2.3 Publicly available benchmark statistics reflect differences in anomaly detection data.

Another problem of the difficulty of anomaly detection in real applications is that the datasets are not the same. Reliable benchmark studies show that the data for multivariate time-series anomaly detection vary greatly in terms of average dimension, sequence length, anomaly proportion, and anomaly type; thus, the same model may perform differently on various datasets due to differences in anomaly density, dimensional structure and event length [3]. Therefore, the best result of an intelligent operation and maintenance platform should not be achieved by a single model on a small number of datasets; instead, greater attention should be paid to data adaptability, indicator stability and cross-scenario migration capabilities. The following two statistical charts are based on the statistical fields of multivariate datasets in public benchmarks and can show the differences in anomaly proportion and data scale intuitively.

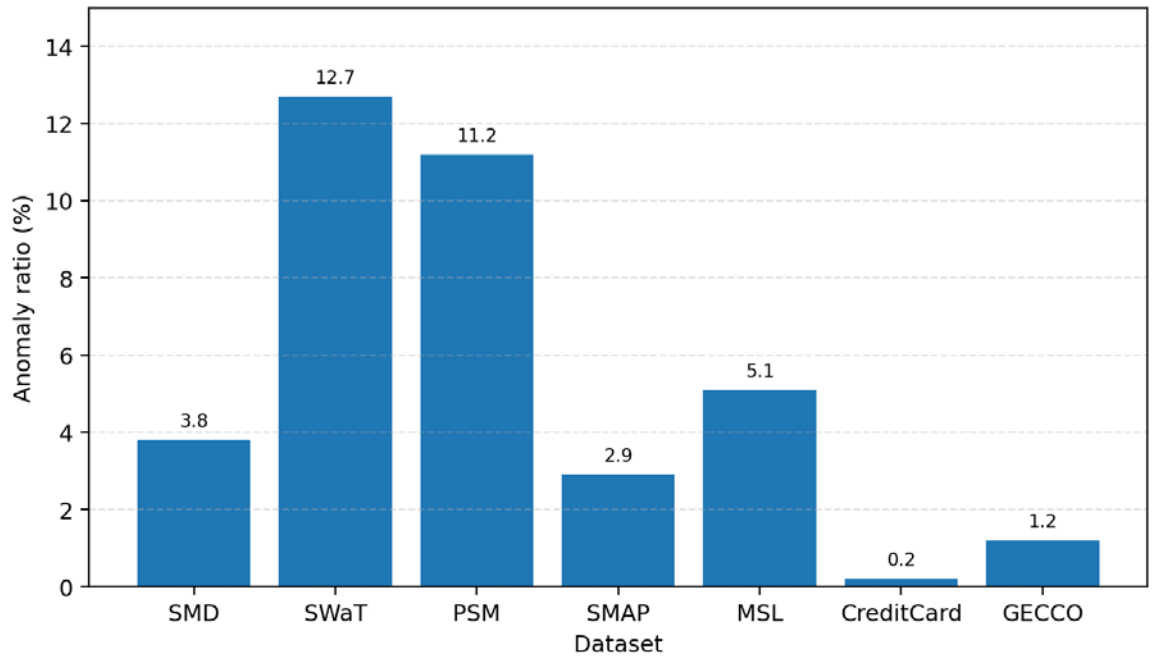

Figure 1. Statistics on the Percentage of Anomalies in the Multivariate Anomaly Detection Benchmark Dataset (Data source: [3])

As shown in Figure 1, the percentages of anomalies in the various multivariate time-series datasets are not the same. SWaT and PSM have a relatively high number of anomalies; CreditCard and GECCO have a lower number. Therefore, the choice of model threshold and the interpretation

of evaluation results will vary. Datasets with a high proportion of anomalies are more likely to increase recall; therefore, those with a lower anomaly proportion need stronger false positive control. Therefore, when deploying models in the large data platform, different thresholds should be set for each business area rather than using a single threshold for all subsystems.

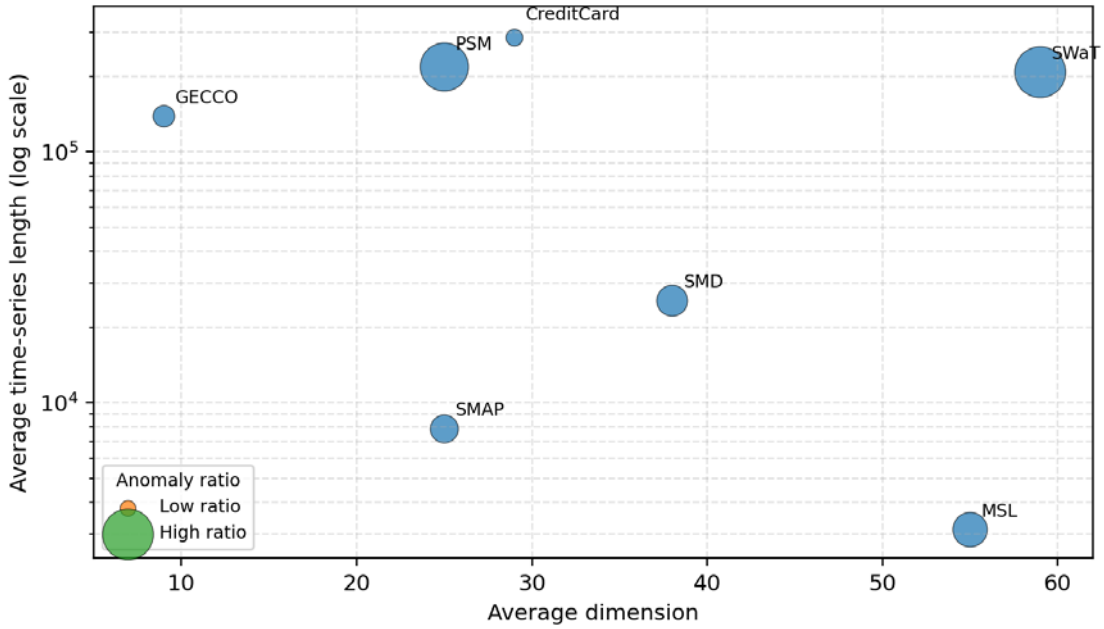

Figure 2. Relationship between dimension and sequence length in multivariate anomaly detection benchmark dataset (Data source: [3])

Figure 2 shows the difference in dataset structure for average dimensionality and average sequence length, and point size represents anomaly density. SWaT is high-dimensional and has a long sequence, suitable for testing high-dimensional association models; PSM and CreditCard have long sequences but moderate dimensionality, suitable for assessing long-term drift identification; MSL is high-dimensional but has a short sequence, and thus more prone to window-length limitations in the model. Therefore, the platform model should consider dimensions, time intervals and anomaly sparsity at the same time.

2.4 A gap remains between academic benchmarks and industrial deployment.

In recent years, industrial benchmarks have begun to pay more attention to the deployment adaptability of time-series models in new areas, new businesses and multiple evaluation scenarios. TimeSeriesBench indicates that most existing algorithms have been trained on only one curve and thus do not support large-scale systems with tens of thousands of curves. Continuous release and updates of distributed systems will also lead to new indicators appearing continuously [4]. TAB further enhances the time series anomaly detection benchmark in terms of dataset coverage, method categories, unified evaluation process and reproducible experimental pipeline, and includes multiple methods such as non-learning, machine learning, deep learning, large models and pre-trained time series models [5]. These studies have shown that intelligent operation and maintenance for big data platforms need to address many issues in addition to the offline accuracy comparison, such as model maintenance cost, cross-domain generalisation, online latency, interpretation credibility and the closed-loop effect of alarms.

3. Raise questions

3.1 The complexity of the platform makes it difficult to identify the abnormal propagation chain.

Anomalies in cloud-native big data platforms are very spreadable. For example, disk jitter in a storage node will first appear as an increase in read/write latency, followed by a wait for computing tasks, accumulation of resource queues, backlog of upstream messages, and timeout of user queries. If the monitoring system only detects independently within each component, a large number of local alarms will be triggered simultaneously, and the earliest point of instability will not be known. Contrastive Learning and Attention Mechanisms can enhance the distinguishing capability of anomaly representations. DCdetector improves the detection of time-series anomalies via dual attention and contrastive representation learning [6]. The Service Topology and Configuration of the actual platform also affect anomaly propagation. Only time-based representations are used, and the root cause cannot be identified.

3.2 Low annotation and concept drift limit the long-term effectiveness of the model

The fault labels of big data platforms generally come from manual work orders, post-event reviews and alarm confirmations, and there are problems such as a lack of labels, inconsistent granularity and blurred time boundaries. The "normal samples" that can be obtained during the model training stage are not pure; weak anomalies, grey releases and short-term congestion are often mixed in the historical running data. DTAAD is a model that enhances the performance of multivariate temporal anomaly detection by combining temporal convolution, attention and autoencoder structures, focusing on lightweight and fast reasoning and diagnosis [7]. However, as the platform is used for a longer time, the amount of business, behaviour of tenants, resource quotas and component versions will all change; thus, the normal mode itself will drift. If the model does not have an online update and drift perception mechanism, threshold aging, false alarms and false negatives will all increase.

3.3 There is a contradiction between alarm accuracy and operational feasibility.

The above three models generally have a high precision and recall. However, given the demands of operation and maintenance (O&M), the requirements for interpretability in alerts and the ability to pinpoint specific faulty components for a low-risk response plan have risen. A high-recall model will output a larger number of candidate anomalies, increasing the work for on-call staff; a high-precision model may fail to identify early, weak anomalies and thus allow the fault to escalate. In multi-tenant big data platforms, if model outputs cannot be mapped to tasks, queues, nodes, services and change records, it will be difficult to trigger O&M actions. Therefore, intelligent O&M needs to address the problem of detection as well as the integrated issues of "anomaly discovery", "scope of impact determination" and "governance strategies".

4. Problem Solving and Strategies

4.1 Constructing a Layered and Collaborative Intelligent Operation and Maintenance Platform Architecture

The five layers in the paper serve as a reference for the actual application: Data Acquisition Layer, Feature Governance Layer, Intelligent Analysis Layer, Disposal Collaboration Layer and

Feedback Learning Layer. The data acquisition layer obtains second-level or event-level data via an indicator collector, log broker, tracing and configuration auditing system; the feature governance layer performs time alignment, missing data imputation, log templating, vectorisation and topology mapping; the intelligent analysis layer detects anomalies, identifies root causes, assesses impacts and aggregates alarms; the disposal collaboration layer interfaces with work orders, orchestrates scripts, manages grey-scale rollbacks, adjusts resources and notifies users; finally, the feedback learning layer saves manual confirmation results and disposal effects for threshold optimisation, sample reweighting and model retraining.

Table 2 Layered Architecture and Key Technologies of Intelligent Operation and Maintenance Platform

Layer	Component	Key method	Output	Deployment note
Collection	Metrics agent, log agent, trace collector	Time alignment, buffering	Telemetry stream	Low overhead
Governance	Feature store, topology registry	Cleaning, embedding, graph mapping	Unified feature view	Schema control
Analysis	Detector, correlator, RCA engine	Forecasting, reconstruction, graph reasoning	Anomaly score and root cause	Online inference
Collaboration	Ticket, alert router, runbook	Deduplication, priority ranking	Action proposal	Human approval
Feedback	Label store, drift monitor	Active learning, threshold tuning	Updated model and rules	Continuous loop

Table 2 shows that the five levels of the intelligent operation and maintenance platform are shown, as well as the closed-loop connection from data acquisition to feedback learning. The collection layer ensures data consistency; the governance layer rectifies any irregularities in the various data; the analysis layer spots problems and determines the reason for such a problem; the collaboration layer converts model results into work orders and plans; and the feedback layer sends back-end verification results for model optimisation. The above design will reduce the risk of the algorithm being detached from the operational and maintenance cycle, thereby preventing the failure of anomaly detection.

4.2 Establish an anomaly scoring mechanism for predictive reconstruction fusion

A single reconstruction error is taken to be an abnormality and thus incorrectly identified as such, whereas a single prediction error may be excessively sensitive to changes in the long-term trend. Therefore, this paper introduces an anomaly scoring method based on prediction and reconstruction errors to determine whether the current state can be reconstructed from a normal pattern and whether changes in the next time step follow the historical dynamic pattern.

$$s_t = \alpha \|\mathbf{x}_t - \hat{\mathbf{x}}_t\|_2^2 + (1-\alpha)\|\mathbf{x}_{t+1} - f_\theta(\mathbf{X}_t)\|_2^2 \tag{3}$$

Equation (3) is shown below: E_t^{rec} is the reconstruction error, E_t^{pred} is the prediction error, s_t is the comprehensive anomaly score at time t, x_t is the true state vector, $\hat{x}_t$ is the model reconstruction state, f_θ represents the time-series prediction model, w is the historical window length, and α is the weighting coefficient between the reconstruction error and the prediction error. The value of α can be adjusted according to the business circumstances to increase the weight of the prediction item for

resource indicators with strong periodicity, and increase the weight of the reconstruction item for log semantics and configuration status.

4.3 Adaptive thresholds are used to mitigate false alarms and false negatives.

Fixed thresholds in cloud-native big data platforms are unable to adapt to fluctuations in weekdays and holidays, different times of day (daytime/nighttime), as well as peak batch processing and interactive query periods. Exponentially weighted moving average is used to update the distribution of outlier scores online, and then a dynamic threshold is generated.

$$\tau_t = \mu_t + \lambda\sigma_t, \quad \mu_t = \mathrm{EWMA}(s_t), \quad \sigma_t = \mathrm{EWMStd}(s_t) \tag{4}$$

As shown in equation (4), μ_t is the dynamic mean of the anomaly score, σ_t is the dynamic standard deviation, ρ is the smoothing coefficient, λ is the alarm sensitivity coefficient, and τ_t is the dynamic threshold at time t. A candidate anomaly occurs when s_t exceeds τ_t. A larger ρ has a relatively stable threshold but responds slowly to sudden changes; a larger λ reduces false alarms but increases the probability of missing a true alarm. Different Parameters may be used for various components in different actual deployments based on service level, tenant priority and past false alarm rates.

4.4 Integrating log semantics and topological correlations for root cause localization

Log data has strong semantic information for anomaly detection. LogGPT learns log sequence patterns through generative models and improves the discrimination target by using reinforcement learning strategies for anomaly detection [8]; LogLLM attempts to leverage large language models to process log semantics and can thus still understand the meaning of text in unstable log situations [9]. Semantics of logs need to be combined with service topology, call chains and task dependencies in the big data platform. Candidate root causes are taken as services, nodes, queues, jobs, configuration items, etc., and the posterior probabilities of root causes are calculated based on multi-source evidence in this paper.

$$\pi_i(t) = \mathrm{softmax}_i\big(q_i(t)\big), \quad \hat{c}_t = \mathrm{argmax}_i \pi_i(t) \tag{5}$$

Equation (5) is as follows: $\pi_i(t)$ is the posterior probability of the i-th candidate root cause at time t; h_t is the comprehensive semantic and temporal representation of the anomaly window; g_i is the structural representation of the candidate component in the service topology; Δ_i represents the evidence of changes, logs and metric offsets of the component relative to the anomaly window; w is a learnable weight vector; and $\hat{c}_t$ is the final root cause prediction result. The purpose of the above formula is to map "anomalies" and "component evidence" to a common score space, and thus generate an ordered list of the root causes rather than a simple yes/no determination of whether it is an anomaly.

4.5 Establish an alarm compression and closed-loop processing mechanism

Only when the results of anomaly detection are added to the handling process will they be practically useful for operation and maintenance. LogRCA has proposed a way to find the smallest set of root cause evidence from logs in a distributed service to solve the problem of locating the actual root cause in a large number of anomaly logs, and has verified this on large-scale production log data [10]. Based on the above, the three kinds of mechanisms at the alarm level in a big data platform are introduced [11-13]. First, a mechanism for aggregating alarms is used to group alarms that belong to the same topological neighbourhood, the same job chain or the same configuration

change in a given time window and only trigger one notification for them [13-16]. Second, the impact assessment mechanism determines the alarm priority according to the affected tenant, job priority, data service level and failure propagation path. Third, the handling orchestration mechanism will automatically manage low-risk operations, such as expanding replicas, isolating abnormal nodes, restarting failed executors and pausing low-priority tasks; and offer suggestions for high-risk operations that still require manual approval [17-20].

Table 3 Evaluation Indicators for Anomaly Detection and Intelligent Operation & Maintenance.

Metric	Measurement focus	Target direction	Operational meaning	Caution
Precision	Correctness of generated alerts	Higher	Reduces alert fatigue	May hide missed incidents
Recall	Coverage of true incidents	Higher	Improves early warning	May increase alert volume
F1 score	Balance between precision and recall	Higher	Compares detector quality	Sensitive to labels
VUS-PR	Event-oriented ranking performance	Higher	Supports robust TSAD evaluation	Needs benchmark protocol
MTTA	Time to acknowledge	Lower	Reflects alert routing efficiency	Depends on workflow
MTTR	Time to recovery	Lower	Measures remediation effectiveness	Influenced by human action

Table 3 shows the model evaluation indicators and operational process indicators in the same table. Precision, recall and F1 score show the performance of the detection model, but do not fully reflect the governance effectiveness of the platform; VUS-PR is more suitable for assessing the quality of time-series anomaly ranking, and mean confirmation time and mean recovery time indicate the closed-loop capabilities of alarm distribution, root cause localisation and handling [21-23]. Offline detection indicators and online work order efficiency should be used in the actual acceptance test to avoid a discrepancy between model accuracy and business value [24-25].

4.6 Project Implementation Path and Risk Control

The Construction of an intelligent operations and maintenance platform should be implemented in stages. The first stage is to build a unified data foundation by synchronizing indicators, logs, call chains and configuration changes; at the same time, construct service topology and job dependency graphs. The second stage will deploy unsupervised or weakly supervised anomaly detection models in a bypass mode to prevent directly triggering high-risk operations, and these models will be used to monitor the impact of false positives, missed positives and alarm aggregation. The third stage will identify the cause and propose a countermeasure; it will then be added to the work order system for staff approval prior to execution. The fourth stage is for autonomous operation under stable conditions, and the range of actions and rollback conditions should be limited; audit records must also be saved. The fifth stage builds a continuous learning loop, periodically adjusts the thresholds, sample weights and model parameters according to work order results, manual annotations and handling effects [26].

The four problems in risk control are as follows: First, data security: Logs and call chains may contain user identifiers, access paths or business fields, which need to be anonymized, have access control and be subject to minimal data collection. Second, model interpretability: The platform

should show the abnormal score as well as key indicators, abnormal log fragments, related components and change records. Thirdly, detect model drift: If the distribution significantly deviates from the normal distribution, either a new model should be trained or the threshold needs to be adjusted. Fourth, automation boundaries: Any operation that may affect data consistency, task results or tenant isolation should not be fully offloaded to the model for automatic execution. With the above control measures, intelligent operation and maintenance will be carried out efficiently, and engineering control will still be possible [27].

5. Conclusion

This paper presents AI-based intelligent operation and maintenance and anomaly detection technology for big data platforms. A multi-source state modelling method for cloud-native scenarios is proposed, a predictive reconstruction fusion anomaly scoring mechanism, an adaptive threshold strategy, a root cause posterior ranking model, and a hierarchical collaborative platform architecture. The research proposes that the core of anomaly detection in large-scale data platforms is no longer about improving offline model scores, but rather the collection and centralised management of metrics, logs, traces, configuration changes, scheduling events, etc. Thus, the cause of the anomaly can be located and addressed in a closed-loop manner.

First, the foundation for intelligent operation and maintenance (O&M) is multi-source observable data. By combining the above metrics, logs, topology and changes to create an integrated state vector, the deficiency of a single monitoring view has been addressed, and supporting data conditions for cross-component anomaly detection have been established. Second, anomaly detection models should be freed from fixed upper and lower bounds to build all-weather benchmarks. To deal with load fluctuations, periodic drifts and local anomalies, combinations of prediction error, reconstruction error, robust standardisation and adaptive thresholds are used. Third, to localize the root cause, combine semantic evidence and topological structure. Large language models can enhance the text-understanding abilities of models; however, only when combined with call chains, service dependencies and configuration changes can they be used to generate actionable root-cause rankings. Fourth, the effectiveness of intelligent O&M should be measured by closed-loop efficiency. Transform the model's output into an alarm compression, impact assessment, work order linkage and handling suggestion, and continuously improve through human feedback.

The three directions for further research have been put forward as follows: First, to improve the comparability of evaluation results, a unified cross-cluster benchmark for big data platforms should be established. Second, in order to enhance interpretability while maintaining low latency, a collaborative mechanism for large language models and time series basic models can be explored. Third, strengthen safe and controllable automated handling strategies so that intelligent operation and maintenance gradually shift from auxiliary diagnosis to trusted closed-loop governance.

References

[1] *Zhang, L.; Jia, T.; Jia, M.; Wu, Y.; Liu, A.; Yang, Y.; Wu, Z.; Hu, Zamanzadeh Darban, Z.; Webb, GI; Pan, S.; Aggarwal, CC; Salehi, M. Deep Learning for Time Series Anomaly Detection: A Survey. ACM Computing Surveys, 2024, 57(1): Article 15. DOI: 10.1145/3691338.*

[2] *Guo, X. (2025, April). Research on Financial Trading Algorithms Based on Deep Reinforcement Learning. In 2025 IEEE 3rd International Conference on Control, Electronics and Computer Technology (ICCECT) (pp. 1711-1715). IEEE.*

[3] *Guo, X. (2025, March). Research on Blockchain-Based Financial AI Algorithm Integration Methods and Systems. In 2025 IEEE International Conference on Electronics, Energy Systems and Power Engineering (EESPE) (pp. 816-821). IEEE.*

[4] *Yuan, Y. (2026). Research on Memory Management and Dynamic Reasoning Methods for Intelligent Agents Based on Large Language Models.*

[5] *Zhang, Z. (2026). Research on the Optimization of Payment Risk Dynamic Monitoring Models Driven by High-dimensional Behavioral Features. Advances in Computer and Communication, 7(3).*

[6] *Xin, H. (2026). Research on Distributed Data Cleaning and Standardized Mapping for Heterogeneous Logistics Data. Advances in Computer and Communication, 7(2).*

[7] *Su, J. (2026). Research on Machine Learning-based Performance Prediction and High-reliability Software Evolution Mechanisms for Android Communication Systems.*

[8] *Wang, Z. (2026). Valuation Enhancement Pathways for Resource Stocks Driven by the Growth of Energy Storage Demand. Engineering Advances, 6(3).*

[9] *Ma, X. (2026). Research on Elastic Scaling and Resource Scheduling Strategies for Distributed Systems Facing Traffic Fluctuations.*

[10] *Wang, N. (2026). Research on Digital Marketing Conversion Improvement Strategies and Predictive Analysis Based on User Behavior Segmentation. Journal of Humanities, Arts and Social Science, 10(7).*

[11] *Wang, Z. (2026). Data-driven Pathways for Optimizing Supply Chain Operational Efficiency in Physical Industries. Advances in Computer and Communication, 7(2).*

[12] *Jin, C. (2026). Research on the Optimization Strategy of Retail Enterprise Product Portfolio Based on Association Rule Analysis. Advances in Computer and Communication, 7(2).*

[13] *Ma, W. (2026). Reliable Data Infrastructure Supported by Automated Data Pipelines. Engineering Advances, 6(2).*

[14] *Wang, Z. (2026). Earnings Quality of Resource Enterprises Under Commodity Price Volatility Transmission.*

[15] *Xiaoyu Gu. (2025) Research on Generative AI Psychological Intervention Models Based on Multimodal Emotion Recognition. Advances in Computer and Communication, 6(5), 304-309.*

[16] *Li, J. (2026). Analysis of Advertising Creativity Generation and User Response Mode Based on AIGC.*

[17] *Li, J. (2026). Research on the Path and Efficiency of Empowering Accurate Advertising Delivery with Data Infrastructure.*

[18] *Zhang, Y. (2026). Research on LLM-Driven Intelligent Architecture Design and Autonomy Mechanism Construction for Cloud-Native Control Planes.*

[19] *Liu, X. (2026). Research on the Application of Artificial Intelligence in Consumer Privacy Data Protection. Procedia Computer Science, 279, 956-965.*

[20] *Liu, X. (2026). Construction of Consumer Data Privacy Protection System Based On Blockchain Technology. Procedia Computer Science, 282, 2082-2091.*

[21] *Zhang, Y. (2026). Exploration of Enterprise Big Data Microservice Architecture Based on Domain-Driven Design (DDD). Procedia Computer Science, 282, 1994-2003.*

[22] *Chi, C. (2026). Research on Mortgage Asset Cash Flow Simulation and Risk Transmission Mechanisms across Structural Tranches Based on Loan-Level Data.*

[23] *Wang, Z. (2026). Research on Data Analysis and Quality Prediction of Manufacturing Processes Based on Improved Deep Learning Algorithms. Procedia Computer Science, 281, 1328-1337.*

[24] *Ma, X. (2026). Distributed Fault Root Cause Localization and Self-Healing Strategy Generation Based on Causal Inference. Procedia Computer Science, 281, 753-760.*

[25] Wang, N. (2026). Research on Integrated Analysis Method of Sales and Operations Data for Management Decision Support.

[26] Han, X. (2026). Research on Adaptive Optimization Methods for Multi-Objective Manufacturing Process Parameters in Complex Assembly Processes. Procedia Computer Science, 279, 366-374.

[27] Sun, J. (2026). Automated Feature Engineering and Screening System for Large-Scale Factor Libraries. Procedia Computer Science, 281, 1282-1290.