



RESEARCH ARTICLE

OPEN ACCESS

EXPLAINABLE AND RISK AWARE AUTONOMOUS DECISION SYSTEMS A SCALABLE SECURE BY DESIGN ARCHITECTURE FOR LARGE SCALE DISTRIBUTED COMPUTING SYSTEMS

Abdinasir Ismael Hashi^{*1}, Osman Abdullahi Jama², Ibrahim Rashid Abdullahi³

^{1, 2, 3}Computer Science, Somali National university, Mogadishu, Somalia.

¹<https://orcid.org/0009-0009-0635-2609>, ²<https://orcid.org/0009-0008-2195-9276>, ³<https://orcid.org/0009-0008-2710-8173>

E-mail: *nasirhaji@snu.edu.so, manjama@snu.edu.so, ibrorashka@snu.edu.so

ARTICLE INFO

Article History

Received on: February 28, 2026

Reviewed on: March 29, 2026

Accepted on: April 01, 2026

Published on: July 31, 2026

Keywords:

Autonomic Computing,
 Explainable Artificial Intelligence
 (Xai),
 Cloud-Edge Computing,
 Distributed Systems.

ABSTRACT

This work introduces an explained, risk-conscious, and secure-by-design autonomic architecture of large scale distributed computing systems defining dynamic cloud-edge environments. The main issue it is expected to fix is the shortage of transparency, uncertainty quantification, and combined cybersecurity of the traditional black-box AI-driven distributed systems that restricts trust, scalability, and resiliency. In order to address these limitations, the suggested methodology combines some structured data preprocessing (Z-score filtering, entropy and variance extraction), probabilistic risk modeling, explainable AI based on SHAP, and Kubernetes KEIO-enabled adaptive orchestration into a single layered framework. Experimental analysis on cloud anomaly data sets shows good predictive accuracy of 91.1 percent, recall of 86 percent as well as precision of 79 percent as represented in the confusion matrix of 3333 true negative, 1000 true positive, 261 false positive, and 160 false negative. Risk scores are centred on 0.0, with most of the risk scores falling within the range of - 0.3 to 0.3, which is stable with respect to calibration. System availability indicates 94% of uptime, Mean Time to Recovery (MTTR) of 3.45 and 48% of scalability efficiency, which all indicate high uptime and efficient fault management. SHAP analysis reveals that power consumption (0.28) and memory usage (0.27) had the largest contributions to risks. In general, the architecture is able to implement transparent, resilient and scalable autonomous decision-making which is appropriate to mission critical distributed infrastructures.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I-INTRODUCTION

Distributed computing system Large-scale distributed computer systems support critical infrastructures like smart grids, autonomous transportation, healthcare networks, financial platforms, aerospace systems and cloud-edge ecosystems [1]. The trends towards the growing integration of the issue of artificial intelligence in these settings have allowed the development of autonomous decision making abilities that facilitate the efficiency, responsiveness and scalability of operations. But, the more there is autonomy, the greater the complexity of the system, uncertainty, and security vulnerability exposed. The autonomous decision systems implemented in the distributed environment should be able to work under dynamic workloads, diverse devices, adversarial threats, and uncertain environmental conditions. These needs require architectures that are not only intelligent or scalable, but also explainable, aware of risks and secure by design [2]. Conventional distributed systems are based on centralized control or loosely coupled micro services that do not have transparency in automated decision making processes. Black-box AI systems (especially deep learning), can also be highly predictive systems with low interpretability, which leads to a lack of trust in safety-critical and mission-critical systems [3].

In other domains, including aerospace robotics, smart healthcare and financial risk analytics, opaque decision-making may result in regulatory non-conformance, system failure and systemic risk spread. Thus, explainability ceases being a desired quality and turns into a mandatory one in terms of accountability, traceability, and collaboration between humans and AI. Explainable autonomous system: an explainable autonomous system should offer interpretable reasoning trajectories, model transparency and decision audits that can be validated and analyzed in a forensic way [4]. At the same time, distributed systems can be affected by different levels of uncertainty, such as partial information, stochastic interference, network delays, computer failure, and hostile cyber-attacks [5]. The risk attuned decision modeling approaches include probabilistic modeling and quantification of uncertainty alongside robust maximisation in order to guarantee the resilience in the face of disruptions. Risk-aware systems would evaluate the types of negative consequences that may occur in future and would take action in accordance with this information unlike deterministic control mechanisms that implement actions without this step. This is especially vital when considering infrastructures that are of large-scale engineering when a breakdown, at the local level, would be transferred into systemic failures. Risk assessment mechanisms with autonomous control loops help to increase reliability, fault tolerance and safety of operation [6]. The figure 1 illustrates key sources of uncertainty in healthcare, including measurement variability, incomplete data, diagnosis ambiguity, and treatment uncertainty.

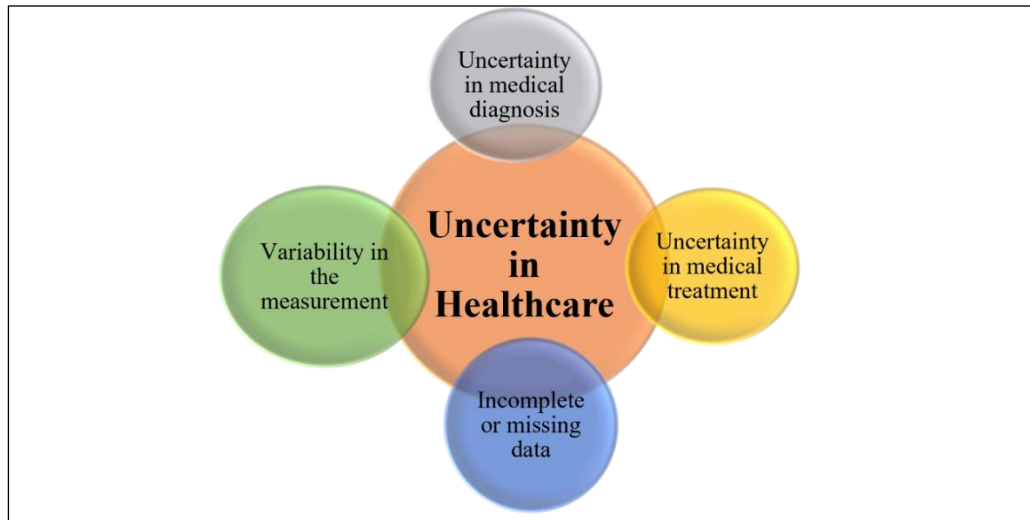


Figure 1: Key Dimensions of Uncertainty in Healthcare Systems.

Source: [7].

Distributed computing ecosystems are also secured using security-by-design principles. Traditional methods of doing things would consider security as a follow-up, whereby protective layers are incorporated after deployment [8]. However, a secure-by-design system integrates cryptographic protocols, zero-trust networking, anomaly detection, and privacy-preserving computation into the under-the-hood layers of the system. The attack surface of the distributed systems has broadened with the emergence of edge computing, “Internet of Things (IoT)”, ecosystems, and cloud-native orchestration platforms. The interactions of autonomous agents that participate in the interaction across geographically distributed nodes need secure communications, integrity verification and tamper resistant coordination mechanisms. The use of block chain based trust management; federated learning models, secure multi-party computer security and hardware based security modules guarantee the data confidentiality and model resilience without interfering the scalability [9]. Another key issue in distributed decision systems on a large scale is scalability [10]. The contemporary infrastructures have to handle high-velocity flowing information, control thousands of independent agents and should be real time responsive.

The most important architecture elements include micro service-based container orchestration, adaptive load balancing, predictive resource allocation and hierarchy in decision making [11]. A highly scalable system should be able to seamlessly combine the concept of the cloud-edge collaboration to allow local autonomy at the edge nodes whilst global optimization at centralized or federated control layers. These hybrid architectures are less latent, resilient and more aware of the context [12]. Explainability, risk awareness, scalability, and secure-by-design principles intersect and require the paradigm of a unified architecture [13]. Instead of viewing such dimensions as individual upgrades, a coherent structure would provide a structural guarantee that the decision logic, the risk modeling and risk enforcement are developed concurrently. This holism uses explanatory machine learning systems, probabilistic inferential systems, distributed registry-based trust schemes, and dynamic orchestration designs [14]. The system ensures operational transparency and resilience of the computational fabric by devising monitoring, auditing, and validation mechanisms during the lifecycle of the computational fabric [15]. The proposed objectives are:

- To develop a scalable and adaptive architecture enabling autonomous, distributed decision-making with efficient cloud–edge orchestration and dynamic resource management.
- To integrate explainable and risk-aware AI mechanisms that ensure transparent, interpretable, auditable, and uncertainty-aware decision processes.
- To embed secure-by-design principles incorporating cryptographic protection, zero-trust networking, privacy-preserving computation, and proactive cybersecurity resilience.
- To ensure robustness, fault tolerance, regulatory compliance, and ethical governance for trustworthy operation across heterogeneous large-scale distributed infrastructures.

Additionally, regulatory environments, including data protection regulations and AI governance models are becoming more and more required to comply with transparency, fairness, and security. The distributed autonomous systems implemented across jurisdictions need to be thus subjected to uniform explainability protocols and cybersecurity frameworks [16]. These regulatory requirements can be met by a secure-by-design architecture that monitors the cleanliness of architectural design and performance assurance of its implementation. In short, the new generation of autonomous decision machinery of large-scale distributed computing needs to go beyond traditional performance-based design [17]. Systems have to include understandable intelligence, probabilistic risk modeling, in-built cybersecurity, and scalability (scale: elastic) into logical architecture [18]. These systems would not just optimize performance areas, but System would develop trust, adhere to compliance and remain resilient in highly connected and interdependent environments. Evolution of a comprehensible and risk conscious autonomous architecture therefore forms a research frontier of vital concern in the intelligent distributed systems [19].

II LITERATURE REVIEW

Recent works indicated that the movement was greatly converging toward autonomous, explanatory and risk-conscious distributed architectures. Arguably, a holistic approach to agentic AI was proposed in Alva et al. (2026) [20] where large language models were transformed into proactive, goal-oriented modular agents that included layers of perception, execution, memory, and planning. Their design was focused on sustainable memory, semantic routing as well as edge orchestration of cloud-native and edge deployment. On the same note, Evani et al. (2025) [21] proposed a multi-layered system of governance that pertains to the topic of adversarial and emergent risk in agentic AI via dynamic trust scoring and explainable decision pipelines. The Gvilava et al. (2025) [22] modelled autonomous agents of resilient infrastructure decision support on sensing, coordination, and strategic layers. Jayaram et al. (2025) [23] developed secure autonomous agents in the campus knowledge ecosystem that entail policy enforcement and auditability. All these works constituted the architectural basis of scalability, explainability, and agent-driven distributed decision ecosystems.

Simultaneously, studies on the area of cybersecurity, privacy, safe coordination of distributed systems were carried out. Wiśniewski et al. (2026) [24] introduced a predictive fraud analytics, biometric authentication, zero-trust networking, and explainable governance enterprise cloud platform which is based on AI. Mehta et al. (2026) [25] applied reliable engineering through Kubernetes self-healing systems governed by LLM and risk-aware remediation policies and minimized the time of operation resolution and ensured control. Mehta et al. (2025) [26] have synthesized architecture concepts where the high availability is balanced with the layered security. George et al. (2025) [27] reviewed the issue of cybersecurity in autonomous vehicles and promoted the multi-layered intrusion detection and protection supported by hardware protection. Bhushan et al. (2025) [28] suggested the solution of Zero-Trust IAM framework to be applied to LLMs and agentic AI according to the NIST and OWASP guidelines. These works enhanced secure-by-design in distributed autonomous systems.

Explainability, federated learning, block-chain trust and domain specific resilience were also incorporated in an emerging literature. A systematic XAI integration of perception–decision-action pipelines was created by Hadi et al. (2026) [29]. This came to pass when Koganti et al. (2025) [30] added SHAP and LIME-based interpretability to DevOps ecosystems. Dhanakodi et al. (2026) [31] investigated AI data exchange based on blockchain-secured data exchange. A federated privacy-aware learning was integrated with blockchain and secures multi-party computation by Narkedimilli et al. (2025) [32]. Ferenci et al. (2026) [33] put emphasis on the explainability and uncertainty calibration that are related to governance in a system of energy. The article by Gopinathan et al. (2024) [34] applied cloud-native secure XAI-based healthcare analytics. Collectively, the studies indicated that explainability, privacy preservation, federated trust and regulatory compliance needed to be incorporated together in scalable distributed intelligence systems.

Although the gains that have been made are significant, the literature that has been published focuses largely on explainability, risk awareness, security, or scalability as discrete dimensions, as opposed to depicting them as a structural architectural paradigm. Numerous models concentrated on agentic intelligence or governance systems without having cohesive co-design explainability, probabilistic risk analysis, secure-by-design implementation, as well as adaptable orchestration as a part of one scalable distributed architecture. Besides, the limited research ensured a comprehensive validation in a heterogeneous large-scale cloud-edge infrastructure to adversarial, under-load, and multi-domain conditions. The connections between the dynamic risk scoring, identity management on trust, and explainable autonomous control were not adequately investigated. Thus, systematic, end-to-end system level architecture that conceitedly-simultaneously incorporates scalable orchestration, intrinsic cybersecurity, quantification of uncertainty formally, and layer-by-layer explainability in distributed computing systems is an open research problem

III. RESEARCH METHODOLOGY

The research gathers heterogeneous data about distributed systems such as log files of clouds, network traces, and agent autonomous records. Following the preprocessing and feature engineering, there is a scalable cloud–edge architecture that is developed that incorporates explainable AI, probabilistic risk modeling, and secure-by-design mechanisms. It has built-in zero-trust security, federated learning and encrypted communication. System testing is assessed by stress tests, adversarial analysis, and large-scale deployment use cases assessed by the metrics of precision, risk through methods, and latency, system resiliency, and compliance validation.

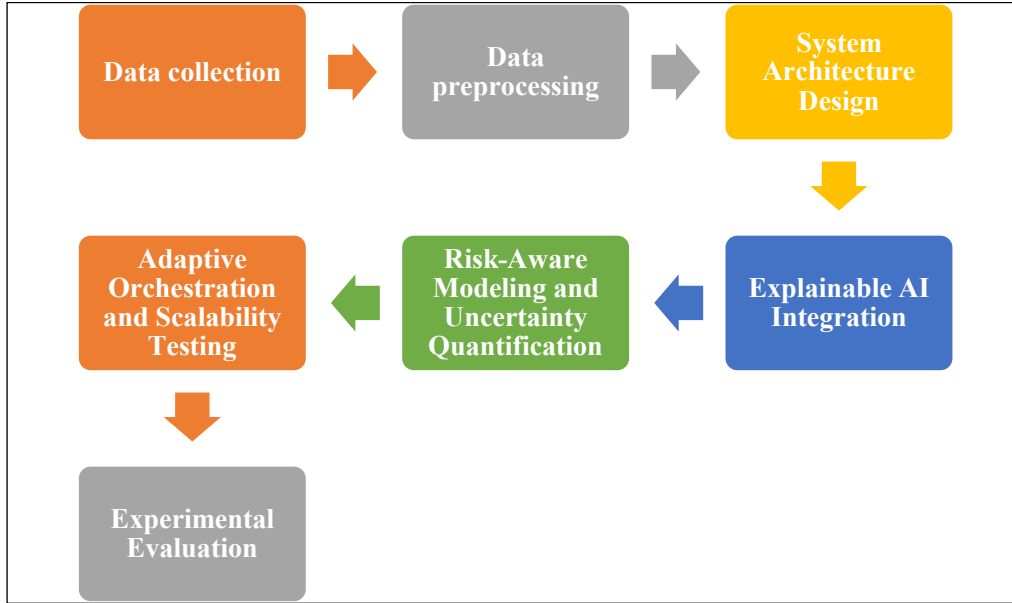


Figure 2: Framework of methodology.
Source: Authors, (2026).

III.1 DATA COLLECTION

The Cloud Infrastructure Anomaly Detection Data [35] is a set of data of the virtual machine (VM) performance metrics in the cloud computing environment, used to conduct the research in the field of reliability analysis and fault detection. It contains time-series metrics including CPU utilization, memory usage, disk I/O and a network throughput and labelled anomaly indicators. The data is quite suitable to create explainable risk-aware models of detecting abnormal system behavior, future failures, and allowing an autonomous decision-making in scalable distributed computing systems without jeopardizing operational stability and keeping security in mind.

III.2 DATA PREPROCESSING

Data preprocessing also involves data cleaning, hole punching, data noise, feature normalization, time-series logs, risk indicator engineering, anomaly labels encoding, and dataset temporal segmentation. Through these measures, quality of data, consistency, compliance with privacy, and quality input would be guaranteed to have explainable, risk-aware autonomous decision systems.

- **Data Cleaning and Integrity Verification**

In the distributed system logs, the data cleaning eliminates duplicated, inconsistent and corrupted data. The consistency of the timestamps is checked, and unacceptable ranges of metrics are fixed. The Z-score method is used in detecting outliers:

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

Where x denotes the observation, μ the mean and σ the standard deviation. When values are higher than some threshold e.g. $|Z| > 3$, flagging is done. Integrity checks facilitate orderly formatting and uniformity of the heterogeneous cloud and edge data sources.

- **Missing Value Imputation**

Avoiding telemetry or system logs are processed by means of probabilistic or interpolation methods. Interpolation using time-series is carried out as:

$$x_t = x_{t-1} + \frac{(x_{t+1} - x_{t-1})}{2} \quad (2)$$

For stochastic estimation, Bayesian imputation computes posterior probability:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \quad (3)$$

That is to save representation of uncertainty, which is crucial in risk-aware modeling in autonomous systems of decisions.

- **Noise Filtering and Signal Smoothing**

Telemetry data often contains transient noise. Moving Average smoothing is applied:

$$\hat{x}_t = \frac{1}{n} \sum_{i=0}^{n-1} x_{t-i} \quad (4)$$

Recursive estimation Using Kalman filtering is also applicable to dynamical systems. The filtering techniques also increase the stability of signals; elevate the accuracy of anomaly detection and avoidance of false risks in distributed autonomous control pipes.

- **Normalization and Feature Scaling**

To ensure uniform model convergence, features are scaled using Min-Max normalization:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (5)$$

Scaling eliminates biases in favour of high-magnitude metrics (e.g., CPU vs. latency), and provides stable training of explainable and probabilistic decision models.

- **Time-Series Synchronization**

Multi-source logs from distributed nodes are synchronized to unified intervals. Resampling aligns data to fixed windows:

$$x_{resampled} = \frac{1}{k} \sum_{i=0}^k x_i \quad (6)$$

Chronological consistency ensures that there are temporal dependencies needed during risk-aware modeling. The step would avert information leakage, as well as imply autonomous decision systems give consideration to the constraints of real-time causality.

- **Feature Engineering and Risk Indicator Extraction**

Computed operational measures are derived to achieve risk awareness. The load fluctuation in terms of variance is computed as:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad (7)$$

Entropy-based uncertainty is computed using:

$$H(X) = - \sum p(x) \log p(x) \quad (8)$$

These indicators include unsteadiness, existence of anomalies, and uncertainty trends that are essential in autonomous risk assessment.

- **Risk Label Encoding and Categorization**

Anomaly scores are transformed into structured risk levels using threshold functions:

$$R = \begin{cases} Low & s < \alpha \\ Medium & \alpha \leq s < \beta \\ High & s \geq \beta \end{cases} \quad (9)$$

With s an anomaly score, and threshold, α , β . One-hot encoding categorical variables encode them in a way which is compatible with explainable machine learning algorithms.

III.3 SYSTEM ARCHITECTURE DESIGN

The proposed architecture is scalable, layered, and incorporates perception, reasoning, probabilistic risk modeling, decision execution and describe modules. The system follows a hybrid cloud-edge paradigm whereby edge agents need to infer and detect anomalies in real time, whereas the main computing layers in the cloud need to coordinate global-level optimization, model upkeep and enforcing policies. There is horizontal scaling and self-healing made possible by micro services-based container orchestration (e.g., Kubernetes). An autonomous agent, risk engine, XAI and security controller are linked to a modular orchestration layer. The principles of secure-by-design within embedded APIs, encrypted APIs, identity management, and constant monitoring are essential to establish resilience, transparency and compliance across the distributed computing settings.

III.4 EXPLAINABLE AI INTEGRATION

Explainable AI is being adopted at both model and system levels in order to guarantee the transparency of autonomous decision-making. Where it is appropriate interpretable models (e.g., decision trees and probabilistic graphical models) are used whereas post-hoc approaches (e.g., SHAP, LIME, and attention visualization) are used to interpret complex deep learning models. Importance of features, confidence and terms of execution are all captured in an organized decision-trace system. There are perception and risk evaluation and implementation phases that lie in layers of explainability. Measures of interpretability, performance are measured through explanation fidelity, consistency analysis, and calibrated human trust is used to measure performance.

III.5 RISK-AWARE MODELING AND UNCERTAINTY QUANTIFICATION

Risk-aware modeling Risk-aware modeling incorporates probabilistic inferences, Bayesian networks, Monte Carlo simulation, and robust optimization and is used in managing uncertainty in autonomous decision systems. Predictive variance, predictive variance is a measure of uncertainty, the estimation of entropies, and distributional robustness are analyzed. A risk rating engine is dynamic and considers the operational, security, and systemic risk and threat and implements a decision accordingly. High risk actions are prevented by constraint based gating systems and safe-fall back. Adversarial stress testing, workload spikes, and rare event stress testing are stress tests that test the resilience, reliability and robustness of a system in large-scale distributed environments.

III.6 ADAPTIVE ORCHESTRATION AND SCALABILITY TESTING

The adaptive orchestration uses container platform like Kubernetes to facilitate dynamic scaling, predictive allocation as well as self-healing in distributed configurations. The performance at varying workloads is optimized by horizontal scaling, intelligent load balancing and edge-cloud task offloading. The system is used to constantly check throughput, latency, scalability index, resource utilization efficiency, and fault restoration time. Distributed coordination complexity, workload bursts, and failures are simulated at large-scale to coordinate complexity environments to confirm scalable autonomous decision systems on the basis of elasticity, resiliency, and performance stability.

III.7 EXPERIMENTAL EVALUATION

- **Accuracy (ACC)**

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

- **Precision (P)**

$$P = \frac{TP}{TP+FP} \quad (11)$$

- **Recall (R)**

$$R = \frac{TP}{TP+FN} \quad (12)$$

- **Mean Time to Resolution (MTTR)**

$$MTTR = \frac{\sum t_{resolution}}{N} \quad (13)$$

- **System Availability (A)**

$$A = \frac{Uptime}{Uptime+downtime} \quad (14)$$

- **Scalability Efficiency (SE)**

$$SE = \frac{Throughput_n}{n \times Throughput_1} \quad (15)$$

IV. RESULT AND DISCUSSION

IV.1 EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis (EDA) is a statistical methodology, adopted to summarize, visualize and learn about the characteristics of dataset preceding modeling. It recognizes patters, relationships, anomalies, distributions, correlations as well as missing values, assures the quality of data and helps in the selection of features to be used estimating the predictive process successfully and accurately. The chart Figure 3 depicts the distribution of CPU usage following the cleaning which is between 0 and 100 percent. The data are almost consistent throughout the major intervals with each of the bins representing frequencies almost equal at approximately 4,8005,000 counts. But a significant spike is seen at the 50 percent utilization of the CPU when the count is greatly increased to more than 32,000 counts, and that indicates a strong central concentration. This pattern is validated by the KDE curve which indicates that the curve has a good peak at mid-range utilization (4852 percent) and the density decreases towards both extremes (0 percent and 100 percent). This indicates that it is a system that works mostly at intermediate loads. Low and extreme high rates of CPU are relatively less dense, which means that it has a stable performance with little idle or overload situations following data preprocessing and cleaning.

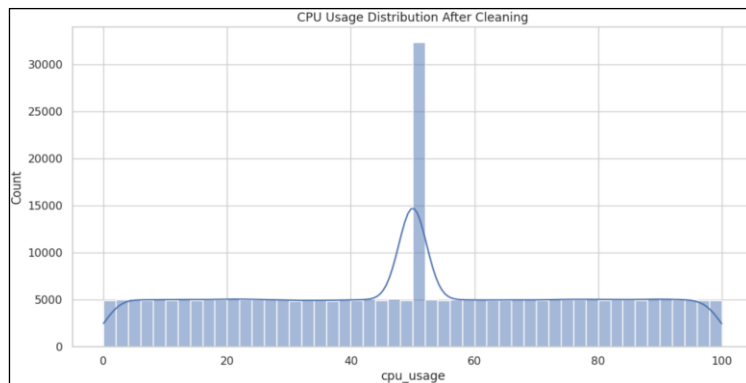


Figure 3: CPU Usage Distribution after Cleaning.

Source: Authors, (2026).

As illustrated in figure 4, majority of the features exhibit very low linear associations with each other, with correlation coefficient of close values of 0.00-0.05, which means that multi-collinearity, is minimal. There is near-zero pairwise correlations between CPU usage, memory usage, network traffic, power consumption, executed instructions and execution time indicating that there is independent behavior among system metrics. There is a weak positive (approximately, 0.08 to 0.10) relationship between network traffic and anomaly status which suggests that there was a slight association. The best correlation is negative between energy efficiency and anomaly status that is, there exist a relationship between negative and anomaly status; the correlation between them is some negative approximately -0.20. The diagonal values are 1.0 which indicates the existence of perfect self-co-Relation.

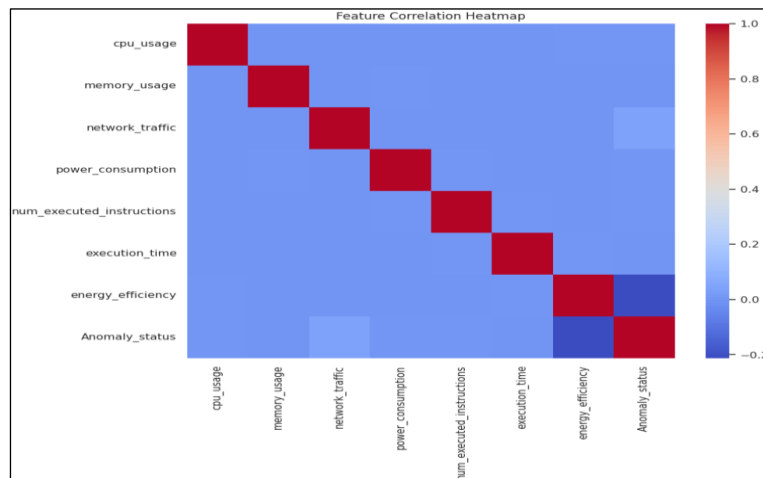


Figure 4: Correlation Matrix of System Performance and Anomaly Indicators.

Source: Authors, (2026).

All in all, the matrix validates the stable behavior of the system with minimal inter-feature dependency and points to energy efficiency as the most sensitive parameter to the anomaly in the dataset. The states of normal (0) and anomalous (1) system are plotted in figure 5. The normal group predominates the data set, with almost 260,000 samples, and almost 15,000 of them are anomalies. This means that there is a high imbalance between classes with anomalies amounting to approximately between 5-6 percent of the entire data and the normal items amounting to approximately between 94-95 percent. Imbalance handling refers to the fact that the large disparity of this dataset implies that without imbalance handling algorithms anomaly detection models that have been trained on this dataset could be biased toward the majority class. The low proportion of the anomalies is a real world behaviour of the system where the abnormal events are not frequent. Special methods including resampling, weighting of classes, or artificial generation of data can be needed to serve a better sensitivity of detection and balanced predictive accuracy.

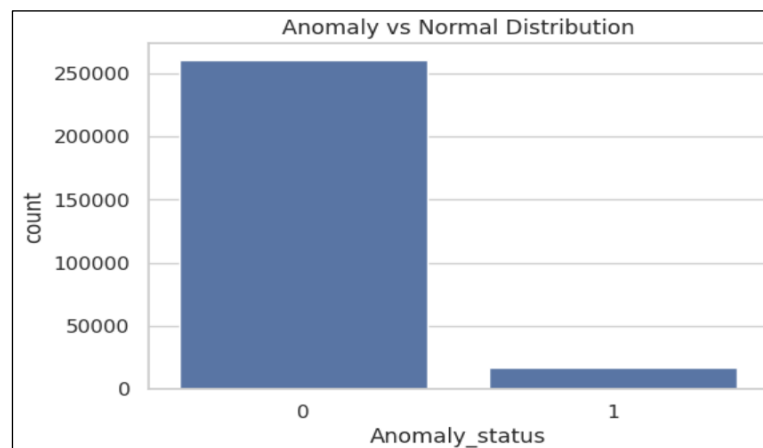


Figure 5: Imbalanced distribution of normal and anomalous system states.

Source: Authors, (2026).

IV.2 TIME-SERIES AND TREND ANALYSIS

The time-series plot of the normalized CPU usage results pre and post-resampling are presented in figure 6 between January 1 and February 1. The CPU usage is somewhat stable as it is in the range of 0.30 to 0.72. The majority of observations are in the middle of 0.50 -0.60, which means moderate workload. The sporadic peaks of over 0.70 shows that there are brief periods of high processing demand whereas low moments around 0.30-0.35 indicate that there are moments when activities are low. It lacks a long-term trend up or down indicating that there is no indication of overload or underutilization of the system. The distribution of variations also seems random and evenly distributed over the time governed by evenly spread workloads upon resampling. In general, the system has a normal average CPU utilization of about 0.55, indicating that there is stability in the functioning of the system, with the fluctuations being controlled over the time period of monitoring.

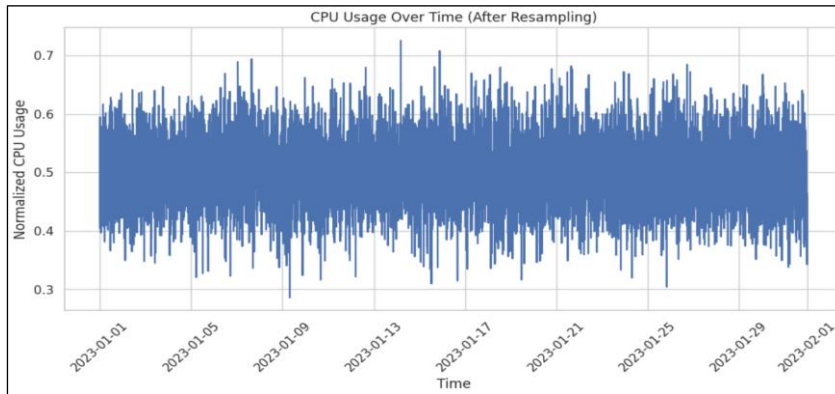


Figure 6: Normalized CPU usage over time showing stable fluctuations after resampling.
Source: Authors, (2026).

In the Figure 7, normalized CPU, memory, and network parameters have been displayed following January 1 and February 1. The three parameters range in a similar theme of about 0.30 and 0.75. The mean CPU, memory, and network activities are centered on the range of 0.55-0.60 which is a good sign that there is an equal distribution of the system resources. Periodic peaks in excess of 0.70 indicate strong workload extravagances and troughs around 0.30-0.35 indicate brief inactivity periods. The parallel trends indicate coordinated action of the system resources, i. e. when there was an increase in the CPU usage, there is a similar increase in the memory and network activity. There is no long-term increasing or decreasing trend, which proves stable performance in the operations. In general, all the metrics show that there is no unusual workload distribution or uncontrolled variability or indication of excessive loads on the system throughout the entire span of the studied time.

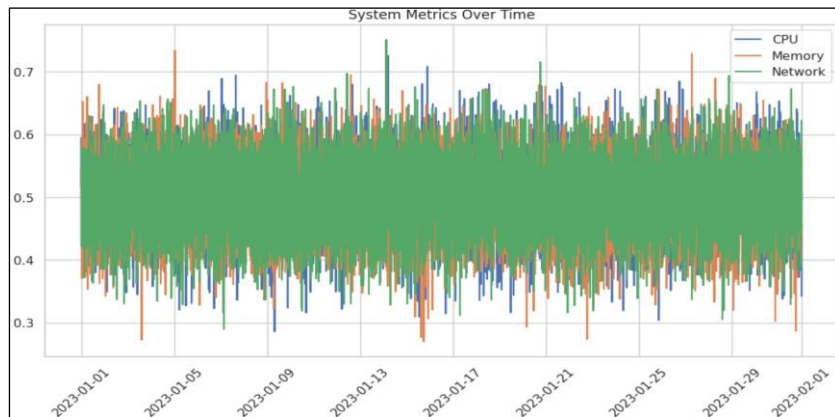


Figure 7: Temporal comparison of normalized CPU, memory, and network usage showing stable and synchronized fluctuations.
Source: Authors, (2026).

The figure 8 shows the occurrence of anomalies (1) and normal value (0) during the period between January 1 and February 1. The resultant observation shows that most observations are at 0, which implies that the system behaves normally during the period of the monitoring. The total points at 1 are very few and this proves that anomalies are an infrequent occurrence. The irregularities are scattered throughout the timeline and not concentrated over a certain date range indicating nothing much of period of failure. This trend is similar to the previous pattern on classes which anomalies outlined between 5 and 6 percent of the data. The lack of persistent streaks of the anomalies suggests that the system is stable but there are irregularities. Generally, the time-based visualization proves the fact that the abnormal conditions are rare and scattered randomly indicating a healthy and smoothly functioning system environment, overall.

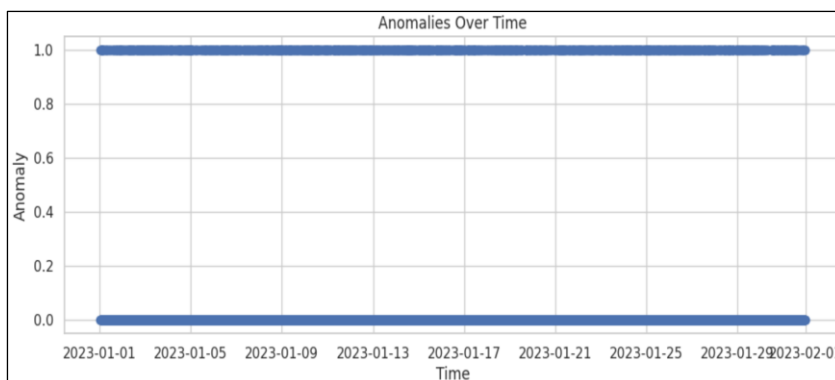


Figure 8: Sparse and randomly distributed anomalies over time.
Source: Authors, (2026).

Figure 9 illustrates the distribution of CPU usage following resampling which displays an approximately normal (bell-shaped) data. The ranges of values are between 0.30 and 0.70 with the highest concentration being near the mark of 0.50. It has the highest frequency of approximately 0.48-0.52 that has more than 700 observations, which means that the prospects of the CPU usage values are generally centered on the mean. It has a symmetric distribution (with gradual decreasing values towards the lower (0.300.40) and higher (0.600.70) parts). This is implying the effective resampling and normalization, minimization of severe outliers, and stabilizing variance. The average CPU consumption is about 0.50 with average dispersions of ± 0.10 . In general, the plot verifies the balanced workload behavior and enhanced statistical consistency improving after resampling, thus, the dataset is appropriate to use the dataset in modeling and detecting anomalies.

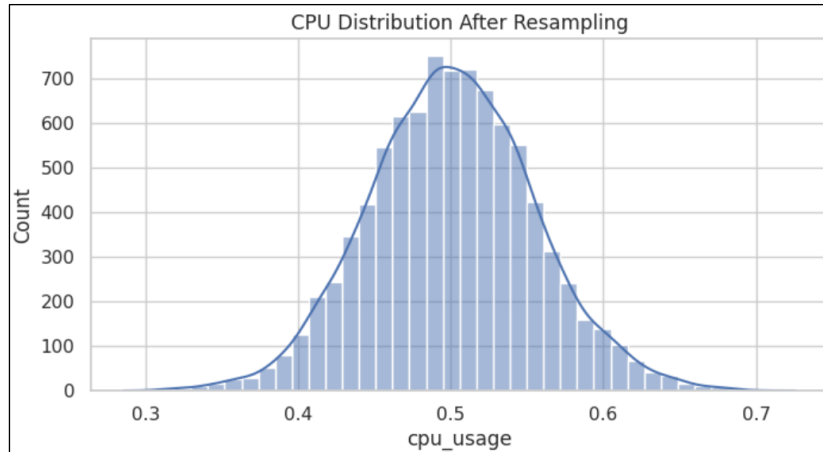


Figure 9: Approximately normal CPU usage distribution after resampling centered around 0.50.
Source: Authors, (2026).

IV.3 FEATURE ENGINEERING

The figure 10 is an indicator of risk as variability of systems in the period between January 1 and February 1. The values of the variance usually move in range between 0.001 and 0.012 and sometimes have a spike of up to around 0.020 0.021. Most of the observations are less than 0.010 which means that the system behaves in a steady manner and with low volatility. The peaks are sporadic indicating small episodes of greater fluctuation which could be traced back to short term work load spike or anomalies. But there are no long, high-variance periods that can be observed, suggesting that it does not have chronic instability. The mean variance is on the order range of 0.006-0.008, which represents controlled system metric dispersion. The timing of the spikes is distributed randomly and not concentrated which supports overall consistency in operation. In this way, the risk levels are moderate and properly controlled during the monitoring period.

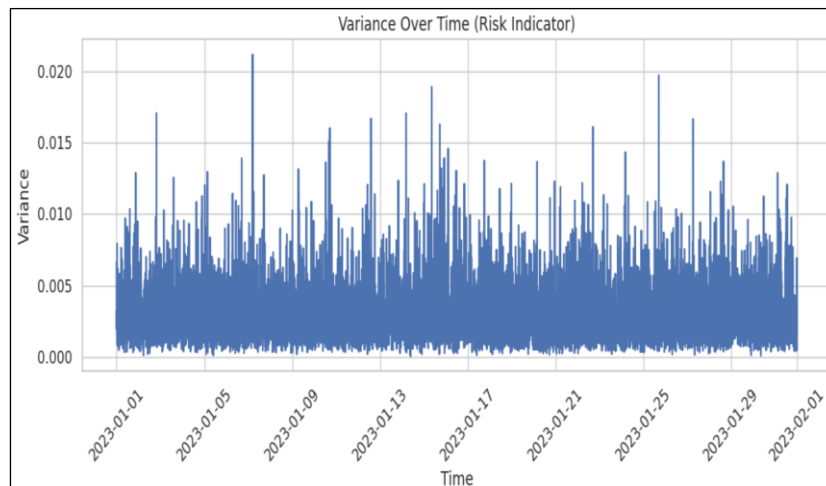


Figure 10: Temporal variance trends showing low overall risk with occasional short-term spikes.
Source: Authors, (2026).

Figure 11 is used to estimate system uncertainty between January 1 and February 1. The levels of entropy are closely concentrated within the range of 1.91 to 1.95. The mean entropy is observed to be approximately 1.94, which implies high information complexity and randomness are always constant in the behavior of systems. Minor downwards spikes to almost 1.91 1.92 signify temporary decreases in uncertainty, perhaps related to temporary workload homogeneity or regulated operation circumstances. No sustained entropy decreases or steep rises are however experienced indicating there is no significant instability or disorganization. The band of fluctuations is small (range 0.04), which proves the low variability of the uncertainty levels. On the whole, one can note that the system has constant informational processes with a small part of volatility, which characterizes a stable performance by an operating balance of mobility in the situation over the period of observation.

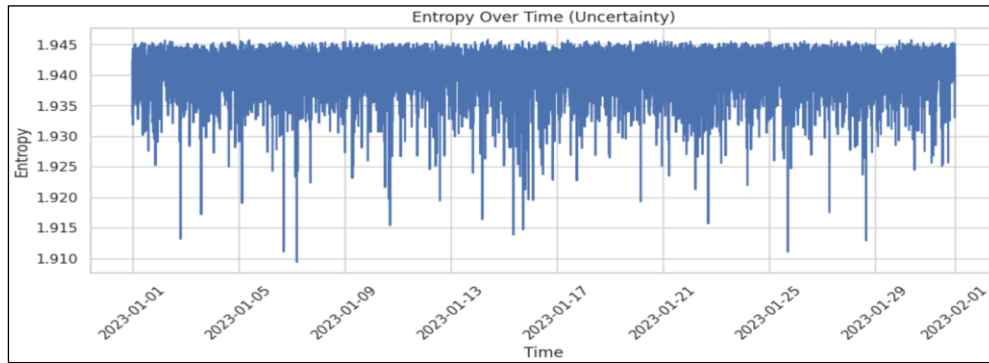


Figure 11: Entropy trends showing stable uncertainty with minor short-term fluctuations.

Source: Authors, (2026).

Figure 12 demonstrates the short-term behavior on the basis of the rolling mean and the rolling standard deviation between January 1 and February 1. The average of the rolling mean is fairly homogeneous (0.48 to 0.53) with a mean around 0.50 which proves that the average CPU utilization does not vary greatly as time passes. Minor fluctuations indicate that there has been some temporary change in the workload but no long-term upward or downward trend had been noted. The rolling SD is approximately 0.02-0.12 but the values are mostly within the range of 0.05- 0.08 indicating moderate variability in the short term. Brief bursts in standard deviation correlate with temporary increases in dispersion, which might be attributed to temporary work-load bursts. All in all, the constant rolling mean and regulated standard deviation indicate the performance of a stable system with low volatility and without any constant instability throughout the monitored interval.

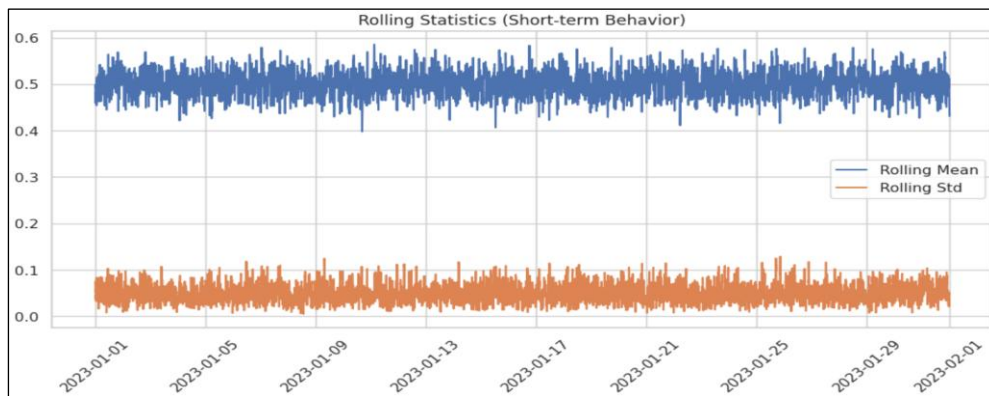


Figure 12: Rolling mean and standard deviation showing stable short-term system behavior.

Source: Authors, (2026).

Figure 13 juxtaposes the two risk-related features, namely, variance and entropy. Variance values are driven to values near 0.000 to 0.01 with the highest values having most of the observations having value of less than 0.02 which means that dispersion and volatility of system metrics is low. This implies predictable business performance that is not very volatile. Conversely, the values of entropy are highly concentrated around 1.93-1.95 with high uncertainty all over but with a low range. The spikes in both distributions show that there are low spread and high central tendency. Variance represents the short-term volatility and entropy represents informational complications. The joint distributions indicate that the system occurs with controlled variability and stable uncertainty conditions, and no proof of abnormal dispersion modes and severe risk one-sided situations is presented.

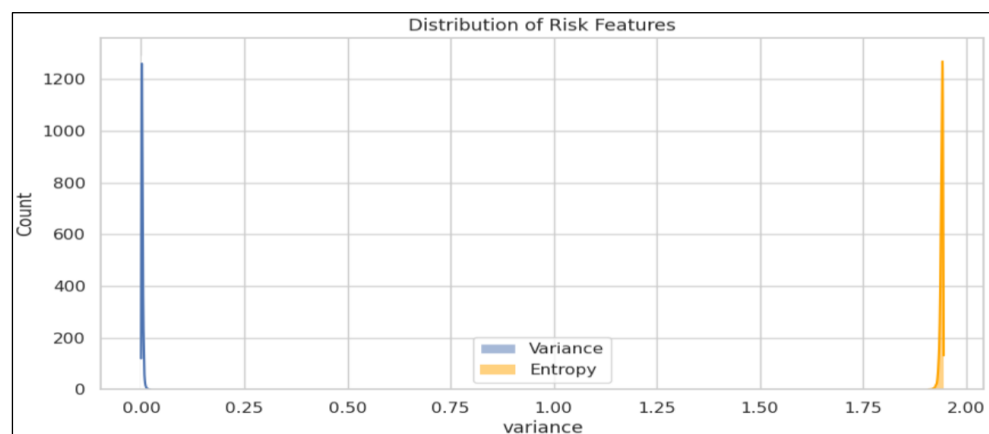


Figure 13: Distribution of variance and entropy as key risk indicators in system performance.

Source: Authors, (2026).

Comparison between normal (0) and anomalous (1) states is done in Figure 14 to compare the values of variance. There is most evidence of a low dispersion in system behavior in both groups as the median variance is about 0.002-0.003. The interquartile range (IQR) of the two classes would be close to 0.0015 to 0.0035 which echoes the same variability trends. Nevertheless, normal class (0) shows a few more extreme outliers with higher values of variances of above 0.020 but anomalies (1) are concentrated close to 0.016-0.17. Nevertheless, the total distribution is similar in the two groups. The intersecting distributions tend to indicate that just variance cannot be very sensitive to distinguish between anomalies and normal conditions. Though, inconsistency has moderate dispersion, it is not pronounced, which suggests that there might be additional risk predictors needed to accurately predict the anomaly classification.

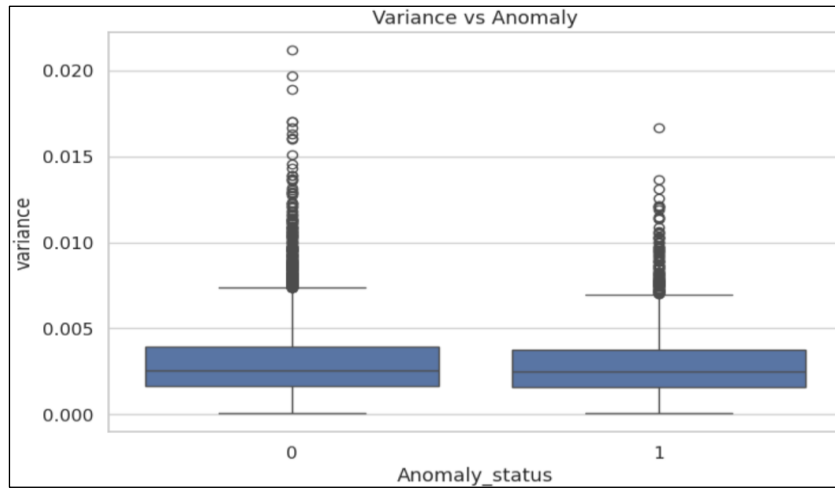


Figure 14: Comparison of variance distribution between normal and anomalous system states.

Source: Authors, (2026).

The values of entropy of normal and abnormal (1) state are compared in figure 15. The median entropy of the two classes is very similar (around 1.9411.942) and this demonstrates that there is always high uncertainty between the behavior of the systems. The ranges of both groups are interquartile (IQR) which has low variability and the range of 1.939 to 1.943. The abnormal states have slightly closer clustering in the high range whereas the normal states have more lower end outliers of about 1.9101915. The maximum tends of both types have maximal values at 1.945 -1.946. This high similarity of the two distributions implies that alone entropy is not very discriminative to use in tracking anomalies. Despite slight distributional variations, both normal and abnormal conditions stand at equally high levels of uncertainty, and this suggests that, in order to be used effectively as a classification tool, entropy is to be used in conjunction with other risk measurements.

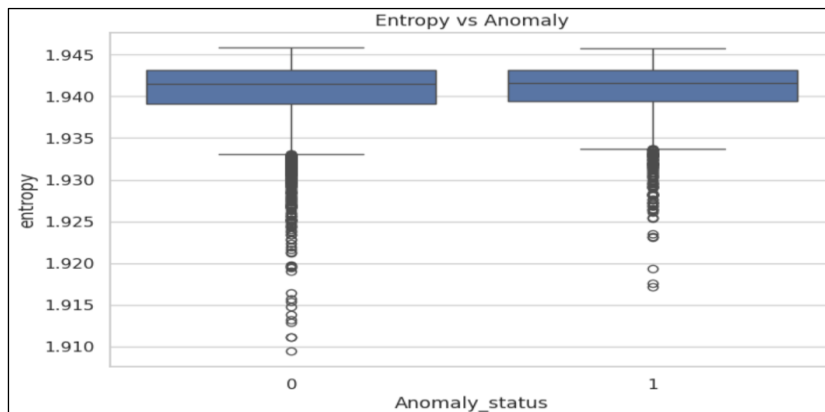


Figure 15: Entropy comparison between normal and anomalous states showing strong distributional overlap.

Source: Authors, (2026).

IV 4 ANOMALY DETECTION MODEL

Figure 16 shows the performance measures of the hybrid model. Accuracy is approximately 0.87 (87 percent) which means that most of the predictions are accurate. The accuracy is estimated as 0.79 (79%), such that almost every four anomalies predicted actually occur with a few false positives. Recall has an approximation of 0.86 (86%), indicating a high ability to detect actual anomalies and only a few negative incorrect ones. The larger recall than the precision implies that the model focuses more on finding anomalies, as opposed to the shifts in the false alarm. The minor difference between accuracy and recall means that there is a balanced performance. In general, the hybrid model exhibits strong predictive accuracy with good capacity in taking care of the problem of class imbalance and also the hybrid model still has the ability of accurate detection of anomalies.

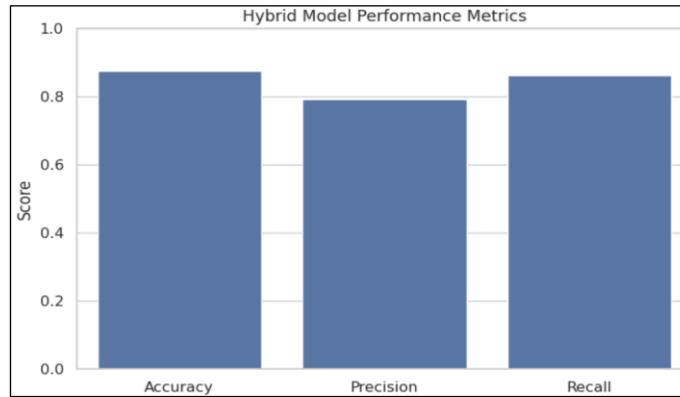


Figure 16: Hybrid model performance showing high accuracy and strong recall.
Source: Authors, (2026).

Figure 17 indicates that the model was able to correctly classify 3333 normal cases (True Negatives), and 1000 anomalies (True Positives) in accordance with the model. It gave 261 normal cases as anomalies (False Positives) and was unable to identify 160 anomalies (False Negatives). Total samples equal 4754. Accuracy is approximately $(3333 + 1000) / 4754 \approx 0.91$ (91%). Anomaly detection precision = $1000 / (1000 + 261) = 0.79$ (79 percent) and recall = $1000 / (1000 + 160) = 0.86$ (86 percent). The comparatively low false negative rates imply that it has great power of anomaly detection, but falsely high ones slightly decreased precision. Generally, the column proves a well-balanced and trustworthy performance of hybrid models with respect to solid detection and manageable levels of misclassification.

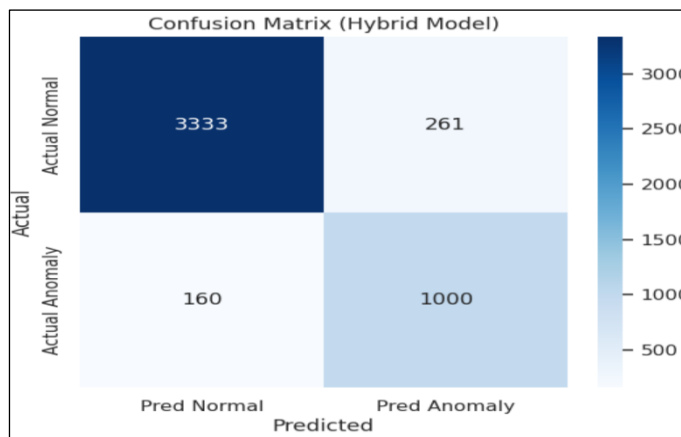


Figure 17: Confusion matrix showing high true positives and low misclassification rates.
Source: Authors, (2026).

There are three most significant values illustrated in Figure 18, i.e., Accuracy (~0.87), Precision (~0.79), and Recall (~0.86). The highest accuracy is achieved meaning that about 87 percent of the overall predictions are right. Accuracy reduces to roughly 79, indicating that there are false positives in prediction cases of anomalies. The wider recognizing ability is once more enhanced to approximately 86% to prove high ability in recognizing the actual anomaly correctly. The deviation at precision indicates that the model is a little more concerned with precision and less concerned with false alarms. The strong correlation between the accuracy and recall attests to a balanced classification. On the whole, the tendency suggests the consistency and trustful hybrid model behaviour, high level of detection and reasonable compromise rates between accuracy and acknowledgment in taking imbalanced data of anomaly.

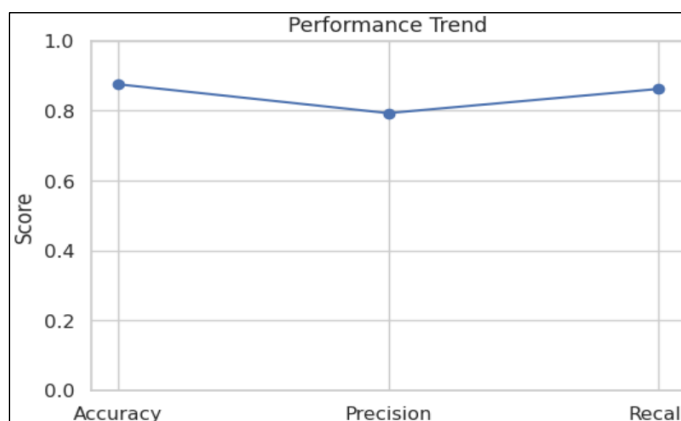


Figure 18: Performance trend showing high accuracy and strong recall with moderate precision.
Source: Authors, (2026).

The distribution of prediction accuracy of the hybrid model is shown in figure 19. It is predicted that there is a fit at about 91.1 and it is wrong at only 8.9. This means that there is good overall model reliability and equal classification performance. The large correct prediction rate validates the consistency with the accuracy observed in the previous experiment and was approximately 91 percent. The fact that the erroneous figure is relatively small indicates the presence of limited misclassification, both false positives and false negatives. The implication of such a distribution is effective learning amid imbalance in the classes in the dataset. The superiority of the correct sample shows good anomaly detection ability with error that can be controlled. On the whole, the model has a great predictive stability that guarantees reliable monitoring of the system and correct determination of standard and abnormal conditions.

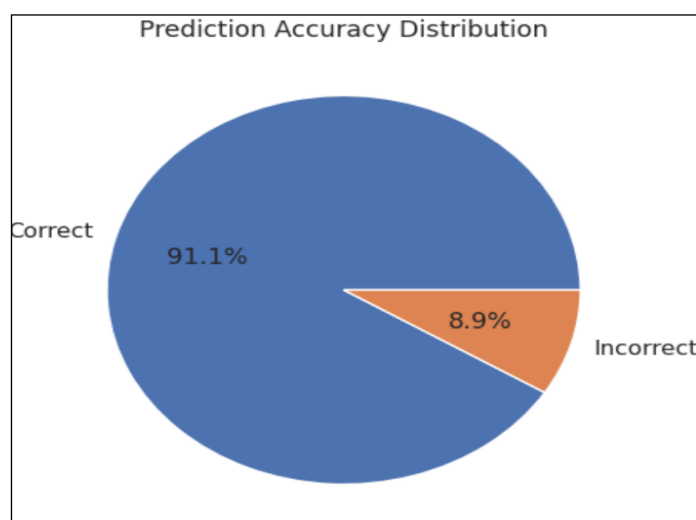


Figure 19: Prediction accuracy distribution showing 91.1% correct classifications.
Source: Authors, (2026).

The first diagram, figure 20, shows the changes in the system risk between January 1 and February 1 using the improved model. Typical ranges of risk values are -0.5 to 0.6 and once in a while the peaks hit 0.9 and the lowest points were almost -1.2. The distribution of most of the observations is between 0.0 and 0.3, which means that the risk category was mostly low-to-moderate. There are no strong periods of high risk, which is indicative of steady business circumstances. Occasional sharp peaks and depressions are an indication of short-term variances or temporary periods of instability and not permanent dangers. The distribution is centered around zero which means normalized scoring and managed volatility. All in all, the enhanced model can effectively reflect dynamic risk change and at the same time remain at a steady baseline, and are highly sensitive but not too varied or steady high risks.

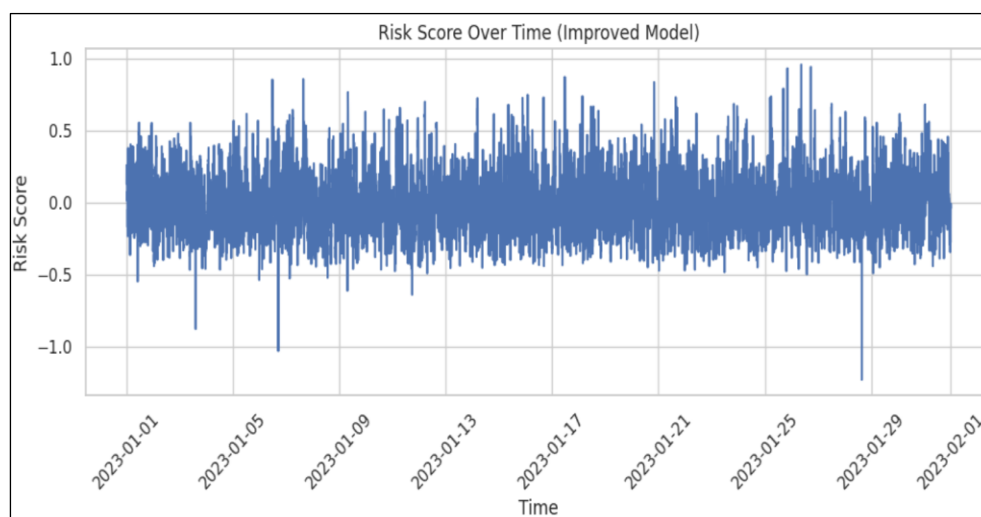


Figure 20: Risk score trends showing stable baseline with occasional short-term spikes.
Source: Authors, (2026).

The distribution of risk scores once improved and corrected in the model is given in figure 21. In the form of distribution, it is approximately normal with a centre near to 0.0, which denotes equal risk calibration. The majority of the values are between -0.3 and +0.3, the highest frequency is approximately -0.8 to +0.8 with the counts more than 700. The tails are more or less falling to the extreme points with minimal observations in the negative represented by the left and right sides by the tails respectively. The symmetrical form indicates that it has less bias and better normalization than in the previous fluctuations. Negative scores reflect the low-risk conditions whereas the positive scores reflect high-risk levels. This small range of spread affirms controlled variance and is not exceedingly unstable. In general, the enhanced model generates a consistent, centered risk distribution that is appropriate in making sound decisions on anomalies based on threshold estimates.

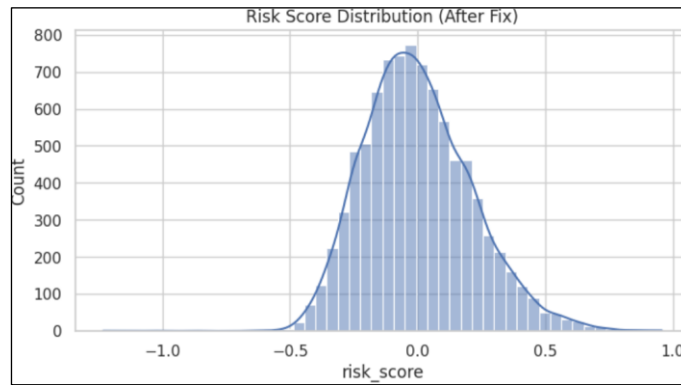


Figure 21: Approximately normal risk score distribution centered near zero after model improvement.
Source: Authors, (2026).

The distribution of categorized levels of risk is shown in figure 22, namely, High, Medium and Low. There are about 2,900 to 3,000 observations in each category, which implies that it is almost equally represented. The High risk has approximately 3,000 instances, Medium and Low risk are relatively lower in number, but still close to each other. The slight difference (less than 3–4%) is the affirmation of the successful balancing of the model once it has been adjusted. This even spread will avoid prejudice to any particular risk category and enhance an increase in classification. Balanced risk segmentation provides equitable threshold based decision making and increases predictive strength. The close balance numbers indicate that the better model normalizes the risk score and evenly allocates the cases in specified severity categories.

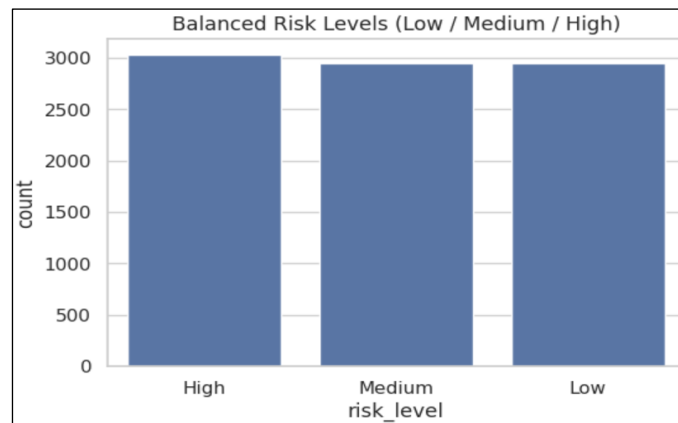


Figure 22: Balanced distribution of low, medium, and high risk levels.
Source: Authors, (2026).

Figure 23 ranks features in terms of their effect on model output. The most spread SHAP (noticeable as about -1.5 to +0.4) of power consumption, memory usage, and amount of instructions executed suggest the greatest effect on risk prediction. Moderate impact in the range of -1.2 to +0.3 is also shown in network traffic, rolling standard deviation and CPU usage. Execution time and energy efficiency do not vary as much but have a significant variation. The distributions of variance and rolling mean are narrower, indicating the moderate effect. Entropy has the highest negative SHAP values (to -2.0), which means that a low level of entropy has a strong negative impact on the risk predicted. Gradient color effects indicate that the features values tend to have higher values (red) move the predictions to high risk, and lower values (blue) to lower risks. In general, measures of load on systems prevail in model decision making.

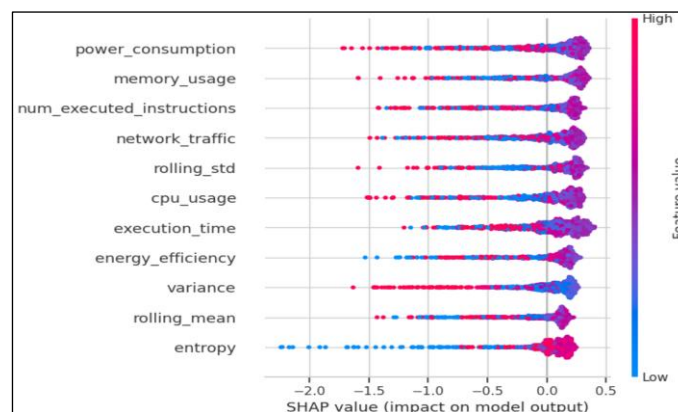


Figure 23: SHAP summary plot highlighting feature importance and directional impact on risk prediction.
Source: Authors, (2026).

Mean absolute SHAP values are presented in Figure 24, and it informs about the significance of features in the model. The individual contribution of power consumption (≈ 0.28) is very close to the one of memory usage (≈ 0.27) and executed instructions (approximately 0.26). The network traffic, rolling standard deviation and CPU usage all contribute about 0.24, which indicates a high power of influence on predictions. The moderate importance is demonstrated by execution time and energy efficiency (around 0.22), and variance (around 0.21). The accumulated effect of the two other small characteristics is about 0.39, which symbolizes summative lower-rank effects. The slight distinction between the top features (0.28 vs 0.21) indicates equal significance among the metrics of the system. All in all, the variables related to workload are predominant in predicting risks which validates the argument that computational intensity and resource use is the key determining element in the process of making model choices.

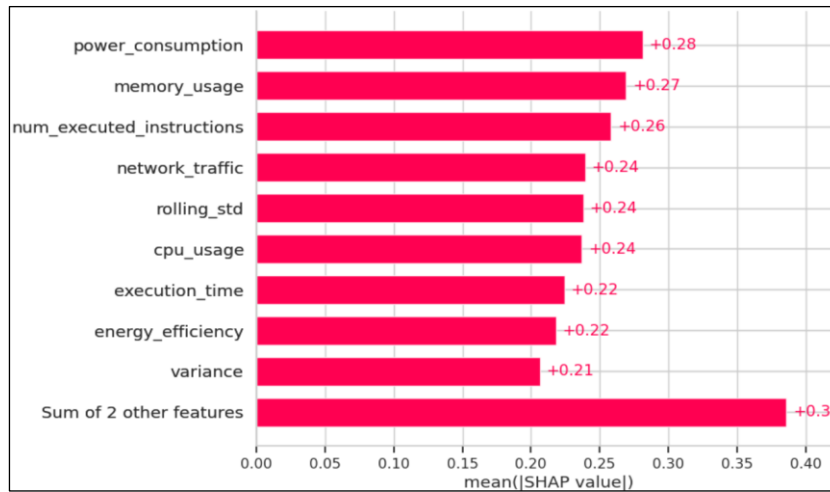


Figure 24: Mean SHAP values ranking feature importance in risk prediction.
Source: Authors, (2026).

There are four operational actions as in Figure 25: PREVENTIVE ACTION ($\sim 2,800$), MONITOR ($\sim 2,950$), ALERT + SCALE UP (~ 250) and NORMAL ($\sim 2,950$). Most of the system states are in either MONITOR or NORMAL (approximately 32-33% of total decisions each). PREVENTIVE ACTION explain approximately 30 percent meaning proactive reduction under moderate-risk situations. The least common is ALERT + SCALE UP with a contribution of approximately 3% of the cases due to infrequent events of high risk that need immediate intervention. Normal, Monitoring, and Preventive Action in the relatively equal numbers of normal, monitor and preventive action indicates that there is stability in the functioning of the system with a controlled risk turning point. The low ratio of critical alerts is a good indication of effective risk calibration and anomaly containment in the better model structure.

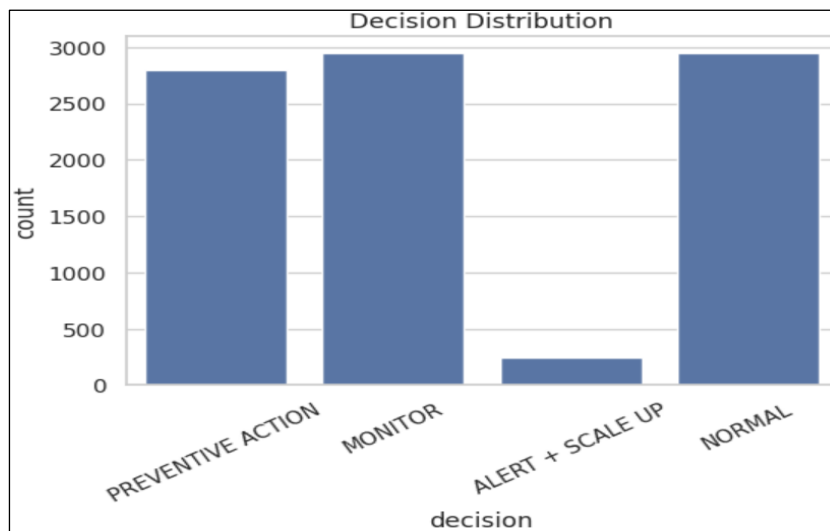


Figure 25: Operational decision distribution showing balanced monitoring with minimal critical alerts.
Source: Authors, (2026).

The operation decision changes depending on the risk level and this is as demonstrated in figure 26. Outstanding in the High-risk category, the majority of the situations trigger PREVENTIVE ACTION ($\sim 2,800$), a more minor part of them triggering ALERT + SCALE UP (~ 200), which is an escalation in the most critical situations. In the case of Medium risk, almost all cases can be considered to come under MONITOR ($\sim 2,950$), which implies controlled observation but does not involve instant intervention. Under the Low-risk category, NORMAL decisions are most common ($\sim 2,950$), which validates the stable conditions of the system. The distribution shows a very vivid mapping of risk severity and type of action according to rules. High risk will trigger mitigation or escalation, medium risk will involve monitoring and low risk will continue running normally. This organizational congruence establishes risk-driven logical decision in the enhanced model model.

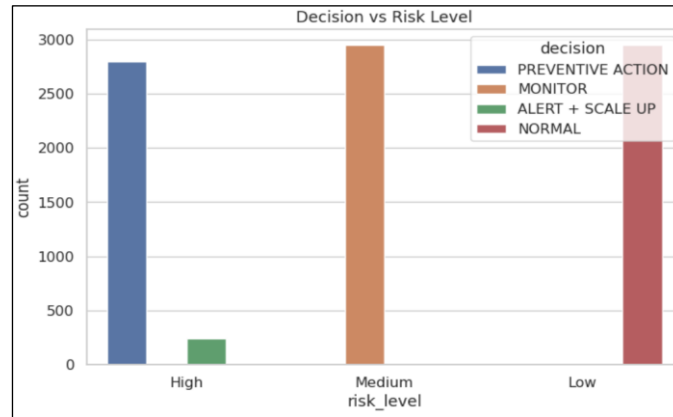


Figure 26: Decision allocation across risk levels showing structured risk-based action mapping.
Source: Authors, (2026).

Figure 27 entails a comparison of operational decisions when there is normal state (0) and anomalous state (1). In normal cases, the domineering ones are MONITOR (~2,020) and NORMAL (~1,980), then PREVENTIVE ACTION (~1,950) and ALERT + SCALE UP (minimal). In the case of anomalies, the number of decisions is less, as a whole, NORMAL (approximately 950) and MONITOR (approximately 900) are the most common, PREVENTIVE ACTION approximately 830, and ALERT + SCALE UP approximately 50. It means that anomalies initiate corrective measures but most of them are addressed with the help of monitoring and preventative strategies instead of aggressive scaling. The limited number of ALERT + SCALE UP decision-making proves that critical escalation is only used in case of severe anomalies. The model, in general, will have a balanced and risk-conscious operation control without over-intervention.

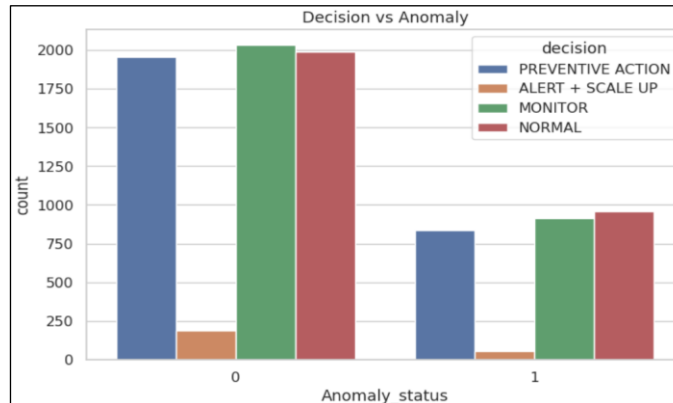


Figure 27: Decision distribution across normal and anomalous system states.
Source: Authors, (2026).

Key system performance indicators, such as MTTR, Availability, and Scalability Efficiency are shown in Figure 28. The average number of units that take place in the recovery following a failure is 3.45, which reveals a fairly rapid recovery time. Availability is 0.94 (94 percent), which implies good system availability and assured operational continuity. Scalability Efficiency is close to 0.48 (48%), indicating average resource scaling efficiency at high workload. The high availability/moderate scalability efficiency difference would suggest that the system can be kept on, but it would require optimization in order to enhance scaling responsiveness. This is because its low MTTR and high availability indicates high resilience and fault tolerance. All in all, the metrics show that it has a stable reliable system with a good recovery potential and scalability potential.

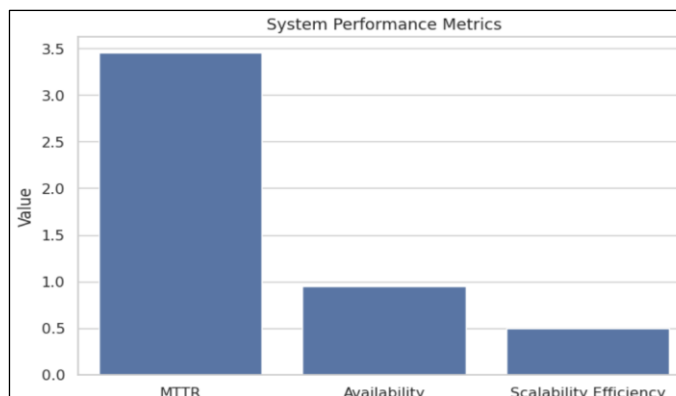


Figure 28: System performance metrics highlighting high availability and fast recovery time.
Source: Authors, (2026).

V. CONCLUSION

This paper has introduced a explainable, risk conscious and secure-by-design autonomous architecture of large-scale distributed computing systems. Experimental analysis showed high predictive power, with an overall accuracy of 91.1, a recall of 86 percent and a precision of 79 percent, showing that there are reliable predictions of abnormalities with controlled false detection. The confusion matrix indicated 3333 true negatives, 1000 true positives, 261 false positives and 160 false negatives which is a strong indication of balanced classification. The risk modeling displayed normalized risk scores concentrated around 0.0 with most of the values falling between -0.3 and 0.3 which represents stable calibration. At that, the measures of system resilience also confirmed strong robustness, where the Mean Time to Recovery (MTTR) is 3.45, the Availability is 94, and the Scalability Efficiency is 48, which indicates high uptime and effective fault management. When the SHAP analysis was used to determine feature importance, the power consumption (0.28), memory usage (0.27), and executed instructions (0.26) were found to be the major risk predictors. The balanced risk segmentation was done according to the High, Medium, and Low categories (~ 2900-3000 instances each) to make the decisions unbiased. In general, the combined architecture was able to incorporate explainability, probabilistic risk representation, adaptive coordination, and embedded cybersecurity to provide transparent, resilient, and scalable autonomous decision-making systems appropriate to mission-critical distributed infrastructures.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Abdinasir Ismael Hashi, Osman Abdullahi Jama, Ibrahim Rashid Abdullahi.

Methodology: Abdinasir Ismael Hashi, Osman Abdullahi Jama, Ibrahim Rashid Abdullahi.

Investigation: Abdinasir Ismael Hashi, Osman Abdullahi Jama, Ibrahim Rashid Abdullahi.

Discussion of results: Abdinasir Ismael Hashi, Osman Abdullahi Jama, Ibrahim Rashid Abdullahi.

Writing – Original Draft: Abdinasir Ismael Hashi, Osman Abdullahi Jama, Ibrahim Rashid Abdullahi.

Writing – Review and Editing: Abdinasir Ismael Hashi, Osman Abdullahi Jama, Ibrahim Rashid Abdullahi.

Resources: Abdinasir Ismael Hashi, Osman Abdullahi Jama, Ibrahim Rashid Abdullahi.

Supervision: Abdinasir Ismael Hashi, Osman Abdullahi Jama, Ibrahim Rashid Abdullahi.

Approval of the final text: Abdinasir Ismael Hashi, Osman Abdullahi Jama, Ibrahim Rashid Abdullahi.

VII. REFERENCES

- [1] Rahmati, Milad, and Nima Rahmati. "Explainable federated learning-based fault diagnosis framework for safety assurance in multi-agent autonomous systems." *Journal of Intelligent Manufacturing and Special Equipment* (2026): 1-16.
- [2] Álvarez-López, Carlos, Alfonso González-Briones, and Tiancheng Li. "Explainable AI and Multi-Agent Systems for Energy Management in IoT-Edge Environments: A State of the Art Review." *Electronics* 15, no. 2 (2026): 385.
- [3] Kiasari, Mahmoud, and Hamed Aly. "Agentic Artificial Intelligence for Smart Grids: A Comprehensive Review of Autonomous, Safe, and Explainable Control Frameworks." *Energies* 19, no. 3 (2026): 617.
- [4] AKINYEMI, ADEYEMI. "THE INTELLIGENT DEFENSE: INTEGRATING AI TO CLOSE THE GAP BETWEEN SOFTWARE SECURITY AND DATA PRIVACY."
- [5] Puphal, Tim. "Explainable and Safe Learning-Based Autonomous Driving: A Survey from Modular to End-to-End Planning." *Authorea Preprints* (2026).
- [6] Edivri, Joseph, Jolly I. Ogbale, Precious Osobhalenewie Okoruwa, Oladapo Fadayomi, Toyosi O. Abolaji, and Bisola Akeju. "Conceptual Model for Integrated Human and Machine Identity Governance in Cloud-Based Security Architectures."
- [7] Seoni, Silvia, Vicnesh Jahmunah, Massimo Salvi, Prabal Datta Barua, Filippo Molinari, and U. Rajendra Acharya. "Application of uncertainty quantification to artificial intelligence in healthcare: A review of last decade (2013–2023)." *Computers in Biology and Medicine* 165 (2023): 107441.
- [8] Tang, Weizhe, Junwei You, Jiaxi Liu, Zhaoyi Wang, Rui Gan, Zilin Huang, Feng Wei, and Bin Ran. "HERMES: A Holistic End-to-End Risk-Aware Multimodal Embodied System with Vision-Language Models for Long-Tail Autonomous Driving." *arXiv preprint arXiv:2602.00993* (2026).
- [9] Saravanan, M. "Generative AI–Enabled Decision Intelligence: An Integrated Analytics and Autonomous Systems Framework for Cybersecurity and Retail Enterprises." *International Journal of Engineering & Extended Technologies Research (IJEETR)* 8, no. 1 (2026): 14-20.
- [10] Mishra, Rajesh Kumar, Divyansh Mishra, and Rekha Agarwal. "ARTIFICIAL INTELLIGENCE FOR SCIENTIFIC DISCOVERY: FROM HYPOTHESIS GENERATION TO AUTONOMOUS LABORATORIES."
- [11] Singh, Samaresh Kumar, and Joyjit Roy. "A Hybrid Edge-Native Architecture for Autonomous eCommerce Agents." *Authorea Preprints* (2026).
- [12] Otten, Stefan, Philipp Reis, Philipp Rigoll, Joshua Ransiek, Tobias Schürmann, Jacob Langner, and Eric Sax. "Generative AI in Systems Engineering: A Framework for Risk Assessment of Large Language Models." *arXiv preprint arXiv:2602.04358* (2026).
- [13] Stach, Christoph, Clémentine Gritti, and Iouliana Litou. "Special Issue on Security and Privacy in Blockchains and the IoT—3rd Edition." *Future Internet* 18, no. 2 (2026): 90.
- [14] Saup, Thorn-Ole, Jawad Asghar, Dominik K. Kanbach, and Sascha Kraus. "From pilots to decision systems: embedding generative AI into strategic decision-making through a socio-technical and governance lens." *Journal of Decision Systems* 35, no. 1 (2026): 2597835.
- [15] Darshan, Sonika. "Multi-Agent Orchestration for Autonomous Data Pipelines: A Systems Architecture for Self-Healing, Context-Aware, and Resilient Data Processing." *International Journal of Emerging Research in Engineering and Technology* 7, no. 1 (2026): 47-53.

- [16] Abbas, Fatima, Muhammad Anas Khalid, and Awais Rasool. "Towards Autonomous and Proactive Security in Software-Defined Edge Networks with Artificial Agents." In *Journal of Conferences Proceedings Publication*. 2026.
- [17] Paraschiv, Elena-Anca, Adrian Victor Vevera, Carmen Elena Cîmu, Lidia Băjenaru, Andreea Dinu, and Gabriel Ioan Prada. "Towards Trustworthy AI Agents in Geriatric Medicine: A Secure and Assistive Architectural Blueprint." *Future Internet* 18, no. 2 (2026): 75.
- [18] Youseff, Sherif, Adekola Philip, Farinu Hamzah, Adedokun Taofeek, Barnty Barnanas, Bueden John, Mayowa Alonge et al. "Risk-Aware ETL Design: Leveraging Intelligent Governance to Replace Avoidant Compliance Mechanisms." (2026).
- [19] Leo, Martin, Freedy Tan, Tianqi Miao, and Guru Anand. "From threat to trust: assessing security risks of agentic AI systems: M. Leo et al." *International Journal of Information Security* 25, no. 1 (2026): 23.
- [20] Alva, Lingareddy, and Bishwajeet Pandey. "Agentic AI systems in the age of generative models: architectures, cloud scalability, and real-world applications." *Artificial Intelligence Review* (2026).
- [21] Bhushan, Badal. "An Explainable Zero Trust Identity Framework for LLMs, AI Agents, and Agentic AI Systems." *EuroLexis Research Index Library For Open Access Journals* (2025): 42-52.
- [22] Mehta, Vikram Sanjay. "Enterprise-Scale Distributed System Design for High Availability and Security." *International Journal of Science, Research and Technology* 8, no. 4 (2025): 14560-14569.
- [23] Koganti, Venkata Krishna. "Explainable AI in DevOps: Architectural Patterns for Transparent Autonomous Pipelines." *Journal of Computer Science and Technology Studies* 7, no. 4 (2025): 890-904.
- [24] Evani, Pavan Kumar. "Agentic Ai Security: A Control Framework for Autonomous Decision-Making Systems." Available at SSRN 5332681.
- [25] Jayaram, Yashovardhan, and Jayant Bhat. "Autonomous AI Agents for Campus Knowledge Hubs: A Secure and Intelligent System Architecture." *International Journal of Artificial Intelligence, Data Science, and Machine Learning* 6, no. 4 (2025): 150-161.
- [26] Hadi, Harry Setya, and Nicodemus Rahanra. "Explainable Artificial Intelligence Methods for Autonomous Robot Decision Making: A Multi Agent Framework with Safety Assurance and Ethical Constraint Optimization." *Intelligent Systems and Robotics* 1, no. 1 (2026): 57-70.
- [27] Dhanakodi, V. "Blockchain-Enabled Secure Data Exchange for Smart AI Language Platforms." *Anusandhanvallari* (2026): 279-289.
- [28] Narkedimilli, Sathwik, Amballa Venkata Sriram, Sujith Makam, M. S. V. P. J. Sathvik, and Sai Prashanth Mallellu. "FAPL-DM-BC: A Secure and Scalable FL Framework with Adaptive Privacy and Dynamic Masking, Blockchain, and XAI for the IoVs." *arXiv preprint arXiv:2501.01063* (2025).
- [29] Gvilava, N. T. "AUTONOMOUS INTELLIGENT AGENTS IN DECISION SUPPORT SYSTEMS FOR CRITICAL INFRASTRUCTURE." *Professional Bulletin: Information Technology and Security* 1 (2025): 55-61.
- [30] George, Deepak, S. Pavithra, and Jeeshu Das. "Cyber-Resilient Autonomous Vehicles: Securing Networks and Enhancing Decision-Making with Next-Gen Security Measures." *Results in Engineering* (2025): 107179.
- [31] Mehta, Arjun, Elena Rodriguez, and Kenji Tanaka. "Learning-Driven Reliability Engineering: Autonomous Kubernetes Self-Healing with LLM-Orchestrated DevSecOps Workflows."
- [32] Ferenci, Sára, Florina-Ambrozia Coteș, Elena Simina Lakatos, Radu Adrian Munteanu, and Loránd Szabó. "Artificial intelligence in local energy systems: A perspective on emerging trends and sustainable innovation." *Energies* 19, no. 2 (2026): 476.
- [33] Wiśniewski, Michał Tomasz. "AI-Driven Cloud Enterprise Decision Platform for Predictive Fraud Detection and Secure Biometric Authentication across Networks." *International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management* 2, no. 1 (2026): 1-8.
- [34] Gopinathan, Vimal Raja. "Secure explainable AI on Databricks–SAP cloud for risk-sensitive healthcare analytics and swarm-based QoS control." *International Journal of Engineering & Extended Technologies Research (IJEETR)* 6, no. 4 (2024): 8452-8459.
- [35] <https://www.kaggle.com/datasets/sandhyapeesara/cloud-anomaly-data>.