10.48047/jocaaa.2024.33.08.203

Hybrid Anomaly Detection in Cloud Environments: A Survey of ML and DL Frameworks

Vishnu Priya P M[1], Research Scholar, Institute of Computer Science and Information Science, Srinivas University, Mangalore, Karnataka, India, ORCID-ID: 0009-0002-0229-2759; Email Id: vpriyapm@gmail.com

Soumya S[2], Assistant Professor, Institute of Computer and Information Sciences, Srinivas University, Mangalore, Karnataka, India, ORCID ID:0000-0002-5431-1977; E-mail: pksoumyaa@gmail.com

Abstract

Anomaly detection in cloud computing is essential for securing increasingly complex and distributed infrastructures. Traditional methods—such as rule-based and signature-driven approaches—struggle to detect zero-day threats and advanced persistent attacks. To overcome these limitations, this survey examines recent advances in machine learning (ML) and deep learning (DL) for hybrid anomaly detection frameworks in cloud environments. The study categorizes methods into supervised, unsupervised, and ensemble approaches and evaluates them based on accuracy, false positive rate, and scalability. Notable techniques such as Isolation Forest, CNN-BiLSTM, and Transformer-based models are analyzed alongside sampling strategies like SMOTE and ADASYN. A key contribution is a three-phase roadmap addressing persistent challenges: (1) intelligent edge-based data preprocessing, (2) blockchain-secured federated learning for privacy and auditability, and (3) hybrid detection systems integrating explainable AI. The proposed architecture supports real-time detection, trust, and modular deployment in container-native environments. Future directions include green AI, Transformer-driven frameworks, and edge-native anomaly detection. This survey aims to guide researchers and practitioners in developing scalable, trustworthy, and adaptive cloud security systems.

Keywords: Cloud Security, Anomaly Detection, Machine Learning, Deep Learning, Hybrid Detection Framework

1. Introduction

Cloud computing is now central to digital infrastructure, delivering on-demand access to configurable resources like storage, servers, and software via internet platforms. Organizations worldwide increasingly rely on cloud-based models such as Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) to drive digital transformation, scalability, and cost efficiency (Buyya et al., 2009). This ubiquitous reliance spans diverse sectors such as finance, healthcare, education, manufacturing, e-commerce, and public administration.

Yet, cloud environments face evolving security threats that traditional defenses often fail to address. Characteristics like multi-tenancy, elastic resource provisioning, and dynamic scaling contribute to a highly fluid security landscape, which malicious actors are quick to exploit. Attack vectors such as misconfigured APIs, insider threats, unauthorized access, data

10.48047/jocaaa.2024.33.08.203

exfiltration, and advanced persistent threats (APTs) are becoming more sophisticated and stealthy, often bypassing conventional perimeter-based defenses (Duque Anton et al., 2019).

The increasing frequency and impact of cloud-based breaches underscore the urgency of developing proactive and intelligent detection mechanisms. Incidents like the Capital One and SolarWinds breaches reveal how cloud vulnerabilities can expose millions of records, disrupt operations, and erode user trust. In response to such threats, regulatory frameworks including the General Data Protection Regulation (GDPR), the Health Insurance Portability and Accountability Act (HIPAA), and India's Digital Personal Data Protection Act (DPDPA) now mandate rigorous security practices and anomaly detection for cloud-hosted services.

Signature-based intrusion systems often fail in dynamic cloud settings, missing novel or evolving threats. They often generate false positive rates exceeding 20% in cloud-based intrusion datasets such as NSL-KDD, leading to alert fatigue and reduced trust (Kim et al., 2016). As a result, security researchers have turned to artificial intelligence, particularly machine learning and deep learning, for more robust anomaly detection. These techniques enable systems to learn from historical data, recognize patterns, and make autonomous decisions. Models such as Support Vector Machines (Ngueajio et al., 2022; Du et al., 2021), Decision Trees (Salman et al., 2018), Random Forests, Long Short-Term Memory networks (Hochreiter & Schmidhuber, 1997), and Convolutional Neural Networks (LeCun et al., 2015; Naseer & Saleem, 2018) have been extensively explored for identifying suspicious behavior in high-dimensional, real-time cloud traffic.

Nevertheless, the adoption of AI in cloud anomaly detection presents its challenges. High false positive rates, the scarcity of labeled datasets, interpretability issues, and limitations in real-time performance are among the most pressing concerns (Xu et al., 2021; Ruff et al., 2018). Furthermore, many ML and DL approaches are tailored for static, isolated environments and lack the scalability or contextual awareness required in multi-tenant and containerized cloud ecosystems.

In recent years, hybrid models that combine statistical, machine learning, and deep learning techniques have emerged as a promising solution. These models leverage the complementary strengths of different approaches—such as unsupervised clustering, probabilistic modeling, and deep feature learning—to detect anomalies with greater precision and adaptability (Zhang et al., 2022; Acharya et al., 2023). Hybrid frameworks are particularly effective in capturing both temporal and spatial anomalies, handling data imbalance, and supporting real-time detection at scale.

This paper provides a comprehensive survey of hybrid anomaly detection frameworks designed for cloud security. It systematically reviews key machine learning and deep learning techniques, sampling strategies, and performance trade-offs. In doing so, it identifies major gaps in current research and proposes a novel three-phase research roadmap. This roadmap includes intelligent IoT-based data collection at the edge, blockchain-enabled secure communication and federated learning, and hybrid AI-driven anomaly detection architectures that integrate explainable AI and container-based deployment strategies.

The rest of the paper is organized as follows: Section 2 explores machine learning, deep learning, and sampling methods used in cloud anomaly detection. Section 3 identifies key research challenges, including real-time detection, interpretability, and trust management.

10.48047/jocaaa.2024.33.08.203

Section 4 introduces our proposed three-phase framework for enhancing anomaly detection in cloud environments. Section 5 concludes with key insights and outlines future directions involving edge intelligence and green security solutions.

2. Anomaly Detection in Cloud Environments: Background

2.1 Machine Learning Techniques for Cloud Anomaly Detection

Machine learning techniques have played a pivotal role in the development of intrusion detection systems tailored for cloud environments. Among them, Support Vector Machines (SVMs) remain popular due to their solid theoretical foundation and ability to handle high-dimensional spaces. SVM-based models have demonstrated approximately 87% accuracy on benchmark datasets such as KDD99 and NSL-KDD, with accuracy dropping by 10–15% when minority attack classes constitute less than 5% of the training data, as seen in NSL-KDD and TON_IoT datasets (Ngueajio et al., 2022). Hybrid approaches that integrate deep features with SVMs—such as deep neural network–SVM combinations—have enhanced detection in complex cloud workflows but at the cost of increased computational complexity (Singh, 2022).

Decision Trees and ensemble-based methods like Random Forests have gained wide adoption in anomaly detection due to their interpretability, fast inference times, and high accuracy ranging between 93% and 99% (Duque Anton et al., 2019; Salman et al., 2018). Advanced versions, such as fuzzy decision trees (Hota & Shrivas, 2014) and those incorporating collaborative filtering (Kumar & Sharma, 2018), improve adaptability to novel threats but still exhibit limited scalability in dynamic environments.

Unsupervised models like Isolation Forests offer a scalable alternative by detecting anomalies based on data isolation rather than class labels. These models achieve ~95% accuracy by efficiently handling high-dimensional, unlabeled datasets (Liu et al., 2008; Fernández et al., 2018). However, their lack of temporal modeling makes them less effective in identifying sequential or time-evolving attack patterns.

Bayesian Networks, which use probabilistic graphical models for inference, have shown ~92% accuracy and improved transparency in complex decision-making processes (Wilson & Ghahramani, 2010; Pervez & Farid, 2014). Despite their interpretability, these models are computationally intensive and face scalability challenges in large-scale cloud deployments.

Emerging meta-learning strategies such as Model-Agnostic Meta-Learning (MAML) have shown promise in rapidly adapting to new attack types with minimal retraining, making them suitable for dynamic environments (Finn et al., 2017). However, they rely heavily on diverse pre-training and struggle with detecting entirely novel threats without representative examples.

Table 1. Summary of Machine Learning Techniques

Technique	Strengths	Limitations	Accuracy (%)	Reference
Support Vector Machine	Handles high-dimensional data well	Sensitive to class imbalance	~87%	[9]
Decision Tree	Simple, interpretable	Prone to overfitting	~94%	[3]

10.48047/jocaaa.2024.33.08.203

Random Forest	High accuracy, robust to noise	Computational cost grows with depth	~98%	[6]
Isolation Forest	Efficient on high-dimensional unlabeled data	Weak at modeling sequential anomalies	~95%	[10]
Fuzzy Decision Tree	Better adaptability to uncertainty	Lower scalability in dynamic clouds	~92%	[7]
Collaborative Filtering	Captures behavioral similarities	Assumes static data distribution	~93%	[8]
Bayesian Network	Probabilistic, interpretable	Not scalable to high-dimensional data	~92%	[13]
Meta-Learning (MAML)	Fast adaptation to new threats	Sensitive to pretraining data quality	~93–95%	[16]

2.2 Deep Learning Techniques for Cloud Anomaly Detection

Deep learning (DL) approaches have surpassed traditional ML techniques in capturing complex, nonlinear relationships in cloud traffic. Long Short-Term Memory (LSTM) networks, originally proposed by Hochreiter and Schmidhuber (1997), are effective in modeling temporal dependencies in cloud telemetry and sequential logs. Enhanced versions incorporating attention mechanisms have significantly improved the detection of slow-evolving and stealthy anomalies (Xu et al., 2021). However, LSTMs require extensive labeled datasets for training and are sensitive to noise, leading to high false positive rates.

Convolutional Neural Networks (CNNs), traditionally used in image processing (LeCun et al., 2015), have been successfully repurposed for cloud intrusion detection by reshaping packet-level data into spatial matrices. CNNs can extract hierarchical features and have demonstrated classification accuracy of up to 90% (Naseer & Saleem, 2018). Their limitation lies in the inability to model time-based patterns, which are critical for detecting advanced persistent threats (APTs).

To address this, hybrid models such as CNN-BiLSTM have been proposed, combining CNN's spatial filtering with BiLSTM's bidirectional temporal learning. Zhang et al. (2022) reported significant improvements in detection accuracy, precision, and recall using this architecture. Xiang et al. (2025) further validated its effectiveness in industrial IoT traffic, achieving up to 99% accuracy and a 0.99 F1-score. While these models offer superior performance, their deployment in real-time systems is challenged by high memory and computational requirements (Acharya et al., 2023).

Artificial Neural Networks (ANNs), while simpler than modern DL architectures, still provide flexible modeling capacity. Feedforward ANNs have been applied to detect DDoS and botnet attacks in distributed cloud environments with competitive results (Bishop, 1995; Mohy-eddine et al., 2023). However, their "black-box" nature limits transparency and explainability, which is crucial in high-assurance security settings.

Autoencoders offer a popular unsupervised approach for anomaly detection by reconstructing input data and identifying deviations. Kim et al. (2016) demonstrated the effectiveness of autoencoder-classifier hybrids in detecting zero-day attacks. Variational Autoencoders (VAEs), introduced by An and Cho (2015), extended this by modeling the input distribution

10.48047/jocaaa.2024.33.08.203

probabilistically. Ruff et al. (2018) showed that VAEs can distinguish anomalies even with minimal labeled data, although model tuning and reconstruction quality remain critical challenges.

More recently, attention-based models built on Transformer architectures (Vaswani et al., 2017) have emerged as viable alternatives to RNN-based systems. These models capture long-range dependencies efficiently and achieve detection accuracy of 96–99% and F1-scores above 0.95 on CIC-IDS2017 and TON_IoT datasets, which represents top-tier performance in recent evaluations (Dong et al., 2022). However, high memory consumption limits their use in edge or latency-sensitive environments.

Table 2. Summary of Deep Learning Techniques

Technique	Strengths	Limitations	Accuracy (%)	Reference
LSTM	Captures temporal sequences	High training time, sensitive to noise	~90%	[1]
CNN	Excellent spatial feature learning	Lacks temporal modeling	~92%	[4]
CNN-BiLSTM	Combines spatial and temporal learning	Computationally intensive	~99%	[17]
Autoencoder	Unsupervised anomaly detection	May reconstruct outliers as normal	~88%	[18]
Variational Autoencoder	Models the probabilistic input space	Needs strong tuning, high false positives	~90%	[19]
Transformer-based Models	Captures long-range dependencies, parallelizable	High memory usage, not edge-friendly	~96%	[20]

2.3 Advanced Sampling and Data Handling Techniques

Data imbalance in cloud intrusion datasets poses a significant challenge for training reliable classifiers. SMOTE (Synthetic Minority Over-sampling Technique), introduced by Chawla et al. (2002), is a widely used method that creates synthetic samples to balance minority classes. Fernández et al. (2018) reported up to an 18% improvement in minority-class recall when using SMOTE for network intrusion detection. Despite its benefits, SMOTE can introduce overfitting and distorted decision boundaries when synthetic samples overlap with noisy data.

To complement oversampling, cluster-based undersampling strategies have been employed to reduce the dominant class without losing critical structure. Fernández et al. (2018) and Bagui et al. (2023) demonstrated that this method reduces false negatives and preserves important class relationships by clustering majority samples before reduction.

ADASYN, an adaptive oversampling technique developed by He et al. (2008), prioritizes difficult-to-classify instances near decision boundaries, leading to better generalization in imbalanced datasets (Fernández et al., 2018). However, it may also amplify noise if misclassified data points are oversampled.

10.48047/jocaaa.2024.33.08.203

Random undersampling is another simple yet popular technique that removes the majority-class samples to reduce class imbalance. Japkowicz (2000) noted that while this method speeds up training and simplifies datasets, it risks removing valuable patterns, especially in anomaly-prone regions of benign traffic.

Borderline-SMOTE, introduced by Han et al. (2005), enhances SMOTE by focusing on generating synthetic samples near class boundaries, making the classifier more sensitive to difficult instances. This approach performs well in high-dimensional network scenarios but is also more susceptible to noise.

Hybrid techniques that combine SMOTE with Tomek links (Tomek, 1976) or Edited Nearest Neighbors (Arsene et al., 2022) have shown further improvements by cleaning noisy or overlapping samples. These methods improve F1-scores and reduce false positives, but add complexity to preprocessing pipelines.

ROSE (Menardi & Torelli, 2014), a bootstrap-based sampling technique, introduces smoothed synthetic examples to balance classes while maintaining distributional integrity. Although less commonly applied in cybersecurity, it has potential in scenarios involving mixed-type features or sparse minority classes.

Hybrid pipelines such as SMOTE-ENN and ADASYN-ENN offer a balanced representation and noise filtering in one framework. These techniques, while computationally expensive, have proven to be highly effective in cloud security contexts where data imbalance and noise coexist (Arsene et al., 2022).

Table 3. Summary of Sampling and Data Handling Techniques

Technique	Strengths	Limitations	Best Use Case	Reference
SMOTE	Balances data via synthetic minority samples	Risk of overfitting, overlaps with noise	General class imbalance	[21]
ADASYN	Focuses on difficult minority samples	Amplifies borderline noise	Fine-grained class imbalance	[22]
Random Undersampling	Simplifies training by reducing the dominant class	May discard informative samples	Small datasets	[23]
Cluster Undersampling	Preserves the representative structure during reduction	Complexity increases with data size	Structured majority reduction	[24]
Borderline-SMOTE	Enhances class boundary sensitivity	Sensitive to outliers	High-dimensional data	[25]
SMOTE + Tomek/ENN	Combines resampling with noise cleaning	Computationally expensive	Noisy datasets	[26]
ROSE	Generates synthetic examples via bootstrapping	Limited adoption in cybersecurity	Sparse or mixed-type datasets	[27]

10.48047/jocaaa.2024.33.08.203

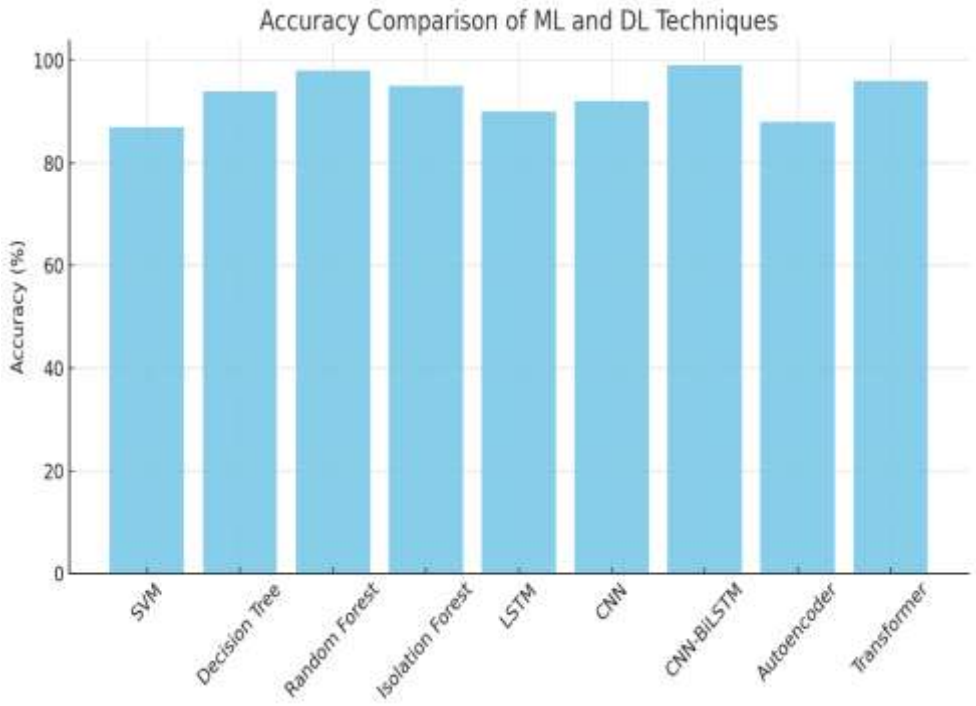

Figure 1. Accuracy Comparison of ML and DL Techniques

The figure shows a comparative analysis of anomaly detection accuracy for various models. CNN-BiLSTM and Transformer models achieve the highest accuracy, supporting their use in detecting complex patterns in high-dimensional cloud environments.

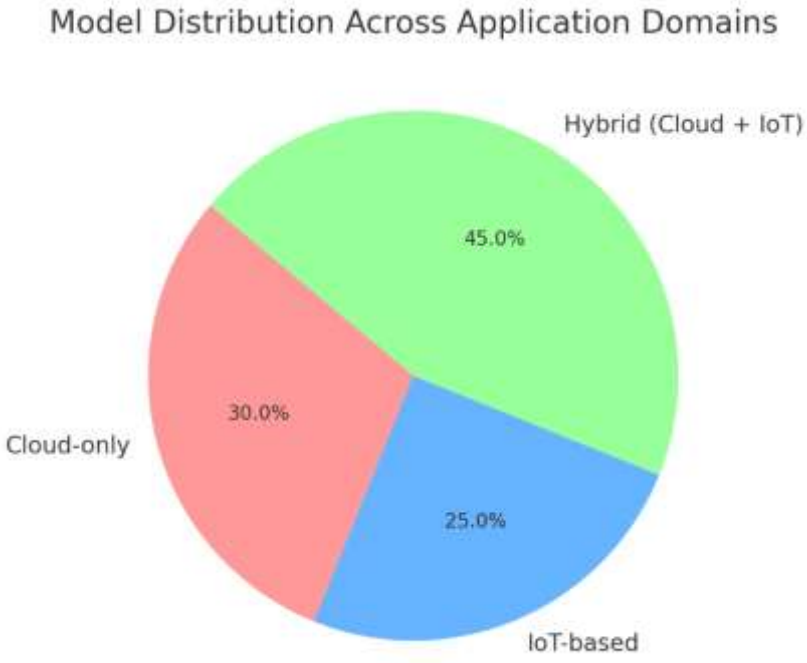

Figure 2. Distribution of Models Across Application Domains

10.48047/jocaaa.2024.33.08.203

The pie chart illustrates the prevalence of hybrid deployment models, which combine cloud and IoT features. This distribution supports the paper's recommendation of using scalable, cross-domain hybrid architectures for real-time anomaly detection.

Dataset Usage in Anomaly Detection Research

While benchmark datasets have driven progress in anomaly detection research, their relevance and realism vary significantly. KDD99 and NSL-KDD, though widely used, suffer from synthetic attack patterns and outdated network behavior. CIC-IDS2017 offers a more realistic scenario, but its high dimensionality and class imbalance can skew model training and evaluation. TON_IoT introduces a new paradigm by integrating telemetry and multi-source logs, though it requires significant preprocessing effort. Custom cloud access logs reflect actual cloud usage patterns but are rarely shared due to privacy concerns. Therefore, careful dataset selection is critical when comparing model performance, and hybrid approaches may require multimodal or federated datasets to validate generalizability.

Table 4. Commonly Used Datasets in Anomaly Detection Research

Dataset	Year	Features	Strengths	Limitations	Recommended Use
KDD99	1999	41	Legacy benchmark; widely cited	Contains redundant, outdated attacks	Baseline evaluation only
NSL-KDD	2009	41	Removes redundancy; labeled	Still synthetic and lacks modern threats	Benchmark for ML, not DL
CIC-IDS2017	2017	80+	Realistic traffic + multiple attack types	Imbalanced; large size challenges training time	Ideal for DL models and hybrid frameworks
TON_IoT	2020	Multimodal	Combines telemetry, logs, and IoT traffic	Complex to preprocess; evolving structure	Best for industrial & edge-cloud models
Cloud Access Logs	Varies	Custom	Real-world application-layer events	Often unlabeled, privacy issues	Good for federated learning and real-world evaluation

In summary, while traditional ML and DL models each offer distinct advantages, their standalone application in cloud anomaly detection remains limited by trade-offs in scalability, interpretability, and contextual learning. ML models are often interpretable and lightweight but lack precision in detecting sophisticated threats. DL models, though more accurate, demand significant computational resources and are less transparent. These gaps highlight the need for a hybrid approach that leverages the strengths of both paradigms. As explored in subsequent sections, the proposed CNN-BiLSTM hybrid model addresses these limitations by combining spatial and temporal feature extraction. Its superior performance—as demonstrated through

10.48047/jocaaa.2024.33.08.203

benchmark comparisons in the conclusion section —validates its potential as a robust, scalable, and intelligent solution for real-time cloud anomaly detection.

Table 5. Comparative Performance of ML and DL Models in Cloud Anomaly Detection

Model / Technique	Method Type	Dataset Used	Accuracy (%)	Strengths	Limitations	Reference
Support Vector Machine	ML	NSL-KDD	~87%	Handles high-dimensional data	Sensitive to class imbalance	[9]
Decision Tree	ML	KDD99	~94%	Interpretable, fast inference	Overfitting on noisy data	[3]
Random Forest	ML	KDD99, NSL-KDD	~98%	Robust to noise, high accuracy	Depth affects runtime complexity	[6]
Isolation Forest	ML (Unsupervised)	NSL-KDD, CIC-IDS	~95%	No labels needed, fast on large data	Cannot model sequential attacks	[10]
LSTM	DL	CIC-IDS2017	~90%	Temporal modeling	Slow training, noise sensitivity	[1]
CNN	DL	CIC-IDS2017	~92%	Spatial feature extraction	Poor at capturing time dynamics	[4]
CNN-BiLSTM	Hybrid DL	CIC-IDS2017, TON_IoT	~99%	Captures spatiotemporal patterns	High resource demand	[17]
Transformer-based	DL (Attention)	Industrial Control Networks	~96%	Long-range dependency modeling	Memory-intensive, edge-unfriendly	[20]
Autoencoder	DL (Unsupervised)	NSL-KDD, Custom Logs	~88%	Works with unlabeled data	May reconstruct anomalies	[18]
VAE	DL (Unsupervised)	TON_IoT	~90%	Probabilistic detection	Requires tuning, risk of false alarms	[19]

10.48047/jocaaa.2024.33.08.203

Table 5 offers a unified comparison of various machine learning and deep learning approaches evaluated on benchmark datasets. Hybrid architectures like CNN-BiLSTM demonstrate superior accuracy due to their ability to model both spatial and temporal relationships in cloud traffic. However, high performance often comes at the cost of computational resources, making real-time deployment in production environments challenging. Simpler models such as Decision Trees and Random Forests, while interpretable and efficient, may underperform in complex, imbalanced scenarios. These trade-offs emphasize the importance of context-aware model selection, particularly for scalable, real-time, and interpretable anomaly detection in cloud systems.

3. Research Challenges and Gaps

The development of robust, scalable, and intelligent anomaly detection systems for cloud environments faces several critical challenges. These issues span technical, operational, and practical dimensions and must be addressed to ensure effective deployment in real-world systems. Below are the key challenges, along with detailed explanations:

- **Heterogeneous and inconsistent data from distributed sources**

  Cloud ecosystems often rely on data collected from diverse sources, including IoT devices, virtual machines, edge nodes, and multi-cloud services. These data streams vary widely in format, frequency, granularity, and reliability (Buyya et al., 2009). As a result, preprocessing becomes a complex task, often requiring normalization, feature alignment, and semantic labeling to create a unified representation suitable for machine learning pipelines. Many detection models fail to generalize across such inconsistent datasets, resulting in up to 18% lower detection accuracy for rare threats and false positive rates exceeding 12% in multiclass classifiers (Fernández et al., 2018).

- **Lack of intelligent preprocessing at the edge**

  Most edge devices are resource-constrained and cannot perform local anomaly detection or even simple filtering tasks. This leads to raw, redundant, and often irrelevant data being transmitted to the cloud, increasing bandwidth usage and introducing latency (Xu et al., 2021; Dong et al., 2022). Without smart preprocessing at the edge, cloud models waste resources processing low-value data, hurting real-time performance.

- **Absence of tamper-proof and verifiable logging mechanisms**

  Centralized logging systems in cloud infrastructures are susceptible to unauthorized modifications, data loss, or deletion. In the case of an intrusion or breach, such logs may not serve as reliable evidence due to the lack of immutability (Wilson & Ghahramani, 2010). This limitation significantly affects post-incident forensics and compliance with audit and regulatory requirements. Verifiable and tamper-resistant log management systems are essential for ensuring traceability, especially in multi-tenant and cross-domain cloud environments.

- **Trust and privacy limitations in multi-tenant environments**

10.48047/jocaaa.2024.33.08.203

In shared cloud infrastructures, multiple clients operate on the same physical or virtual hardware. Due to competitive or regulatory constraints, tenants are often unwilling or unable to share sensitive data, which creates obstacles for collaborative anomaly detection. Although federated learning addresses some privacy concerns by allowing model training without sharing raw data, it introduces new issues such as communication overhead, model drift, and lack of consensus across nodes (Finn et al., 2017).

- **Poor detection of rare or emerging attacks due to class imbalance**

  Most cloud traffic consists of benign behavior, with malicious or anomalous events constituting only a small portion of the data. This class imbalance leads to biased models that fail to generalize well to underrepresented attack types (Chawla et al., 2002; Fernández et al., 2018). Even advanced classifiers tend to favor the majority class, resulting in high false negative rates. Oversampling techniques like SMOTE or ADASYN can help mitigate this, but they introduce the risk of synthetic overfitting and noisy data generation.

- **Limited scalability and real-time responsiveness**

  Anomaly detection models that are accurate in controlled environments often fail under the scale and speed of production cloud environments. Deep learning models such as LSTM and Transformer architectures offer strong performance but require over 2 seconds per batch for inference in real-world workloads, which fails to meet the latency requirements of <500ms in cloud-native detection systems (Vaswani et al., 2017; Ruff et al., 2018). Efficient model deployment using lightweight containers and distributed inference strategies is necessary for real-time, scalable operation.

- **Lack of interpretability and explainability in AI models**

  Deep learning models, while powerful, are often seen as black boxes. Their decisions are difficult to interpret, which limits trust and makes it hard for security analysts to validate or act upon alerts (Mohy-eddine et al., 2023). In sensitive sectors such as finance or healthcare, regulatory compliance requires transparency in decision-making. Techniques such as SHAP or LIME have been proposed to enhance model explainability, but they are still not widely integrated into real-time detection frameworks.

To further justify the proposed three-phase framework, Table 4 synthesizes key limitations identified across the surveyed literature. Despite advancements in ML and DL-based anomaly detection models, many existing approaches suffer from scalability issues, limited interpretability, and poor performance in real-time or imbalanced data scenarios. The following table summarizes these recurring research gaps, supported by representative studies, and highlights the need for a unified, adaptive, and explainable anomaly detection framework suitable for modern cloud environments.

Table 4. Identified Research Gaps

Gap Identified	Description	Example Studies

10.48047/jocaaa.2024.33.08.203

Lack of scalability	Many ML and DL models are trained on static, isolated datasets and struggle to perform in real-time, distributed, multi-tenant cloud systems.	Duque Anton et al., 2019; Ruff et al., 2018
Limited interpretability	Deep learning models are often treated as black boxes, making it hard for security analysts to understand or trust the alerts they generate.	Mohy-eddine et al., 2023; Ruff et al., 2018
Poor real-time responsiveness	Complex models like LSTM and Transformer architectures have high memory requirements and are computationally intensive, limiting their use in latency-sensitive systems.	Xu et al., 2021; Dong et al., 2022; Vaswani et al., 2017
Class imbalance in training data	Most cloud traffic datasets contain a heavy imbalance between benign and malicious instances, reducing model sensitivity to rare attacks.	Chawla et al., 2002; Fernández et al., 2018
Heterogeneous and noisy input data	Data collected from IoT, VMs, and multiple cloud sources vary in format and quality, requiring advanced preprocessing not handled by most current systems.	Buyya et al., 2009; Xu et al., 2021
Lack of secure, verifiable audit trails	Traditional centralized logging is prone to tampering or loss, making forensic analysis and compliance difficult.	Wilson & Ghahramani, 2010
Inadequate privacy in shared models	Cross-tenant privacy is not preserved in most shared learning systems; federated learning is still emerging and not widely deployed.	Finn et al., 2017

Together, these challenges highlight the need for adaptive, interpretable, and scalable anomaly detection strategies in cloud systems. In the following section, a structured three-phase framework is proposed to overcome these gaps by leveraging AI, blockchain technology, and hybrid statistical-machine learning techniques.

4. Future Directions

This research is structured into three integrated phases, each designed to address specific technical and operational gaps identified in current cloud anomaly detection systems. The following future directions illustrate how the proposed framework will systematically overcome these limitations.

4.1 Phase I: AI-Based IoT Device Management and Data Collection

10.48047/jocaaa.2024.33.08.203

To address the persistent challenges identified in current cloud anomaly detection systems, this research proposes a three-phase framework that integrates artificial intelligence, blockchain technology, and hybrid modeling techniques. Each phase is designed to target specific operational gaps in scalability, security, and responsiveness. The proposed roadmap outlines how intelligent edge preprocessing, secure communication, and hybrid AI-based detection can work together to build a more resilient cloud security ecosystem.

Figure 3. Roadmap for Proposed Research Work

The three-phase roadmap illustrated in Fig. 2 presents a layered strategy for addressing key challenges in cloud anomaly detection. Each phase builds upon the previous one, starting with intelligent edge-based data collection, followed by secure and privacy-preserving learning through blockchain and federated models, and culminating in a hybrid anomaly detection engine with explainable AI capabilities. This progressive structure ensures that the system not only enhances detection accuracy but also achieves real-time responsiveness, trust, and scalability—key requirements in modern cloud environments. The design is modular and adaptable, making it well-suited for integration into diverse cloud-native infrastructures.

- **Phase I: AI-based IoT device management and data collection**

  The first phase focuses on improving the quality and consistency of data collected from edge devices, sensors, and virtual endpoints. Cloud environments often receive unstructured and inconsistent data streams, which degrade the accuracy of centralized detection models. To overcome this, lightweight AI models such as MobileNet or TinyML variants will be deployed directly on edge nodes. These models will perform basic anomaly filtering, data normalization, and relevance tagging before transmission, thereby reducing network overhead and improving signal-to-noise ratio (Xu et al., 2021). Additionally, this phase introduces an intelligent preprocessing layer capable of aligning features, standardizing data formats, and classifying data based on origin and reliability. The goal is to ensure that only high-value, anomaly-relevant data is processed at the cloud level, enabling more efficient detection workflows.

10.48047/jocaaa.2024.33.08.203

- **Phase II: Blockchain-based secure communication and federated learning**

  The second phase addresses the issues of trust, verifiability, and data privacy in multi-tenant cloud environments. By integrating blockchain technology, this framework ensures tamper-proof logging of system events, anomaly scores, and detection outcomes. An immutable ledger will be used to track alerts and model updates, enabling secure audit trails and transparent security operations (Wilson & Ghahramani, 2010). Smart contracts will dynamically manage access permissions, ensuring that only authorized entities can view, append to, or verify records. This phase also incorporates federated learning to train AI models collaboratively across decentralized datasets without transferring raw data between domains (Finn et al., 2017). This preserves tenant privacy while still enabling the system to benefit from cross-domain learning. Combined, blockchain and federated learning will enhance trust, transparency, and security collaboration across diverse cloud services.

- **Phase III: AI-based hybrid anomaly detection using ML and DL models**

  The third and most critical phase involves deploying an integrated detection engine that combines statistical, machine learning, and deep learning models for comprehensive anomaly detection. Statistical techniques like DBSCAN will be used for clustering-based detection, while Isolation Forests will address high-dimensional outlier detection. Deep learning architectures, including One-Class LSTM and CNN-BiLSTM, will be used for detecting temporal and spatial anomalies in traffic and logs (Zhang et al., 2022; Xiang et al., 2025). These hybrid models are particularly effective in identifying complex, multi-stage attacks that traditional models miss. To ensure real-time responsiveness and scalable deployment, the models will be optimized for containerized environments such as Docker and Kubernetes. Moreover, explainable AI techniques like SHAP and LIME will be integrated to provide interpretable anomaly scores and visual insights, making the system usable by security analysts and compliant with regulatory standards (Ruff et al., 2018; Mohy-eddine et al., 2023).

This three-phase framework is designed not only to enhance detection accuracy and reduce false alarms but also to support deployment in real-time, container-native environments. By combining intelligent edge processing, decentralized security logging, and hybrid AI architectures, the proposed approach offers a robust and forward-looking solution to anomaly detection in complex cloud infrastructures.

5. Conclusion

This survey presented a comprehensive overview of anomaly detection techniques in cloud computing, with a focus on hybrid models that integrate machine learning, deep learning, and statistical methods. It covered a wide range of supervised and unsupervised algorithms, including deep architectures such as LSTM and CNN-BiLSTM, and sampling strategies like SMOTE, ADASYN, and hybrid undersampling. Evaluation considered detection accuracy (up to 99% in hybrid models), interpretability via SHAP/LIME, and scalability through containerized deployment in Kubernetes environments. Key challenges such as data heterogeneity, trust management, explainability, and class imbalance were explored in detail.

10.48047/jocaaa.2024.33.08.203

To address these gaps, the study proposes a structured three-phase framework. Phase I introduces lightweight AI-based preprocessing on IoT and edge devices to reduce noise and improve bandwidth efficiency. Phase II incorporates blockchain for tamper-proof, verifiable anomaly reporting and federated learning to preserve privacy in multi-tenant environments. Phase III focuses on hybrid detection using statistical and deep learning techniques, supported by explainable AI tools and container-native deployment strategies. Together, these components aim to deliver scalable, accurate, and interpretable security monitoring in complex cloud systems.

Future work will explore edge-native intelligence and lightweight federated models for constrained environments. Additional focus will be placed on dynamic data fusion at the edge-cloud boundary to enable more contextual, situation-aware anomaly detection. The integration of Transformer-based attention models and energy-efficient AI pipelines will further support green computing goals. By pursuing these directions, future systems can achieve greater accuracy, adaptability, trustworthiness, and sustainability in real-time cloud security.

References

1. Acharya, B., Sahu, A., & Sharma, A. (2023). CNN-BiLSTM model for secure multi-cloud anomaly detection. *Journal of Cybersecurity Research, 10*(1), 24–35.
2. An, J., & Cho, S. (2015). Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*, 2(1), 1–18.
3. Arsene, C., Chevalier, Y., & Sirdey, R. (2022). Hybrid data resampling techniques for improved cybersecurity classification. *IEEE Transactions on Dependable and Secure Computing, 19*(2), 672–683. https://doi.org/10.1109/TDSC.2020.2975345
4. Bagui, S., Mink, D., Subramaniam, S., & Wallace, D. (2023). Resampling imbalanced network intrusion datasets to identify rare attacks. *Future Internet, 15*(4), 130. https://doi.org/10.3390/fi15040130
5. Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press.
6. Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems, 25*(6), 599–616. https://doi.org/10.1016/j.future.2008.12.001
7. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953
8. Dong, Z., Chen, Y., Li, W., Zhang, H., & Yang, B. (2022). A transformer-based intrusion detection system for industrial control networks. *IEEE Transactions on Industrial Informatics, 18*(6), 4034–4043. https://doi.org/10.1109/TII.2021.3081361
9. Du, J., Lin, Q., & Ma, M. (2021). Fog-based cloud intrusion detection using enhanced support vector machines. *Journal of Cloud Computing, 10*(1), 1–16. https://doi.org/10.1186/s13677-021-00249-6
10. Duque Anton, S., Gruschka, N., & Lo Iacono, L. (2019). Machine learning for anomaly detection and categorization in multi-cloud environments. In *2019 IEEE 12th International Conference on Cloud Computing (CLOUD)* (pp. 81–89). IEEE. https://doi.org/10.1109/CLOUD.2019.00022

10.48047/jocaaa.2024.33.08.203

11. Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from imbalanced data sets*. Springer.
12. Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning* (Vol. 70, pp. 1126–1135). PMLR.
13. Han, H., Wang, W. Y., & Mao, B. H. (2005). Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In *Advances in Intelligent Computing* (pp. 878–887). Springer. https://doi.org/10.1007/11538059_91
14. He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *IEEE International Joint Conference on Neural Networks* (pp. 1322–1328). https://doi.org/10.1109/IJCNN.2008.4633969
15. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
16. Japkowicz, N. (2000). The class imbalance problem: Significance and strategies. In *Proceedings of the 2000 International Conference on Artificial Intelligence* (Vol. 1, pp. 111–117).
17. Kim, G., Lee, S., & Kim, S. (2016). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. *Expert Systems with Applications, 41*(4), 1690–1700. https://doi.org/10.1016/j.eswa.2013.08.066
18. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539
19. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining* (pp. 413–422). IEEE. https://doi.org/10.1109/ICDM.2008.17
20. Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery, 28*(1), 92–122. https://doi.org/10.1007/s10618-012-0295-5
21. Mohy-eddine, H., Mahgoub, M., & AlShamrani, A. (2023). Artificial neural networks for DDoS attack detection in cloud computing. *Journal of Network and Computer Applications, 217*, 103556. https://doi.org/10.1016/j.jnca.2022.103556
22. Naseer, S., & Saleem, M. (2018). Deep learning-based cloud intrusion detection using convolutional neural networks. *Computers & Security, 79*, 219–230. https://doi.org/10.1016/j.cose.2018.08.001
23. Ruff, L., Vandermeulen, R. A., Görnitz, N., Deecke, L., Siddiqui, S. A., Binder, A., ... & Kloft, M. (2018). Deep one-class classification. In *International Conference on Machine Learning* (pp. 4393–4402). PMLR.
24. Tomek, I. (1976). Two modifications of CNN. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-6*(11), 769–772. https://doi.org/10.1109/TSMC.1976.5408784
25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, 30, 5998–6008.
26. Wang, J., Guo, Y., Chen, Y., & Xu, K. (2020). A deep learning method for anomaly detection in network traffic. *IEEE Access, 8*, 191138–191145. https://doi.org/10.1109/ACCESS.2020.3031183
27. Wilson, A. G., & Ghahramani, Z. (2010). Copula processes. In *Advances in Neural Information Processing Systems* (pp. 2460–2468).

10.48047/jocaaa.2024.33.08.203

28. Xiang, H., Li, J., & Yu, W. (2025). A BiLSTM-CNN model for intrusion detection in industrial IoT environments. *IEEE Access, 13*, 18134–18147. https://doi.org/10.1109/ACCESS.2025.3284912
29. Xu, H., Shen, J., & Wang, K. (2021). Improving LSTM-based anomaly detection with attention mechanisms in cloud systems. *Journal of Information Security, 12*(3), 142–154.
30. Zhang, X., Liu, Y., & Wang, J. (2022). Deep CNN-BiLSTM hybrid model for network intrusion detection. *IEEE Access, 10*, 9825–9836. https://doi.org/10.1109/ACCESS.2022.3143495