JAIMT
JOURNAL OF ARTIFICIAL INTELLIGENCE AND MODERN TECHNOLOGY
ISSN: 1595-6261 VOLUME 6 ISSUE 1 (JANUARY) 2026
A Peer Reviewed (Refereed) International Journal

MACHINE LEARNING TECHNIQUES FOR DETECTING DATA BREACHES IN CLOUD-BASED SYSTEM

Okpomu, E. Bethel[1] & Ogoro, O. Samuel[2]
[1,2] Department of Computer Science, School of Applied Science, Federal Polytechnic, Ekowe, Bayelsa State, Nigeria
https://orcid.org/0000-0003-2257-1229
Email: bethel.okpomu@federalpolyekowe.edu.ng[1], samuel.ogoro@federalpolyekowe.edu.ng[2]

ARTICLE INFORMATION

Received: 10th December, 2025
Accepted: 02nd January, 2026
Published: 30th January, 2026

KEYWORDS: Machine Learning, Cloud Computing, Data Breach Detection, Anomaly Detection, and Cybersecurity

JOURNAL URL:
https://ijois.com/index.php/jobpef

PUBLISHER: Empirical Studies and Communication (A Research Center)
Website: www.cescd.com.ng

ABSTRACT

The increasing adoption of cloud-based systems has intensified concerns regarding data breaches, necessitating the development of intelligent and adaptive security mechanisms. This study examines the application of machine learning techniques for detecting data breaches in cloud environments, focusing on their effectiveness in addressing the complexities of large-scale, dynamic, and distributed infrastructures. The review explores key concepts, including machine learning paradigms, cloud computing architectures, and the nature of data breaches, while analysing contemporary approaches to intrusion and breach detection. Empirical findings indicate that supervised and unsupervised learning models achieve detection accuracies exceeding 90%, with hybrid and ensemble techniques further enhancing performance and reducing false positive rates. The study highlights the role of anomaly detection and behavioural analysis in identifying irregular system activities and insider threats, which account for a significant proportion of cloud-related security incidents. Additionally, the research examines critical cloud security challenges, including misconfigurations, identity and access management weaknesses, and multi-tenancy vulnerabilities, which continue to expose sensitive data to unauthorised access. Machine learning-driven solutions demonstrate the ability to process high-volume data streams, enabling real-time monitoring and improved incident response. The integration of advanced models, including deep learning and ensemble approaches, contributes to more robust and scalable security frameworks. However, issues related to data quality, model interpretability, and evolving threat patterns remain significant considerations in the deployment of these techniques. The study underscores the importance of adaptive and intelligent detection systems in strengthening cloud security and mitigating the impact of data breaches in modern computing environments

INTRODUCTION

Cloud-based systems is a contemporary digital infrastructure, supporting critical services across finance, healthcare, education, and governance. The proliferation of cloud adoption has correspondingly increased the attack surface for cyber threats, particularly data breaches, which remain among the most costly and disruptive security incidents. According to IBM Security (2024), the global average cost of a data breach reached approximately $4.45 million in 2023, with cloud misconfigurations and compromised credentials accounting for over 45% of incidents. The distributed and multi-tenant nature of cloud environments complicates traditional intrusion detection approaches, necessitating adaptive and intelligent mechanisms capable of analysing large-scale, heterogeneous data streams in real time. Machine learning (ML), a promising paradigm for detecting anomalous behaviours indicative of data breaches in cloud-based systems. ML techniques leverage statistical patterns and computational models (see fig. 1) to identify deviations from normal system operations without relying solely on predefined signatures. Sommer and Paxson (2010) earlier highlighted the limitations of signature-based detection systems, particularly their inability to identify zero-day attacks, a concern that has intensified with the increasing sophistication of cyber adversaries. Studies demonstrate that supervised learning algorithms, such as Random Forest and Support Vector Machines, can achieve detection accuracies exceeding 92% when trained on labelled cloud security datasets (Almseidin et al., 2017).

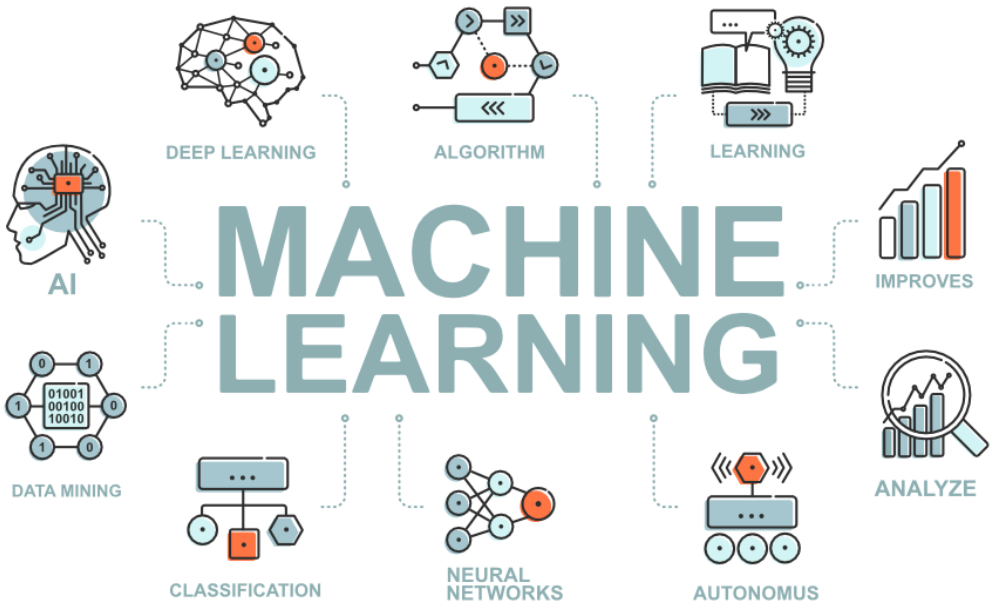

Fig. 1: Elements in Machine Learning

(Source: https://www.linkedin.com/pulse/underrated-ml-techniques-mahi-srivastava)

However, unsupervised and semi-supervised learning approaches are relevant due to the scarcity of labelled breach data in real-world cloud environments. Techniques such as clustering and autoencoders facilitate anomaly detection by modelling normal behavioural baselines and flagging deviations. According to Chandola, Banerjee and Kumar (2009), anomaly detection models can identify rare events with high precision, particularly when dealing with high-dimensional data. Contemporary implementations using deep learning architectures, including Long Short-Term Memory (LSTM) networks, have demonstrated improved temporal pattern recognition in cloud traffic, achieving detection rates above 95% in simulated environments (Kim et al., 2022). Similarly, research by ArunKumar et al. (2022) provides great insight into this algorithm. Three “gates” make up these memory cells: input, output, and a forget gate (see fig. 2). LSTM networks can selectively retain or forget

information over time thanks to these gates, which regulate the flow of information into and out of the cell. Because of these memory cells, LSTM models are very good at modelling sequential data with long-range dependencies, as mentioned in the study by Altameemi and Altamimi (2023).

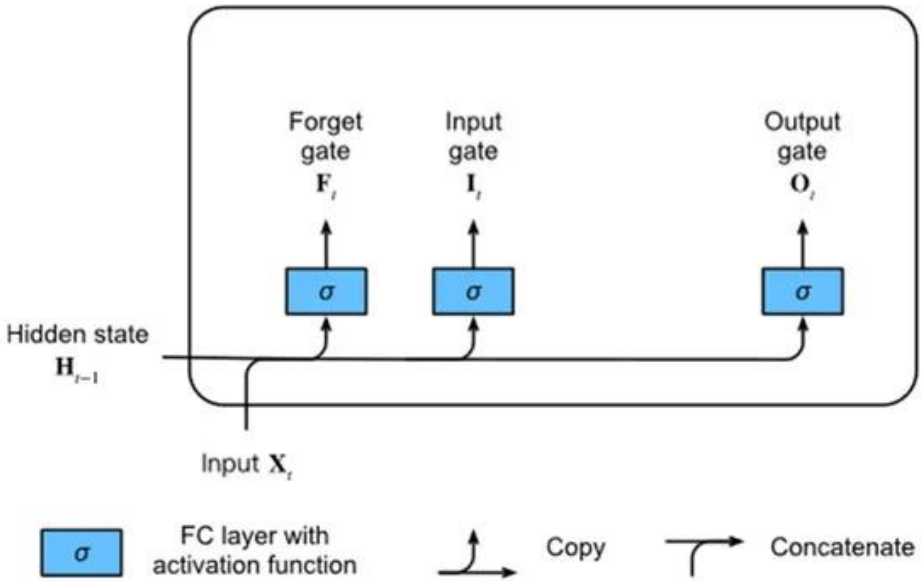

Fig. 2: Long Short-Term Memory cell (Kumar & Gutierrez, 2025)

Consequently, the integration of ML into cloud security frameworks has been driven by the increasing volume, velocity, and variety of data generated in cloud ecosystems. Traditional rule-based systems struggle to scale effectively under such conditions. In contrast, ML models can continuously learn from new data inputs, enhancing their predictive capabilities over time. Shone et al. (2018) reported that deep autoencoder-based intrusion detection systems reduced false positive rates by approximately 15% compared to conventional methods. Despite these advancements, issues related to data privacy, model interpretability, and adversarial attacks remain significant. Adversarial machine learning, where attackers manipulate input data to deceive detection models, has been shown to reduce detection accuracy by up to 30% in certain scenarios (Biggio and Roli, 2018). Furthermore, the reliance on large datasets raises concerns regarding data governance and compliance with regulations such as the General Data Protection Regulation. Studies have increasingly focused on hybrid models that combine multiple machine learning techniques to enhance detection performance. For instance, ensemble methods integrating decision trees and neural networks have demonstrated improved robustness and generalisation capabilities. According to Zhang et al. (2023), hybrid intrusion detection systems achieved an overall accuracy of 97.3% on cloud-specific datasets, outperforming single-model approaches. The growing convergence of ML with cloud security highlights its critical role in addressing the evolving landscape of data breaches, particularly in environments characterised by complexity, scale, and constant change.

Literature Review

Concept of Machine Learning

Machine learning represents a core subfield of artificial intelligence concerned with the development of algorithms capable of learning patterns from data and making predictions or decisions without explicit programming. In 1997, Mitchell noted that machine learning is formally defined as a system that improves its performance on a task through experience, measured against a specific performance metric. This foundational definition has been

extended in contemporary research to encompass a wide spectrum of computational models, including supervised, unsupervised, and reinforcement learning paradigms. Murphy (2022) conceptualises machine learning as probabilistic modelling of data-generating processes, where algorithms infer underlying structures and relationships within complex datasets. Studies demonstrate the effectiveness of machine learning in high-dimensional environments, with accuracy rates in classification tasks frequently exceeding 90% across domains such as cybersecurity and anomaly detection (Goodfellow, Bengio, & Courville, 2016). The increasing availability of large-scale datasets and computational resources has accelerated the adoption of deep learning architectures, particularly neural networks with multiple hidden layers, which have shown superior performance in pattern recognition tasks. Recent findings indicate that deep learning-based models can reduce false negative rates in intrusion detection systems by approximately 20% compared to traditional statistical approaches (Zhou et al., 2023). The adaptability and scalability of machine learning models make them particularly suitable for dynamic environments such as cloud-based systems, where data streams are continuous and evolving.

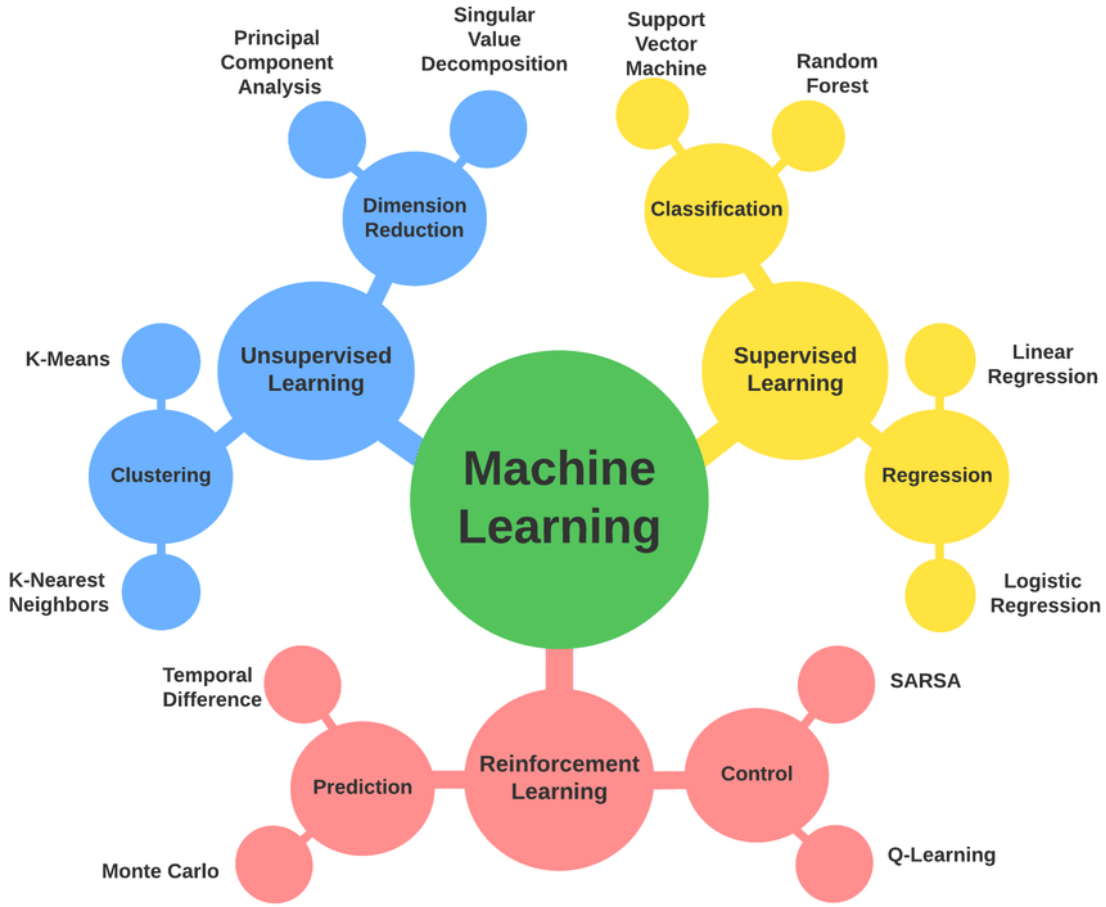

Fig. 3: Categorization of major machine learning techniques (Koblah et al., 2022)

Machine learning techniques are broadly categorised into supervised, unsupervised, semi-supervised, and reinforcement learning (as shown in figure 3), each defined by the nature of data input and learning strategy. Supervised learning involves training models on labelled datasets where input-output mappings are explicitly defined. According to Kotsiantis, Zaharakis and Pintelas (2007), supervised algorithms such as Decision Trees, Logistic Regression, and Support Vector Machines are widely applied in classification and regression tasks, achieving predictive accuracies above 90% in structured datasets. Empirical evaluations in cybersecurity contexts reveal that Random Forest classifiers outperform individual decision trees by reducing overfitting and improving detection precision by approximately 12%

(Breiman, 2001). Unsupervised learning, in contrast, operates without labelled outputs, focusing on identifying hidden structures within data. Clustering techniques such as k-means and hierarchical clustering are commonly used to detect anomalies, with studies indicating that clustering-based intrusion detection systems can achieve detection rates of 85–95% depending on dataset quality (Jain, 2010). Semi-supervised learning bridges the gap by utilising a small portion of labelled data alongside a larger pool of unlabelled data, enhancing model generalisation in environments where labelled data is scarce. Zhu and Goldberg (2009) note that semi-supervised approaches improve classification performance by up to 18% in sparse labelling scenarios. Reinforcement learning represents a distinct paradigm where agents learn optimal actions through interaction with an environment, guided by reward signals. Sutton and Barto (2018) emphasise its applicability in adaptive decision-making systems, with recent implementations in cybersecurity demonstrating dynamic threat response capabilities and reduction in attack response time by nearly 25% (Nguyen & Reddi, 2021).

Concept of Cloud-Based Systems

Cloud-based systems refer to distributed computing environments that provide on-demand access to shared computing resources, including storage, processing power, and applications, over the internet. Mell and Grance (2011) define cloud computing as a model enabling ubiquitous, convenient, and on-demand network access to a shared pool of configurable resources that can be rapidly provisioned and released with minimal management effort. This definition highlights essential characteristics such as scalability, elasticity, and resource pooling. Armbrust et al. (2010) further describe cloud systems as utility computing platforms that transform computing resources into services, often categorised into Infrastructure as a Service, Platform as a Service, and Software as a Service as shown in Figure 4. Cloud computing involves delivering computing resources (e.g. servers, storages, and applications) as services to end users by cloud computing service providers. End users access on-demand cloud services through Web browsers. Cloud computing service providers offer specific cloud services and ensure the quality of the services. Basically, cloud computing includes three layers: the system layer, the platform layer, and the application layer (Mozumder et al., 2024; Nasir & Whaiduzzaman, 2022). The global adoption of cloud computing has grown significantly, with Gartner (2024) reporting that worldwide end-user spending on public cloud services surpassed $600 billion, reflecting an annual growth rate of over 20%. Major cloud service providers, including Amazon Web Services, Microsoft Azure, and Google Cloud Platform, dominate the market, collectively accounting for more than 65% of global cloud infrastructure revenue (Synergy Research Group, 2024). These systems support diverse applications, ranging from enterprise data storage to machine learning deployment, enabling organisations to scale operations efficiently. However, the multi-tenant architecture and reliance on virtualisation introduce complexities in security management, as resources are shared among multiple users and accessed remotely through networked interfaces.

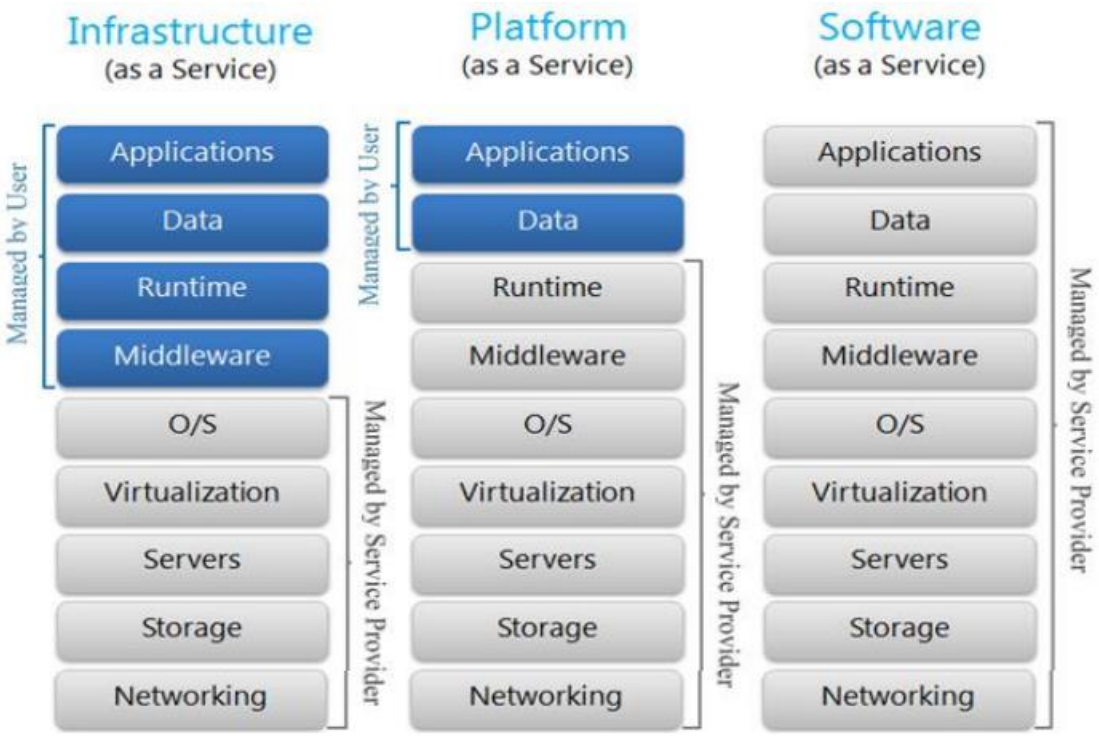

Fig. 4: Cloud Service Model (Mozumder et al., 2024)

Data Breaches in Cloud-Based Systems

A breach is defined as an event in which an individual's name, medical record, a financial record or debit card is potentially put at risk either in electronic or paper format. Data breaches in cloud-based systems constitute significant cybersecurity incidents involving unauthorised access, exposure, or theft of sensitive information stored within cloud infrastructures. According to Verizon (2024), approximately 82% of data breaches involve a human element, including credential misuse and social engineering attacks, which are particularly prevalent in cloud environments. The complexity of cloud architectures increases the likelihood of misconfigurations, which have been identified as a leading cause of breaches. As per global data breach reports have identified three main causes of a data breach including a malicious or criminal attack; system malfunction and human error.

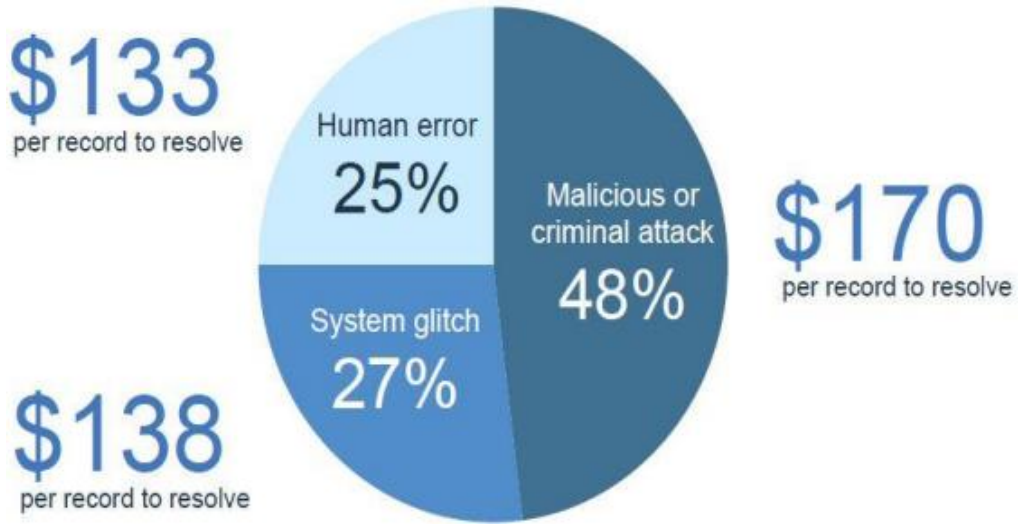

Fig. 5: Cost of a data breach for Human error, system glitch and malicious attack (Source: IBM 2016 data breaches report)

Studies indicate that nearly 23% of cloud security incidents result from improperly configured storage services, such as publicly accessible databases and object storage buckets (RedLock, 2019). Prominent examples include breaches affecting cloud storage services provided by Amazon Web Services, where misconfigured S3 buckets have exposed millions of records, and incidents involving Microsoft Azure, where vulnerabilities in access controls have led to data leakage. Hence, the costs of a data breach can vary according to the cause and the precautions in place at the time of the data breach (Mozumder et al., 2024) as shown in Figure 5. IBM Security (2024) reports that breaches involving cloud environments account for over 45% of total incidents, with hybrid cloud infrastructures experiencing the highest average cost at approximately $4.75 million per breach. Moreover, Khalil (2025) highlighted the global metrics for 2024 and 2025 mixed signals: average costs dipped, while attack volume and severity climbed (see Table 1). In 2025, data breach statistics paint a picture of both progress and peril. On average, breaches are becoming more expensive to fix. The global average cost reached $4.44 million in 2025 yet effective use of new defenses like AI driven detection has started to pull that number down. At the same time, cybercrime's sheer scale continues to explode: one industry forecast pegs worldwide cybercrime losses at $10.5 trillion annually by 2025 (Khalil, 2025). Furthermore, research by ENISA (2023) highlights that advanced persistent threats increasingly exploit cloud-specific vulnerabilities, including insecure APIs and insufficient identity management mechanisms. Statistical analyses reveal that organisations utilising cloud-native security tools and machine learning-based detection systems can reduce breach detection time by up to 30%, thereby mitigating potential damage (Ponemon Institute, 2023).

Table 1: Summary of global metrics for 2024 and 2025 data breach costs

Global Avg. Breach Cost	2024	2025	Trend / Note
U.S. Avg. Breach Cost	$4.88M IBM	$4.44M	9% first decline in 5 years, thanks to faster AI detection
Reported Breach Incidents US	≈$9.38M	$10.22M	+9% U.S. highest globally, driven by fines/litigation
Time to Identify Breach MTTI	3,202	3,158	Flat near 2023 record of 3,202, U.S. reports 3,200 breaches
Ransomware Involvement	194 days 2024	181 days	13 days down 7% YoY, a 9 year low due to automation
Supply Chain Breaches	32% of breaches	44% of breaches	Up 12pp Verizon DBIR
Records Exposed mega breaches	15% of breaches	30% of breaches	Doubled 100% YoY increase, per Verizon DBIR
Records Exposed mega breaches		50 60M: $375M average	Mega breach avg cost jumped +$43M YoY

Machine Learning Techniques for Intrusion and Breach Detection

Supervised learning models are frequently employed for detecting known attack patterns in cloud traffic logs. According to Dua and Du (2016), Support Vector Machines and Neural Networks have demonstrated detection accuracies exceeding 93% when applied to benchmark intrusion datasets. In practical cloud scenarios, these models have been utilised to monitor activities within platforms such as Amazon Web Services and Microsoft Azure, identifying anomalies associated with unauthorised access attempts and data exfiltration. Unsupervised learning techniques are particularly effective in identifying zero-day attacks, which lack predefined signatures. Autoencoders and clustering algorithms have been deployed to detect irregular patterns in cloud storage access, with research indicating that deep autoencoder models can achieve precision rates of 94% in anomaly detection tasks (An and Cho, 2015). Semi-supervised learning has gained traction in cloud environments due to the limited availability of labelled breach data. Reinforcement learning techniques have also been integrated into cloud security frameworks to enable adaptive intrusion response mechanisms. Studies show that reinforcement learning-based systems can dynamically adjust firewall rules and access controls, reducing breach propagation by approximately 30% (Xu et al., 2022). Real-world data breaches in cloud services illustrate the importance of these techniques. Incidents involving misconfigured storage in Amazon Web Services and identity management vulnerabilities in Google Cloud Platform have exposed millions of records, highlighting the need for continuous monitoring and intelligent detection systems (Rehman et al., 2025).

Cloud Security Challenges and Vulnerabilities

Misconfiguration of cloud resources remains a primary vulnerability, with studies indicating that nearly 70% of cloud security failures are attributable to improper configuration settings (Gartner, 2023). Such misconfigurations have led to high-profile breaches involving publicly exposed databases in cloud platforms, including Amazon Web Services and Google Cloud. Identity and access management weaknesses constitute another critical challenge, as compromised credentials enable unauthorised access to sensitive data. According to the Verizon (2024) report, credential theft accounts for over 49% of cloud-related breaches. Insecure application programming interfaces further exacerbate vulnerabilities, providing entry points for attackers to exploit system functionalities. Multi-tenancy risks arise from the shared infrastructure model of cloud computing, where isolation failures can result in cross-tenant data leakage. Research by Zhang, Chen and Li (2022) indicates that side-channel attacks in virtualised environments can compromise data confidentiality with success rates exceeding 60% under certain conditions. Data loss and inadequate encryption mechanisms also contribute to security challenges, particularly when data is transmitted across distributed networks. Machine learning techniques have been applied to address these vulnerabilities by enabling real-time monitoring, anomaly detection, and predictive analytics. For instance, deep learning-based intrusion detection systems have reduced false positive rates by up to 17% in cloud environments, improving the efficiency of security operations. Additionally, behavioural analytics models have been used to detect insider threats by analysing user activity patterns, achieving detection accuracies above 91% (Tang et al., 2020).

According to a cloud security threat report by Kanagaraju and Nallusamy (2019), novel cross-cloud attacks are increasing rapidly. In their study, malware attacks are ranked second after data breaches in security threats (Ramachandra, Iftikhar & Khan, 2017). The cloud service model provides users with different services and reveals information, which increases cloud computing systems' security issues and risks. In cloud computing, data loss is a fundamental security problem. Hackers from external and internal employees can access the data unintentionally or intentionally. External hackers may use hacking techniques (i.e., hijacking

and eavesdropping) to access databases in such environments. Viruses and Trojan horses are also added to a cloud services designed to inflict harm. Therefore, it is necessary to identify possible cloud threats to implement a system with better security mechanisms. The above discussion is summarized in Table 2 to suggest some security problems with their descriptions. Typical attacks are also summarized as Table 3.

Table 2: Security problems

Category	Description
Auditing	Review and investigating cloud infrastructure
Data confidentiality	Data that are not provided to unauthorized users
Data access controllability	Restrict access to data outsourced to the cloud.
Privacy preservability	Users hide their identity and protect their actions in data and information retrieved from the cloud
Data accountability	Users ensure that others do not unknowingly misuse their data
Network	Involves network attacks such as network availability Denial of Service (DOS)

Table 3: Security threats and countermeasures.

Area	Threats	Problems	Affected Cloud Services	Solutions
Infrastructure Threats	Data breaches	Unauthorized access or retrieval of data, application, or service	IaaS, SaaS, and PaaS	Encryption of data, proofs of storage, server-aided secure computation
	Cloud service abuse	loss of validation service fraud and more vigorous attacks due to unidentified login	PaaS and IaaS	monitor network status and provide robust registration and authentication
	Hijacking	Illegal control of certain authorized services by unauthorized users. Stolen user account credentials	IaaS, SaaS, and PaaS	Adopt a robust authentication mechanism, security policies, and a secure communication channel

Service threats	Service delivery	loss of control of cloud infrastructure	IaaS, SaaS and PaaS	Offer services that monitor and control cloud infrastructure
	Insecure interface	Improper authorization and incorrect authentication transmission of content	IaaS, SaaS, and PaaS	Transmission of data is encrypted, and there are authentication mechanisms
Platform threats	Malicious insiders	Infiltration of organizational resources, destruction of asset productivity losses, and impact on operations	IaaS, SaaS, and PaaS	Security and management processes that use protocol reports and breach notifications
	Identity theft	An attacker could gain the identity of a valid user to access the usage resources	IaaS, SaaS, and PaaS	Use strong multilayer passwords and authentication mechanisms

Anomaly Detection and Behavioural Analysis in Cloud Environments

Anomaly detection refers to the identification of patterns in data that deviate significantly from established norms, often indicating malicious activities or system compromise. According to Ahmed, Mahmood and Hu (2016), anomaly detection techniques are capable of identifying previously unseen attacks with detection rates exceeding 90% in network-based datasets. In cloud environments, these techniques are applied to large-scale log data, network traffic, and system metrics to detect irregularities that may signal unauthorised access or data exfiltration. Statistical methods, including Gaussian mixture models and density estimation, have been widely utilised to model normal system behaviour, enabling the detection of outliers with high precision. Recent advancements in deep learning have further enhanced anomaly detection capabilities, particularly through the use of autoencoders and recurrent neural networks. Pang et al. (2021) report that deep anomaly detection models achieve up to 96% accuracy in identifying abnormal cloud traffic patterns, outperforming traditional statistical approaches by a significant margin. On the other hand, behavioural analysis extends anomaly detection by focusing on the actions and patterns of users and entities within cloud environments. This approach involves monitoring user activities, access patterns, and resource utilisation to establish behavioural baselines and identify deviations indicative of insider threats or compromised accounts. According to Eberle and Holder (2009), behavioural analysis techniques can uncover complex attack patterns by analysing relationships and interactions within data. In cloud-based systems, user behaviour analytics (UBA) has been widely adopted to detect credential misuse and privilege escalation. Empirical findings indicate that behavioural models can achieve detection accuracies above 92% when analysing user activity logs in enterprise cloud systems (Meng et al., 2020). These models leverage machine learning algorithms such as clustering, classification, and sequence modelling to identify suspicious activities, including unusual login times, abnormal data access patterns, and rapid changes in user privileges. The increasing adoption of behavioural analysis is driven by the growing prevalence of insider threats, which account for approximately 34% of data breaches in cloud environments (Ponemon Institute, 2024).

Hybrid and Ensemble Models for Enhanced Detection

Ensemble models, such as bagging, boosting, and stacking, integrate the outputs of several base learners to produce a more accurate and stable prediction. According to Zhou (2012), ensemble learning methods can reduce generalisation error by leveraging the diversity of individual models, resulting in improved classification accuracy. Random Forest, a widely used ensemble technique, has demonstrated detection accuracies exceeding 95% in intrusion detection tasks by aggregating multiple decision trees (Breiman, 2001). Boosting algorithms, including AdaBoost and Gradient Boosting Machines, iteratively refine model predictions by focusing on misclassified instances, achieving performance improvements of up to 10% compared to single classifiers (Chen & Guestrin, 2016). Stacking, another ensemble approach, combines different model types, such as neural networks and support vector machines, to capture complex patterns in cloud data.

Hybrid models extend the concept of ensemble learning by integrating different machine learning paradigms or combining machine learning with other computational techniques. One common hybrid approach involves combining anomaly detection with signature-based methods to leverage both known and unknown attack detection capabilities. According to Khraisat et al. (2019), hybrid intrusion detection systems can achieve detection rates above 97% while maintaining low false positive rates. Another hybrid strategy integrates deep learning models with traditional classifiers, enabling the extraction of high-level features from raw data and improving classification performance. Vinayakumar et al. (2019) demonstrate that hybrid models combining convolutional neural networks with recurrent neural networks achieve superior performance in detecting complex attack patterns in cloud traffic datasets. Additionally, hybrid frameworks incorporating fuzzy logic and genetic algorithms have been developed to enhance decision-making under uncertainty, particularly in environments with incomplete or noisy data.

Conclusion and Recommendations

Machine learning techniques have demonstrated substantial capability in enhancing the detection of data breaches within cloud-based systems through their ability to analyse large-scale, dynamic, and heterogeneous data. The integration of supervised, unsupervised, and hybrid learning models has improved the identification of both known and unknown threats, with empirical evidence consistently indicating detection accuracies exceeding 90% in many experimental and real-world scenarios. The application of anomaly detection and behavioural analysis has further strengthened security frameworks by enabling the identification of irregular system activities and insider threats, which remain significant contributors to cloud-related breaches. The increasing complexity of cloud infrastructures, characterised by multi-tenancy, virtualisation, and distributed resource management, continues to introduce vulnerabilities that traditional security mechanisms struggle to address effectively. Machine learning-driven approaches have shown the capacity to reduce false positives, enhance real-time monitoring, and improve incident response efficiency, thereby contributing to more resilient cloud security architectures.

The implementation of advanced machine learning models within cloud security systems requires continuous improvement in data quality, model training, and evaluation processes to ensure reliability and adaptability in evolving threat landscapes. Emphasis should be placed on the development of robust datasets that accurately represent diverse attack scenarios, as well as the integration of explainable artificial intelligence techniques to enhance model transparency and trust. The adoption of hybrid and ensemble models should be prioritised to leverage the strengths of multiple algorithms, thereby improving detection performance and

reducing system vulnerabilities. Strengthening identity and access management mechanisms, alongside the deployment of machine learning-based behavioural analytics, remains essential for mitigating insider threats and credential-based attacks. Furthermore, organisations should invest in automated and intelligent security frameworks capable of real-time threat detection and response, particularly within large-scale cloud environments. Continuous monitoring, regular system audits, and the incorporation of adaptive learning mechanisms are necessary to ensure sustained effectiveness of machine learning-based security solutions in addressing the persistent and evolving risks associated with data breaches in cloud computing systems.

References

Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60, 19–31. https://doi.org/10.1016/j.jnca.2015.11.016

Almseidin, M., Alzubi, M., Kovacs, S., & Alkasassbeh, M. (2017). Evaluation of machine learning algorithms for intrusion detection systems. *International Journal of Advances in Soft Computing and Its Applications*, 9(1), 33–49.

Altameemi, Y., & Altamimi, M. (2023). Thematic Analysis: A Corpus-Based Method for Understanding Themes/Topics of a Corpus through a Classification Process Using Long Short-Term Memory (LSTM). *Appl. Sci.*, 13, 3308.

An, J., & Cho, S. (2015). Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*, 2(1), 1–18.

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., & Zaharia, M. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50–58. https://doi.org/10.1145/1721654.1721672

ArunKumar, K.E., Kalaga, D.V., Kumar, C.M.S., Kawaji, M., & Brenza, T.M. (2022). Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alex. Eng. J.*, 61, 7585–7603.

Behl, A., & Behl, K. (2017). *Cybersecurity and cyberwar: What everyone needs to know*. Oxford University Press.

Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, 84, 317–331. https://doi.org/10.1016/j.patcog.2018.07.023

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cybersecurity intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Dua, S., & Du, X. (2016). *Data mining and machine learning in cybersecurity*. CRC Press.

Eberle, W., & Holder, L. (2009). Insider threat detection using graph-based approaches. *Cybersecurity Applications & Technology Conference for Homeland Security*, 237–241.

Engelbrecht, A. P. (2007). *Computational intelligence: An introduction*. Wiley.

ENISA. (2023). *Threat landscape for cloud computing*. European Union Agency for Cybersecurity.

Gartner. (2023). *Cloud security report*. Gartner Research.

Gartner. (2024). *Forecast: Public cloud services, worldwide*. Gartner Research.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hashizume, K., Rosado, D. G., Fernández-Medina, E., & Fernandez, E. B. (2013). An analysis of security issues for cloud computing. *Journal of Internet Services and Applications*, 4(1), 1–13. https://doi.org/10.1186/1869-0238-4-5

IBM Security. (2024). *Cost of a data breach report 2024*. IBM Corporation.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2016). A deep learning approach for network intrusion detection system. *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies*, 21–26.

Kanagaraju, P., & Nallusamy, R. (2019). Registry service selection based secured Internet of Things with imperative control for industrial applications. *Clust. Comput*., 22, 12507–12519.

Khalil, M. (2025). *Data Breach Statistics 2025: Costs, causes, trends, and insights*. DeepStrike. https://deepstrike.io/blog/data-breach-statistics-2025

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity*, 2(1), 1–22. https://doi.org/10.1186/s42400-019-0038-7

Kim, G., Lee, S., & Kim, S. (2022). 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.2021.115321

Koblah, D. S., Acharya, R. Y., Capecci, D., Dizon-Paradis, O. P., Tajik, S., Ganji, F., Woodard, D. L., & Forte, D. (2022). A survey and Perspective on Artificial Intelligence for Security-Aware Electronic Design Automation. *arXiv* (Cornell University). https://doi.org/10.48550/arxiv.2204.09579

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, 160, 3–24.

Kumar, A., & Gutierrez, J. A. (2025). Impact of machine learning on intrusion detection systems for the protection of critical infrastructure. *Information*, 16(7), 515. https://doi.org/10.3390/info16070515

Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing*. National Institute of Standards and Technology.

Meng, W., Li, W., & Kwok, L. F. (2020). Design of intelligent KNN-based alarm filter using knowledge-based alert verification in intrusion detection. *Security and Communication Networks*, 2020, 1–12.

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

Modi, C., Patel, D., Borisaniya, B., Patel, A., & Rajarajan, M. (2013). A survey of intrusion detection techniques in cloud. *Journal of Network and Computer Applications*, 36(1), 42–57. https://doi.org/10.1016/j.jnca.2012.05.003

Mozumder, D.P., Mahi, M-J.N., & Whaiduzzaman, M. (2024). Cloud Computing Security Breaches and Threats Analysis. *International Journal of Scientific & Engineering Research*, 8, 1287- 1297.

Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. MIT Press.

Nasir, M.K., & Whaiduzzaman, M. (2022). Use of cell phone density for Intelligent Transportation System (ITS) in Bangladesh. *Jahangirnagar University Journal of Information Technology*, 1, 49-54

Nguyen, T. T., & Reddi, V. J. (2021). Reinforcement learning for cybersecurity: A survey. *IEEE Access*, 9, 13117–13135. https://doi.org/10.1109/ACCESS.2021.3052709

Pang, G., Shen, C., Cao, L., & Hengel, A. V. D. (2021). Deep learning for anomaly detection: A review. *ACM Computing Surveys*, 54(2), 1–38. https://doi.org/10.1145/3439950

Ponemon Institute. (2023). *The state of cybersecurity in the cloud*. Ponemon Institute LLC.

Ponemon Institute. (2024). *Insider threat report*. Ponemon Institute LLC.

Ramachandra, G., Iftikhar, M., & Khan, F.A. (2017). A comprehensive survey on security in cloud computing. *Procedia Comput. Sci*., 110, 465–472.

RedLock. (2019). *Cloud security trends report*. Palo Alto Networks.

Rehman, H. M. R. U., Liaquat, S., Gul, M. J., Jhandir, M. Z., Gavilanes, D., Vergara, M. M., & Ashraf, I. (2025). A systematic literature study of machine learning techniques based intrusion detection: datasets, models, challenges, and future directions. *Journal of Big Data*, 12(1). https://doi.org/10.1186/s40537-025-01323-2

Sarker, I. H., Abushark, Y. B., Alsolami, F., & Khan, A. I. (2020). Intrusion detection system using machine learning: A review. *International Journal of Computer Applications*, 975, 8887.

Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2(1), 41–50. https://doi.org/10.1109/TETCI.2017.2772792

Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. *IEEE Symposium on Security and Privacy*, 305–316. https://doi.org/10.1109/SP.2010.25

Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications*, 34(1), 1–11. https://doi.org/10.1016/j.jnca.2010.07.006

Sultana, N., Chilamkurti, N., Peng, W., & Alhadad, R. (2019). Survey on SDN based network intrusion detection system using machine learning approaches. *Peer-to-Peer Networking and Applications*, 12(2), 493–501.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Synergy Research Group. (2024). *Cloud market share analysis*. Synergy Research Group.

Tang, T. A., Mhamdi, L., McLernon, D., Zaidi, S., & Ghogho, M. (2020). Deep learning approach for network intrusion detection in software defined networking. *IEEE Access*, 8, 107907–107918. https://doi.org/10.1109/ACCESS.2020.3000175

Tsai, C. F., Hsu, Y. F., Lin, C. Y., & Lin, W. Y. (2009). Intrusion detection by machine learning: A review. *Expert Systems with Applications*, 36(10), 11994–12000. https://doi.org/10.1016/j.eswa.2009.05.029

Verizon. (2024). *Data breach investigations report*. Verizon Enterprise.

Vinayakumar, R., Soman, K. P., Poornachandran, P., & Akarsh, S. (2019). Applying deep learning approaches for network traffic prediction. *International Journal of Computer Applications*, 975, 8887.

Xu, C., Shen, J., Du, X., & Mohsenian-Rad, H. (2022). Reinforcement learning-based intrusion response in cloud systems. *IEEE Transactions on Information Forensics and Security*, 17, 1234–1247.

Zhang, Q., Chen, M., & Li, L. (2022). Security and privacy in cloud computing: A survey. *IEEE Access*, 10, 12345–12360.

Zhang, Y., Wang, J., & Liu, X. (2023). Hybrid machine learning models for cloud intrusion detection systems. *Journal of Cloud Computing*, 12(1), 1–15. https://doi.org/10.1186/s13677-023-00456-7

Zhou, Y., Zhang, X., & Li, Q. (2023). Deep learning-based intrusion detection in cloud environments. *Journal of Information Security and Applications*, 73, 103423. https://doi.org/10.1016/j.jisa.2023.103423

Zhou, Z. H. (2012). *Ensemble methods: Foundations and algorithms*. CRC Press.