Securing cloud data with machine learning: trends, gaps, and performance metrics

Blessing Ifeoluwa Omogbehin, Tshiamo Sigwele, Thabo Semong, Aone Maenge, Zhivko Nedev, Hlomani Hlomani
Department of Computing and Informatics, Botswana International University of Science and Technology, Palapye, Botswana

Article Info

Article history:

Received Sep 24, 2025
Revised Nov 24, 2025
Accepted Dec 15, 2025

Keywords:

Adversarial attacks
Cloud computing security
Data access control
Deep learning
Machine learning

ABSTRACT

The increasing reliance on cloud computing has raised significant concerns about the security of data access control, as traditional models are insufficient in managing the dynamic and large-scale nature of cloud environments. This review evaluates machine learning (ML)-based approaches to improve cloud data security, with a particular focus on advancements in anomaly detection and insider threat prevention. Deep learning (DL) models emerge as the most dominant, utilized by 47% of the studies due to their superior ability to process large datasets and adapt to real-time environments. Random forest models are also prominent, being adopted in 20% of the studies for their strong performance in anomaly detection and categorization. TensorFlow stands out as the most widely used tool, featuring in nearly 37% of the reviewed works, while datasets like Amazon Access and computer emergency response team (CERT) are employed in 20% and 13% of the research, respectively. Anomaly detection and prevention are critical priorities, accounting for 41.2% of the research objectives. However, gaps remain, with 21.7% of the studies noting adversarial vulnerabilities and 13% identifying limitations in dataset diversity. The review recommends further development of ML models to address these challenges, expanding dataset diversity, and improving real-time monitoring techniques to enhance cloud data security.

Corresponding Author:

Tshiamo Sigwele
Department of Computer Science and Information Systems
Botswana International University of Science and Technology
Plot 10071, Boseja, Palapye, Botswana
Email: sigwelet@biust.ac.bw

1. INTRODUCTION

In the rapidly evolving landscape of cloud computing, securing data access has become a paramount concern [1], [2]. As organizations increasingly migrate their operations to cloud environments, the need for robust data access control mechanisms is more critical than ever [3], [4]. In cloud computing, data access control refers to the processes and technologies used to ensure that only authorized users have access to the data stored in the cloud [5], [6]. This includes everything from authenticating users, authorizing them to access specific data, monitoring and auditing access to ensure no unauthorized access occurs [7], [8]. It is a technique used to regulate user access to data assets in the cloud storage system [9], this is because a single unauthorized access to cloud data can make a global headline.

One prominent and recent example of failed access control in cloud system was the Synnovis attack. On June 3, 2024, Synnovis a pathology service provider for multiple National Health Service (NHS) trust in United Kingdom (UK) experienced a ransomware attack [10], the Qilin ransom gang locked up the patient data stored on the cloud server, offering a severe service disruption. The Health Insurance Portability and

Accountability Act (HIPAA) journal reported that 1,134 scheduled operations were cancelled, 2,194 outpatients' appointments in the first thirteen days were rescheduled and more than 300 million patient interaction information were leaked to the dark web, these consist of highly confidential data such as result for human immunodeficiency virus (HIV) and cancer, blood group test results and more [11]. NHS England and Synnovis captured in their report that the incident cost a financial damage of 32.7 million Euro [10], [12]. The implications of this attack were substantial, thereby emphasizing the necessity for granular data access control mechanisms in cloud-based systems. A significant advancement would arise, exploring research involving ML-driven data access control in cloud computing.

Machine learning (ML) is an area of artificial intelligence (AI) concerned with designing algorithms that enable computers to learn autonomously from experience and adapt their behavior accordingly [13], [14]. Since ML systems are inherently adaptive, they refine their understanding as new data is introduced [15]. ML algorithms are trained on abundant flow of data gathered over time, and they use this data to identify patterns and make predictions about new data [16], [17]. In the context of cloud data security and access management, ML can be used for real time threat detection [18], pattern recognition, anomaly detection [19], log monitoring [20], and behavioral analysis.

This review provides a comprehensive analysis of the application of ML models in data security and access control within cloud computing environments, focusing on challenges, advancements, and future research directions. The literature reveals significant gaps, including vulnerability to adversarial attacks, dataset limitations, and computational inefficiencies. This review is organized into three main sections: Traditional access control mechanisms, ML-based access control models, and performance evaluation metrics. The contributions of this review are: i) the study determines gaps associated with current approaches in cloud computing access control such as adversarial vulnerabilities, dataset limitations, and computational inefficiencies, and gives future directions; ii) the study identified and recommended key ML research datasets and tools, most adopted and ML models, key evaluation metrics suitable for cloud data access control; iii) the study reviews, analyses and deduces patterns, insights and trends on ML-based access control models used in access control; and iv) the review identifies and recommends deep learning (DL) at 47% literature adoption and random forest (RF) at 20% adoption as the most suitable ML models for cloud data access control.

The structure of this article is as follows. Section 1 is the introduction. In section 2, the literature review approach methodology is presented. In section 3, we present and comprehensively analyze the traditional cloud data access control models. In section 4, we present and perform some analysis on the ML based cloud data access control models. Section 4 presents the trends and insights from the review, while section 5 offers conclusions.

2. METHOD: REVIEW APPROACH

The aim of this review paper is to offer a comprehensive understanding of the current research and advancements in ML techniques for data access control in cloud computing environments. The review focuses on recent publications found in peer-reviewed journals, conference proceedings, and other reputable sources. The review process adheres to a systematic approach, involving the following key steps:

i) Literature searching: utilizing online academic databases such as ScienceDirect, SpringerLink, IEEE Xplore, Google Scholar, and Elsevier's Mobile Edge Computing journal to identify pertinent articles. Keywords including "machine learning," "data access control," "cloud computing," and "cloud security," were employed.
ii) Inclusion criteria: articles specifically addressing the scope from 5 years will be included.
iii) Exclusion criteria: articles not focused on data access control or cloud computing or machine learning, or those unrelated to these terms, were excluded.
iv) Data extraction: key information was extracted from selected articles, encompassing title, authors and publication years, ML models, research objectives, evaluation metrics, model limitations, and datasets and tools utilized.
v) Synthesis: extracted information was organized into a summarized table, followed by the analysis with the extraction of trends and patterns concerning the most adopted models, objectives, metrics, tools, and datasets.

3. TRADITIONAL CLOUD DATA ACCESS CONTROL MODELS

Table 1 shows various traditional cloud data access control frameworks, summarizing their research objectives, limitations, and the tools/techniques employed from concurrent studies. These frameworks, including intrusion detection prevention system (IDPS), asymmetric encryption model (AEM), attribute-based access control (ABAC), and hybrid models, aim to enhance data security and prevent unauthorized access. The table highlights key issues such as susceptibility to adversarial attacks, implementation challenges, scalability concerns, and user experience problems, providing comprehensive insights into the effectiveness and drawbacks of each model in different cloud environments.

Table 1. Traditional data access control models

S/N	Authors and years	ML access control models	ML models research objectives	Evaluation metrics and performance
1.	Kumar *et al.* [21]	IDPS	Detection of malicious behavior over the network, Enhance the confidentiality, integrity and availability (CIA).	- Prone to adversarial attacks. - Access denial to legitimate users.
2.	Brandão [22]	AEM	Prevention of unauthorize Access to cloud system.	Prone to adversarial attacks.
3.	Khan [23]	ABAC	Protection of file upload, file download and file deletion.	Ignores software and network security.
4.	He *et al.* [24]	ABAC	Prevention of unauthorized access to cloud environment.	- Utilizes single authority - Prone to Privilege escalation.
5.	Bhatt and Sandhu [25]	ABAC	To secure accesses and data flow between various users in the cyber space.	Neglects real-world application/testing
6.	Prantl *et al.* [26]	ABAC	- To provide data confidentiality. - To prevent illegal sharing of authentication keys.	- Computationally intensive. - Poor user experience.
7.	Kumar and Verma [27]	ABAC	- Time bound data access. - Biometric defense mechanisms.	- Poor user experience. - Utilizes single authority.
8.	Choudhary and Singh [28]	Hybrid model: query-based role and attribute access control (QRAAC), Role based access control (RBAC), task-based access control (TBAC)	Enhance the CIA triad of cloud's data.	- Access denial to legitimate users. - Privilege escalation.
9.	Dayana and Rani [29]	RBAC	- To prevent access policy violation. - Prevention of data linkage in cloud environments.	- Lacks protection against privilege escalation.
10.	Kumar *et al.* [30]	Hybrid model: symmetric/ asymmetric cryptography	Prevention of malicious access to resources in the cloud environments.	- Computational limitation - Implementation problem due to the dynamicity of cloud platforms.

3.1. Analysis of traditional cloud computing data access control frameworks

Figure 1 shows that ABAC is the most utilized model for data access control in cloud computing, representing 50% of the usage. The hybrid model follows at 20%, with RBAC, IDPS, and AEM each at 10%. ABAC's dominance is attributed to its scalability, dynamic nature, flexibility in attribute-based access, efficiency with attribute-based rules engines, fine-grained control, and robust security incident response capabilities. These features make ABAC particularly suited for managing large, dynamic cloud environments.

3.2. Analysis of traditional tools for cloud computing data access control frameworks

Figure 2 shows the distribution of tools and techniques used in traditional cloud computing data access control frameworks. It indicates that 50% of the literature employs the use of attributes, highlighting their importance in cloud security. Cryptography accounts for 20%, while intrusion detection systems (IDS), encryption, and purpose-based trust access control (PbTAC) each make up 10%. The dominance of attribute-based methods underscores their effectiveness in providing fine-grained access control and enhancing security in cloud environments.

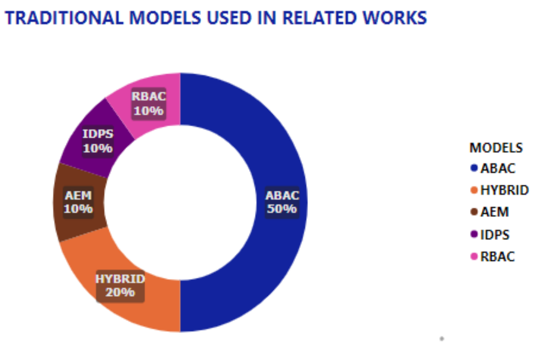

Figure 1. Analysis of traditional cloud computing data access control frameworks

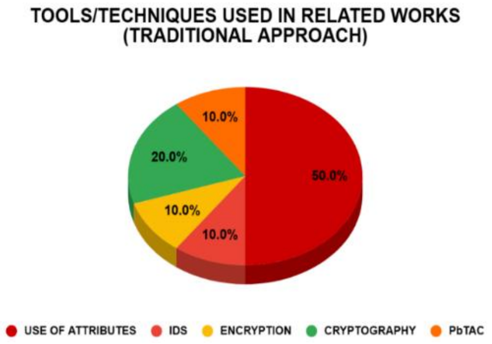

Figure 2. Analysis of traditional tools for cloud computing data access control frameworks

3.3. Analysis of framework objectives in traditional cloud computing data access

Figure 3 shows the research objectives in traditional cloud computing data access control, with 28.6% focusing on unauthorized access restriction, the most dominant objective. Enhancing the CIA triad is the focus of 21.4%, followed by access security and detection of malicious behavior at 14.3% each. Prevention of access policy violation, illegal sharing of keys, and provision of biometric defense each account for 7.1%. This distribution emphasizes the critical importance of robust access control measures in ensuring cloud data security and integrity.

3.4. Analysis of framework gaps in traditional cloud computing data access control

Figure 4 shows the research limitations in traditional cloud computing data access control frameworks. The most significant limitation is vulnerability to adversarial attacks, affecting 20% of the reviewed literature. Computational intensity, poor user experience, and reliance on a single authority each account for 13.3%. Other limitations, each at 6.7%, include focus on theoretical foundations, privilege escalation, neglect of network/software security, and the potential benefits of integrating ML. These limitations highlight the necessity for more robust, efficient, and user-friendly access control solutions.

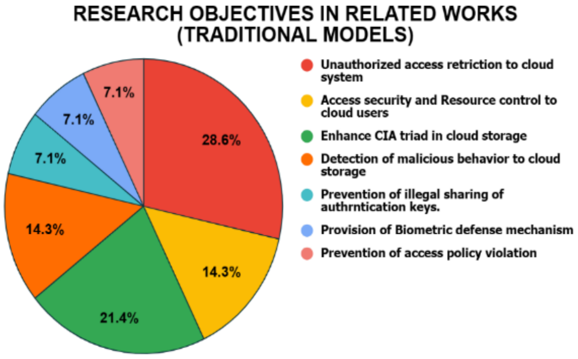

Figure 3. Analysis of framework objectives in traditional cloud computing data access

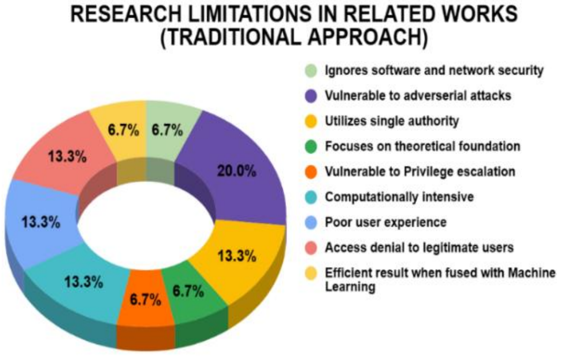

Figure 4. Analysis of framework gaps in traditional cloud computing data access control

4. MACHINE LEARNING BASED CLOUD DATA ACCESS CONTROL MODELS

Table 2 provides a comprehensive overview of various ML models applied to data access control in cloud computing environments. It includes current key studies detailing the specific ML algorithms used, such as LightGBM, multilayer perceptron (MLP), decision trees (DT), RF, linear regression (LR), deep neural networks (DNN), and support vector machines (SVM). The table outlines each study's research objectives, evaluation metrics, performance outcomes, and limitations. Additionally, it outlines the datasets and tools utilized in these studies, highlighting the advancements and challenges in implementing ML for enhancing data access security in cloud systems. This summary aids in understanding the current state and future directions of ML-driven cloud security research. The table will be comprehensively analyzed in the subsequent sections highlighting trends, patterns and insights.

The following sections present an in-depth analysis of the literature review in Table 2. This section provides detailed discussions of the most implemented ML models used in related works and the reasons behind their adoption. It also includes a summary of research objectives, summary of summary of research limitations, and the most adopted datasets and tools.

4.1. Analysis of ML models adopted from literature

As shown in Figure 5, authors at different years used ML models like DT, DL, RF, LR, and SVM to carried out their various research objectives. The chart shows that DL is the most utilized ML model in related works, accounting for 47.06%. RF follows at 17.65%, with DT, LR, and SVM each at 11.76%. This indicates a strong preference for DL due to it is advanced capabilities in handling complex data. DL is ideal for cloud data access control due to it is real-time data processing, flexibility, and ability to learn complex patterns [31], [32]. It excels in anomaly detection, security analysis [33], and creating secure encryption algorithms, making it crucial for protecting against unauthorized access.

Table 2. Data access control models using ML

S/N	Authors and years	ML access control models	ML models research objectives	Evaluation metrics and performance	ML model limitation	ML models datasets and tools
1	Mehmood *et al.* [34]	LightGBM algorithm DT	Cloud insider/internal threat detection	Accuracy: 97%, Precision rate: 97%, F1 score: 0.95, Recall: 83%	- Ignores behavioral biometric attack.	Datasets: CERT datasets. Tools: Not specified
2	Kanaker *et al.* [35]	- Multilayer perceptron algorithm DNN - Regression model	Prevention and detection of malicious access to cloud storage system	Recall: (95.9%), Accuracy: (95.86%), Precision rate: (95.9%), F-measure: (0.955), ROC: (97.1), false positive rate (FPR): (29.1%)	- Accuracy compromise. - Different validation accuracy.	Dataset: 10-fold cross-validation dataset. Tools: Weka.
3	Khilar *et al.* [36]	DT algorithm	To ensure authorized users access trusted cloud resources.	Accuracy: 90.3%, Precision: 90%, F1-score: 0.90, Time: 0.35, Recall: 90%, Root mean absolute error (RMAE): 31.1%	- Prone to overfitting. - Computational limitation.	Dataset: Apache server log for data. Tools: TensorFlow, Scikit learn
4	Akoto and Salman [37]	RF	Detection and categorization of Anomalies in cloud systems.	Accuracy: 99%, Precision rate: 93.6%, FPR: 1.9%, Undetection (UND) rate: 0.4%	- Low backdoor attack accuracy. - False attack detection - Limited dataset.	Dataset: UNSW dataset. Tools: not Specified
5	Afshar *et al.* [38]	- LR algorithm - RF algorithm	- To protect resources from authorized access requests. - To detect and prevent internal breaches	Accuracy: 99.62%, Total number of T-violation: 5 and 2, Total number of P-Violation: 67 and 11	- Limited external threat detection which would restrict comprehensive security.	Dataset: Amazon access sample dataset. Tools: TensorFlow
6	Nguyen *et al.* [39]	DNN	- Accurate detection and prevention of multiple cloud attacks in real time with small computation.	Accuracy: 99.93%, True positive rate (TPR)/Recall: 99.57%, FPR: 0.04, true negative rate (TNR): 99.6%	- Lacks robust adversarial attack defenses.	Dataset: CICIDS2017 dataset. Tools: Keras, TensorFlow, T-shark.
7	Liu *et al.* [40]	RF algorithm	- Enhance access decision-making. - To maintain system performance.	Accuracy: 92.6%, TPR/Recall: 91.6%, Precision rate: 93.4%, F-measure: 0.925	- Requires more trees for accurate prediction	Dataset: Amazon access dataset. Tools: Python 3.6
8	Nobi *et al.* [41]	DL	Unauthorized restriction to data in cloud storage	TPR/Recall:95%, F-measure:0.95, Precision: 95%, FPR: 0.05, Accuracy: 95%	- Adversarial attack - Bias and human errors in training data	Datasets: Amazon dataset, synthetic TensorFlow, Keras
9	El-Kassabi *et al.* [42]	DL	Detection of anomalies in cloud workflow.	Accuracy: 96.14%, Precision rate: 93%, TPR/Recall: 99%, F1-score: 0.96	- Large data collection. - Complex, Challenges.	Datasets: COVID-19. Tools: Pytorch and Scikit-learn
10	Alheeti *et al.* [43]	SVM algorithm	Detection and prevention of malicious access	Accuracy: 99.92%, Precision rate: 96%, Recall: 97%, F1-score: 0.99	- Computational - Limitation	Datasets: CIDD datasets). Tools: Scikit-learn.
11	Anakath *et al.* [44]	Deep belief neural network	Cloud insider/internal threat detection	Accuracy: 99%, F-measure: 0.98, Precision: 100%, Recall: 99%	- Requires intensive computation - Time consuming.	Dataset: Open-source datasets. Tool: TensorFlow
12	Chehab and Mourad [45]	DL	Detection and Prevention of malicious access	Accuracy: 90%, Precision: 96%, Recall: 96%, F-measure: 0.96	- Too complex to implement. - Model not scalable.	Synthetic dataset. TensorFlow, Scikit-learn
13	Jiang *et al.* [46]	DL	Attack detection, classification and prevention.	Accuracy: 99.23%, FPR: 9.86, Recall/TPR: 99.23%	- Lacks robust adversarial defenses. - Dataset limitation	Dataset: KDD99 dataset, Tools: TensorFlow
14	Ferhi *et al.* [47]	DL	Denial of Service detection and prevention	Accuracy: 99.90%, Precision: 95.6, Recall: 99.58, F-measure: 97.58	- Computational limitation. - High training time.	Dataset: CSE-CIC-IDS2018 dataset. Tools: Scikit-learn
15	Padmavathi *et al.* [48]	SVM	Malicious insider threat detection	True detection rate (TDR): 100%, Precision: 100%, F-measure: 100%, Threshold value: 50%	- Limited dataset	Dataset: CERT dataset Tools: Python

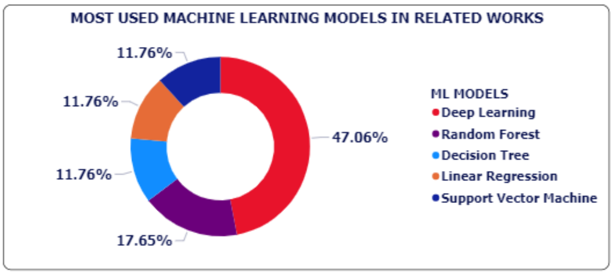

Figure 5. ML models adoption on cloud data access control

4.2. Analysis of ML datasets adopted from literature

Figure 6 illustrates the most used datasets in related works, with the Amazon access sample dataset leading at 20%. This is followed by the CERT dataset at 13.3%, and several datasets like Apache server log data, UNSW datasets, 10-fold cross-validation dataset, CICIDS2017, COVID-19, CIDD, open-source datasets, synthetic datasets, KDD99, and CSE-CICIDS2018 each used at 6.7%. This distribution indicates a diverse usage of datasets in research, with a significant preference for Amazon access sample datasets, followed by CERT datasets. Amazon access sample dataset prevailed due to it is comprehensive, representative nature of real-world data and relevance to access control rule and policy [19]. It is standardized for easy comparison, comes with detailed descriptions, and is publicly and freely available, making it accessible and attractive to researchers on a budget.

4.3. Analysis of ML tools adopted from literature

Figure 7 illustrates the distribution of ML tools employed in related studies on data access control. TensorFlow emerges as the most widely used framework, appearing in 36.84% of the reviewed works. This is followed by Scikit-learn at 26.32%, while both Keras and Python are each utilized in 10.53% of the studies. PyTorch, T-Shark, and Weka are less frequently adopted, each featuring in 5.26% of the cases. These findings underscore TensorFlow’s prominence in this area of research. TensorFlow can develop systems to generate security alerts, monitor and adjust access policies, and modify privileges based on behavior [36], [38]. Compatible with Google Cloud, AWS, and Azure, it creates efficient models. It identifies and blocks malicious activities, processes real-time data, handles large datasets, and categorizes data by sensitivity [41].

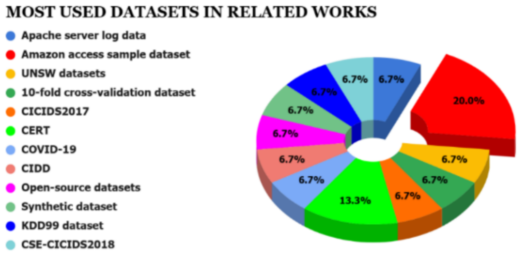

Figure 6. Analysis of ML-based cloud access control datasets adopted from literature

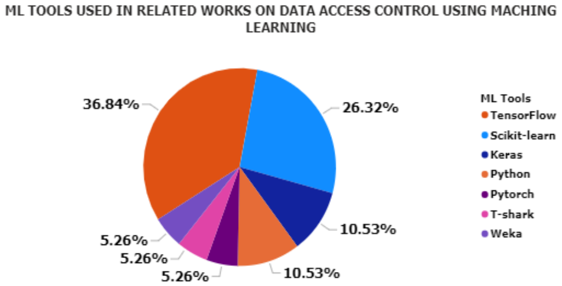

Figure 7. Analysis of ML-based cloud access control tools adopted from literature

4.4. Analysis of ML frameworks objectives adopted from literature

Figure 8 shows the various research objectives adopted in the literature concerning ML frameworks. The primary focus is on anomalies detection and prevention (41.2%), indicating a significant emphasis on identifying and mitigating unusual behaviors in cloud environments. Unauthorized access restriction and denial of service detection and prevention each account for 17.6%, highlighting the importance of safeguarding cloud resources from unauthorized access and service disruptions. Cloud insider/internal threat detection, enhancing decision-making in granting access, behavioral attack prevention, and performance optimization of cloud access control each represent 5.9%. This distribution underscores the diverse challenges in cloud security, with a predominant focus on anomaly detection and preventive measures to ensure robust data access control.

4.5. Analysis of ML cloud access control gaps from literature

Figure 9 shows key research gaps in ML-based cloud access control, with adversarial vulnerabilities at 21.7%, complexity at 17.4%, and dataset limitations at 13%. Accuracy compromise, overfitting, and computational limitations each account for 8.7%. Challenges in implementation, false attack detection, behavioral biometrics attacks, and limited external threat detection each represent 4.3%. The most significant gap is adversarial vulnerabilities, due to the increasing sophistication of attacks that can exploit model weaknesses and the critical need for robust defenses.

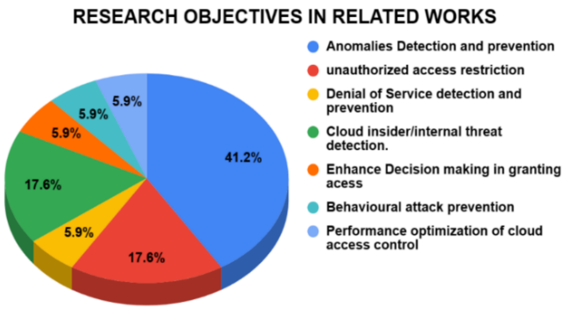

Figure 8. Analysis of ML-based cloud access control framework objectives from literature

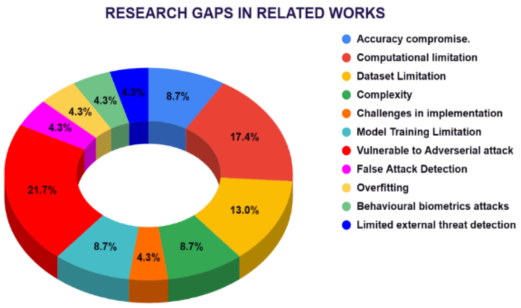

Figure 9. Analysis of ML cloud access control gaps from literature

4.6. Performances metrics for ML-based cloud data access control frameworks

Table 3 shows critical performance metrics for ML-based cloud data access control frameworks, summarizing their definitions, equations, and referenced authors. Metrics like accuracy, precision, FPR, F-measure, recall, TNR, RMAE, false negative rate (FNR), TDR, receiver-operating characteristics (ROC), time-violation (T-violation), privacy-violation (P-violation), training time, and threshold value are essential for evaluating model effectiveness and reliability. These metrics help analyze and identify patterns, providing insights into model performance and data protection, as defined and supported by various studies. Performance metrics are measures that are used to evaluate the efficacy of a ML model in making predictions [49], [50]. Table 3 illustrates the various metrics used to quantify how effective the different models used in related works.

Figure 10 shows the most adopted ML performance metrics in related works. ML-based cloud data access control frameworks, with accuracy being the most utilized at 22%, followed by precision at 20%, recall at 18%, and F-measure at 17%. These metrics are critical for evaluating the effectiveness of data access control models in cloud computing. Accuracy ensures correct predictions and minimizes miscalculations that could lead to data breaches [46], [47], [51]. Precision focuses on the quality of positive predictions, minimizing false alarms [44], [45]. Recall captures unauthorized access events efficiently [46], [47], and F-measure balances precision and recall, optimizing overall model performance [42], [43].

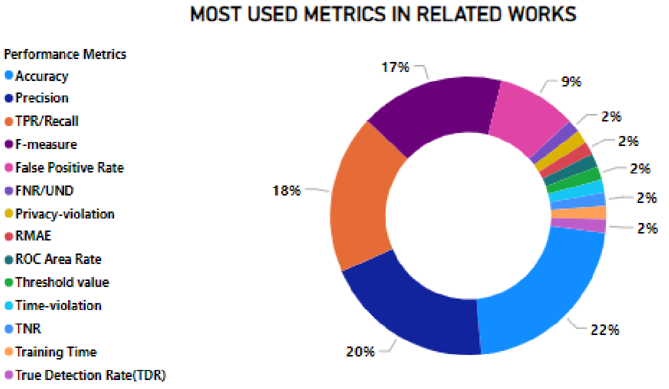

Figure 10. Most adopted ML performance metrics

Table 3. Performances metrics for ML-based cloud data access control frameworks

No.	Metrics	Definition	Equation	Authors
1	Accuracy	Measures the predictive power of the model e.g. ability to predict unauthorize access to data in cloud storage systems [51]. The best value when measured in percentage is 100%.	$\frac{TP+TN}{N}$	[34], [35], [36], [37] [38], [39], [40], [41] [42], [43], [44], [45] [46], [47]
2	Precision	Measures the quality of a positive prediction made by the model that's, the portion of the data point our model says existed in the relevant class that are indeed relevant [20].	$\frac{TP}{TP+FP}$	[34], [35], [36], [37] [39], [40], [41], [42] [43], [44], [45], [47] [48]
3	FPR also known as false alarm rate (FAR)	Measures the fraction of negative access that are misclassified as positive access to data in the cloud storage. This metric evaluates the rate at which an access control system generates FAR [52].	$\frac{FP}{FP+TN}$	[35], [37], [39], [41] [46], [48]
4	F-measure	Measures the average of precision and recall, the best value of F-score is one (1) and its worst value is Zero (0). It shows the actual performance of a model especially in the case of imbalance dataset [14].	$2 \times \frac{Precision \times Recall}{Precision + Recall}$	[34], [35], [36], [40] [41], [42], [43], [44] [45], [47]
5	Recall/TPR detection rate (DR)	Measures the fraction of positive access that are correctly classified as positive access to data in cloud storage [46], [47].	$\frac{TP}{TP+FN}$	[34], [35], [36], [39] [40], [41], [42], [43] [44], [45], [46], [47]
6	TNR	Measures the fraction of negative access (unauthorized access) that are correctly classified [15].	$\frac{TN}{TN+FP}$	[39]
7	RMAE	Ensures that model performance is accessed accurately while adhering to data privacy and compliance with access policy [53].	$\sqrt{\frac{1}{N}\sum_{i-1}^{N}\lvert Y_{true,i} - Y_{pred,i}\rvert}$	[36]
8	FNR undetected rate	Measures the fraction of positive access that are misclassified as negative access to data in the cloud storage [15], [37].	$\frac{FN}{FN+TP}$	[37]
9	TDR	Measures the proportion of positive instances that are not incorrectly classified as negative. It is the complement of the FNR [17].	1- FNR	[48]
10	ROC	Measures the trade-off between FPR and TPR of the security system [54].	No specific formula for calculation.	[35]
11	T-violation	Measures the risk of a security system being compromised by an attacker exploiting a discrepancy in the timing of events [55].	No specific formula for calculation.	[38]
12	P-violation	Measures the action or behavior that violates access policy e.g. privacy of a user. (whether intentional or unintentional) [56].	No specific formula for calculation	[38]
13	Training time	Measures the amount of time that it takes to train a model [57].	Depends on: dataset size, model type, CPU available, optimization techniques	[36]
14	Threshold value	Helps determine the cut-off point for deciding between granting or denying access [58].	No specific formula for calculation	[48]

TP=true positive, FN=false negative, N=total numbers of samples, Y_pred=predicted label, TN=true negative, FP=false positive, Y_true=true label

5. TRENDS, PATTERNS AND INSIGHTS

This section highlights key trends, patterns, and insights identified from the literature review on data access control in cloud computing using ML models. These insights provide a comprehensive understanding of the current state of research and the prevailing directions in this field. They are as follows; i) there is a growing preference for DL models due to their superior ability to process large datasets and adapt to dynamic cloud environments; ii) numerous studies highlight the critical importance of detecting anomalies and insider threats, showcasing a proactive approach to enhancing cloud security; iii) the employment of various datasets and tools underscores the necessity for comprehensive data to effectively train ML models and demonstrates the versatility of tools like TensorFlow, Keras, and Scikit-learn; and iv) accuracy and precision are prioritized as the most essential metrics, indicating the need for reliable and precise models to effectively prevent unauthorized access.

6. CONCLUSION

The literature review on data access control in cloud computing reveals significant trends and insights in the application of ML models. DL models, especially neural networks, dominate ML-based frameworks, comprising 47.06% of the models used, due to their robust performance in handling large datasets, real-time processing, and anomaly detection. The Amazon access sample dataset is the most frequently used, reflecting a preference for comprehensive, real-world data. TensorFlow is the leading tool, used in 36.84% of studies, highlighting its capability to develop and deploy complex ML models across various cloud platforms. In traditional data access control frameworks, ABAC is predominant, representing 50% of usage from 2013 to 2023, owing to its scalability, flexibility, and fine-grained control. This method effectively manages large, dynamic cloud environments by incorporating detailed attribute-based rules and robust incident response mechanisms. Attribute-based methods, used in 50% of traditional frameworks, emphasize precise control over data access. The primary research objective in traditional frameworks is unauthorized access restriction, comprising 28.6% of the objectives, underscoring the need for robust access control measures. Performance metrics for ML-based frameworks highlight the critical importance of accuracy, adopted in 22% of studies, ensuring precise predictions and minimizing errors. Precision and recall, used in 20% and 18% of studies respectively, stress the balance between reducing false positives and comprehensive threat detection. The F-measure, adopted in 17% of studies, provides a balanced view of model performance, crucial for optimizing security in cloud environments. Research directions include enhancing DL models, defending against adversarial attacks, integrating more robust ML with traditional methods, improving datasets, optimizing performance, real-time monitoring, ensuring explainability, and scalability.

ACKNOWLEDGEMENTS

The authors express their sincere appreciation to the Botswana International University of Science and Technology for providing access to vital resources and facilities.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

In an effort to acknowledge the distinct contributions of each author, provide transparency in authorship and strengthen collaborative research, this journal implements the Contributor Roles Taxonomy (CRediT). Each author's contribution is as follows.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Blessing Ifeoluwa Omogbehin	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Tshiamo Sigwele	✓	✓		✓	✓				✓	✓	✓	✓	✓	
Thabo Semong	✓			✓	✓					✓		✓		
Aone Maenge						✓				✓	✓			
Zhivko Nedev				✓		✓				✓		✓		
Hlomani Hlomani				✓	✓					✓		✓		

C	: **C**onceptualization	I	: **I**nvestigation	Vi	: **Vi**sualization
M	: **M**ethodology	R	: **R**esources	Su	: **Su**pervision
So	: **So**ftware	D	: **D**ata Curation	P	: **P**roject administration
Va	: **Va**lidation	O	: Writing - **O**riginal Draft	Fu	: **Fu**nding acquisition
Fo	: **Fo**rmal analysis	E	: Writing - Review & **E**diting		

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

No new data was collected or analyzed for this study. The article is based entirely on previously published literature, all of which is appropriately cited in the References section.

REFERENCES

[1] R. Sikka and M. Ojha, "An overview of cloud computing," *International Journal of Innovative Research in Computer Science and Technology*, vol. 15, no. 3, pp. 135–138, 2021, doi: 10.55524/ijircst.2021.9.6.31.

[2] S. Aboukadri, A. Ouaddah, and A. Mezrioui, "Machine learning based identity and access management systems (ML I&AM): a taxonomy," *Colloquium in Information Science and Technology, CIST*, pp. 657–662, 2023, doi: 10.1109/CiSt56084.2023.10409872.

[3] O. D. Segun-Falade, O. S. Osundare, W. E. Kedi, P. A. Okeleke, T. I. Ijomah, and O. Y. Abdul-Azeez, "Assessing the transformative impact of cloud computing on software deployment and management," *Computer Science and IT Research Journal*, vol. 5, no. 8, pp. 2062–2082, 2024, doi: 10.51594/csitrj.v5i8.1492.

[4] O. C. Adeusi, Y. O. Adebayo, P. A. Ayodele, T. T. Onikoyi, K. B. Adebayo, and I. O. Adenekan, "IT standardization in cloud computing: Security challenges, benefits, and future directions," *World Journal of Advanced Research and Reviews*, vol. 22, no. 3, pp. 2050–2057, 2024, doi: 10.30574/wjarr.2024.22.3.1982.

[5] O. Godwin and M. O. Musa, "Challenges and strategies for enhancing ICT security in public institutions," *International Journal of Innovative Science and Research Technology (IJISRT)*, pp. 2185–2190, 2024, doi: 10.38124/ijisrt/ijisrt24jul1024.

[6] S. Namasudra, "Data access control in the cloud computing environment for bioinformatics," *International Journal of Applied Research in Bioinformatics*, vol. 11, no. 1, pp. 40–50, 2021, doi: 10.4018/ijarb.2021010105.

[7] A. A. S. Alqahtani and T. Alshayeb, "Zero-effort two-factor authentication using wi-fi radio wave transmission and machine learning," in *2023 IEEE 13th Annual Computing and Communication Workshop and Conference, CCWC 2023*, 2023, pp. 313–318. doi: 10.1109/CCWC57344.2023.10099124.

[8] P. Kamboj, S. Khare, and S. Pal, "User authentication using Blockchain based smart contract in role-based access control," *Peer-to-Peer Networking and Applications*, vol. 14, no. 5, pp. 2961–2976, 2021, doi: 10.1007/s12083-021-01150-1.

[9] M. Moravcik and L. Zidekova, "Overview of access control mechanisms in cloud environments," in *ICETA 2024-22nd Year of International Conference on Emerging eLearning Technologies and Applications, Proceedings*, IEEE, 2024, pp. 465–470. doi: 10.1109/ICETA63795.2024.10850806.

[10] M. Dollar, "01 July 2024: update on cyber incident," *Synnovis*. Accessed: Mar. 29, 2025. [Online]. Available: https://www.synnovis.co.uk/cyberattack-information-centre

[11] S. Alder, "NHS Pathology Provider synnovis notifies organizations affected by June 2024 ransomware attack," *The HIPAA Journal*. Accessed: Mar. 30, 2024. [Online]. Available: https://www.hipaajournal.com/care-disrupted-at-london-hospitals-due-to-ransomware-attack-on-pathology-vendor/

[12] NHS England, "Synnovis ransomware cyber-attack," *NHS England-London*. Accessed: Mar. 29, 2024. [Online]. Available: https://www.england.nhs.uk/london/synnovis-ransomware-cyber-attack/

[13] N. T. G. Anthony, M. Shafik, F. Kurugollu, and H. F. Atlam, "Anomaly detection system for ethereum blockchain using machine learning," *Advances in Transdisciplinary Engineering*, vol. 25, pp. 311–316, 2022, doi: 10.3233/ATDE220608.

[14] T. A. Al-Shehari *et al.*, "Enhancing insider threat detection in imbalanced cybersecurity settings using the density-based local outlier factor algorithm," *IEEE Access*, vol. 12, pp. 34820–34834, 2024, doi: 10.1109/ACCESS.2024.3373694.

[15] H. Nandanwar and R. Katarya, "Deep learning enabled intrusion detection system for Industrial IoT environment," *Expert Systems with Applications*, vol. 249, 2024, doi: 10.1016/j.eswa.2024.123808.

[16] J. Su *et al.*, "Large language models for forecasting and anomaly detection: a systematic literature review," *arxiv: 2402.10350*, 2024.

[17] S. Ahmad, S. Mehfuz, S. Urooj, and N. Alsubaie, "Machine learning-based intelligent security framework for secure cloud key management," *Cluster Computing*, vol. 27, no. 5, pp. 5953–5979, 2024, doi: 10.1007/s10586-024-04288-8.

[18] D. A. Winkler, "Role of artificial intelligence and machine learning in nanosafety," *Small*, vol. 16, no. 36, pp. 1–6, 2020, doi: 10.1002/smll.202001883.

[19] B. I. Omogbehin, T. Sigwele, and T. Semong, "Leveraging deep learning for access control and data security in cloud environments," in *2024 IEEE 4th International Conference on ICT in Business Industry and Government, ICTBIG 2024*, Indore, India, 2024, pp. 1–5. doi: 10.1109/ICTBIG64922.2024.10911855.

[20] P. Srinivas, F. Husain, A. Parayil, A. Choure, C. Bansal, and S. Rajmohan, "Intelligent monitoring framework for cloud services: a data-driven approach," in *ACM International Conference Proceeding Series*, 2024, pp. 381–391. doi: 10.1145/3639477.3639753.

[21] K. Kumar, G. Nanak, and K. Kumar, "Intrusion detection and prevention system in enhancing security of cloud environment," *International Journal of Advanced Research in Computer Engineering and Technology*, vol. 6, no. 8, pp. 2278–1323, 2017.

[22] P. R. Brandão, "The importance of authentication and encryption in cloud computing framework security," *International Journal on Data Science and Technology*, vol. 4, no. 1, 2018, doi: 10.11648/j.ijdst.20180401.11.

[23] A. R. Khan, "Access control in cloud computing environment," *ARPN Journal of Engineering and Applied Sciences*, vol. 7, no. 5, pp. 613–615, 2012.

[24] H. He, L. han Zheng, P. Li, L. Deng, L. Huang, and X. Chen, "An efficient attribute-based hierarchical data access control scheme in cloud computing," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, 2020, doi: 10.1186/s13673-020-00255-5.

[25] S. Bhatt and R. Sandhu, "ABAC-CC: Attribute-based access control and communication control for internet of things," *ACM Symposium on Access Control Models and Technologies, SACMAT*, 2020, pp. 203–212. doi: 10.1145/3381991.3395618.

[26] T. Prantl *et al.*, "Towards a cryptography encyclopedia: a survey on attribute-based encryption," *Journal of Surveillance, Security and Safety*, vol. 4, no. 4, pp. 129–54, 2023, doi: 10.20517/jsss.2023.30.

[27] A. Kumar and G. Verma, "Securing cloud access with enhanced attribute-based cryptography," *Computing*, vol. 106, no. 12, pp. 4193–4207, 2024, doi: 10.1007/s00607-023-01212-7.

[28] S. Choudhary and N. Singh, "Analysis of security-based access control models for cloud computing," *International Journal of Cloud Applications and Computing*, vol. 12, no. 1, pp. 1–19, 2022, doi: 10.4018/IJCAC.2022010104.

[29] K. R. Dayana and P. S. Rani, "Trust aware cryptographic role based access control scheme for secure cloud data storage," *Automatika*, vol. 64, no. 4, pp. 1072–1079, 2023, doi: 10.1080/00051144.2023.2243144.

[30] S. Kumar, G. Karnani, M. S. Gaur, and A. Mishra, "Cloud security using hybrid cryptography algorithms," in *2021 2nd International Conference on Intelligent Engineering and Management, ICIEM 2021*, 2021, pp. 599–604. doi: 10.1109/ICIEM51511.2021.9445377.

[31] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. Robinson, "Deep learning for unsupervised insider threat detection in structured cybersecurity data streams," *AAAI Workshop-Technical Report*, pp. 224–234, 2017.

[32] M. Y. Shakor and M. Ibrahim Khaleel, "Modern deep learning techniques for big medical data processing in cloud," *IEEE Access,* vol. 13, pp. 62005-62028, 2025, doi: 10.1109/ACCESS.2025.3556327.

[33] A. R. Al-Ghuwairi, Y. Sharrab, D. Al-Fraihat, M. AlElaimat, A. Alsarhan, and A. Algarni, "Intrusion detection in cloud computing based on time series anomalies utilizing machine learning," *Journal of Cloud Computing*, vol. 12, no. 1, 2023, doi: 10.1186/s13677-023-00491-x.

[34] M. Mehmood, R. Amin, M. M. A. Muslam, J. Xie, and H. Aldabbas, "Privilege escalation attack detection and mitigation in cloud using machine learning," *IEEE Access*, vol. 11, pp. 46561–46576, 2023, doi: 10.1109/ACCESS.2023.3273895.

[35] H. Kanaker, N. A. Karim, S. A. B. Awwad, N. H. A. Ismail, J. Zraqou, and A. M. F. Al ali, "Trojan horse infection detection in cloud based environment using machine learning," *International Journal of Interactive Mobile Technologies*, vol. 16, no. 24, pp. 81–106, 2022, doi: 10.3991/ijim.v16i24.35763.

[36] P. M. Khilar, V. Chaudhari, and R. R. Swain, "Trust-based access control in cloud computing using machine learning," *Cloud Computing for Geospatial Big Data Analytics*. Cham, Switzerland: Springer, 2019, pp. 55–79 doi: 10.1007/978-3-030-03359-0_3.

[37] J. Akoto and T. Salman, "Machine learning vs deep learning for anomaly detection and categorization in multi-cloud environments," in *2022 IEEE Cloud Summit, Cloud Summit 2022*, 2022, pp. 44–50. doi: 10.1109/CloudSummit54781.2022.00013.

[38] M. Afshar, S. Samet, and H. Usefi, "Incorporating behavior in attribute based access control model using machine learning," *2021 IEEE International Systems Conference (SysCon),* Vancouver, Canada, 2021, pp. 1-8, doi: 10.1109/SysCon48628.2021.9447115.

[39] X. H. Nguyen, X. D. Nguyen, H. H. Huynh, and K. H. Le, "Realguard: A lightweight network intrusion detection system for IoT Gateways," *Sensors*, vol. 22, no. 2, pp. 1–18, 2022, doi: 10.3390/s22020432.

[40] A. Liu, X. Du, and N. Wang, "Efficient access control permission decision engine based on machine learning," *Security and Communication Networks*, vol. 2021, 2021, doi: 10.1155/2021/3970485.

[41] M. N. Nobi, R. Krishnan, Y. Huang, M. Shakarami, and R. Sandhu, "Toward deep learning based access control," in *CODASPY 2022-Proceedings of the 12th ACM Conference on Data and Application Security and Privacy*, Association for Computing Machinery, 2022, pp. 143–154. doi: 10.1145/3508398.3511497.

[42] H. T. El-Kassabi, M. A. Serhani, M. M. Masud, K. Shuaib, and K. Khalil, "Deep learning approach to security enforcement in cloud workflow orchestration," *Journal of Cloud Computing*, vol. 12, no. 1, 2023, doi: 10.1186/s13677-022-00387-2.

[43] K. M. A. Alheeti, A. A. A. Lateef, A. Alzahrani, A. Imran, and D. Al Dosary, "Cloud intrusion detection system based on SVM," *International Journal of Interactive Mobile Technologies*, vol. 17, no. 11, pp. 101–114, 2023, doi: 10.3991/ijim.v17i11.39063.

[44] A. S. Anakath, R. Kannadasan, N. P. Joseph, P. Boominathan, and G. R. Sreekanth, "Insider attack detection using deep belief neural network in cloud computing," *Computer Systems Science and Engineering*, vol. 41, no. 2, pp. 479–492, 2022, doi: 10.32604/csse.2022.019940.

[45] M. Chehab and A. Mourad, "LP-SBA-XACML: lightweight semantics based scheme enabling intelligent behavior-aware privacy for IoT," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 1, pp. 161–175, 2022, doi: 10.1109/TDSC.2020.2999866.

[46] F. Jiang *et al.*, "Deep learning based multi-channel intelligent attack detection for data security," *IEEE Transactions on Sustainable Computing*, vol. 5, no. 2, pp. 204–212, 2020, doi: 10.1109/TSUSC.2018.2793284.

[47] A. Mansoor, M. Anbar, A. A. Bahashwan, B. A. Alabsi, and S. D. A. Rihan, "Deep learning-based approach for detecting DDoS attack on software-defined networking controller," *Systems,* vol. 11, no. 6, doi: 10.3390/systems11060296.

[48] G. Padmavathi, D. Shanmugapriya, and S. Asha, "A framework to detect the malicious insider threat in cloud environment using supervised learning methods," in *2022 9th International Conference on Computing for Sustainable Global Development, INDIACom 2022*, 2022, pp. 354–358. doi: 10.23919/INDIACom54597.2022.9763205.

[49] M. N. Nobi, M. Gupta, L. Praharaj, M. Abdelsalam, R. Krishnan, and R. Sandhu, "Machine learning in access control: a taxonomy and survey," *arxiv: 2207.01739*, vol. 1, no. 1, 2022.

[50] S. Aboukadri, A. Ouaddah, and A. Mezrioui, "Machine learning in identity and access management systems: survey and deep dive," *Computers and Security*, vol. 139, 2024, doi: 10.1016/j.cose.2024.103729.

[51] M. El Moudni and E. Ziyati, "A hybrid approach to enhancing personal sensitive information protection in the context of cloud storage," *International Journal of Computing and Digital Systems*, vol. 17, no. 1, pp. 1–14, 2025, doi: 10.12785/ijcds/1571016609.

[52] E. Silambarasan, R. Suryawanshi, and S. Reshma, "Enhanced cloud security: a novel intrusion detection system using ARSO algorithm and Bi-LSTM classifier," *International Journal of Information Technology*, vol. 16, no. 6, pp. 3837–3845, 2024, doi: 10.1007/s41870-024-01887-x.

[53] A. Ye and Z. Wang, "Classical machine learning principles and methods," in *Modern Deep Learning for Tabular Data*, Berkeley, CA: Apress, 2023, pp. 3–93. doi: 10.1007/978-1-4842-8692-0_1.

[54] H. Zhang, M. Zaman, A. Jain, and S. Sampalli, "A hybrid machine learning intrusion detection system for wireless sensor networks," in *20th International Wireless Communications and Mobile Computing Conference, IWCMC 2024*, 2024, pp. 830–835. doi: 10.1109/IWCMC61514.2024.10592535.

[55] A. H. Abdi *et al.*, "Security control and data planes of SDN: a comprehensive review of traditional, AI, and MTD approaches to security solutions," *IEEE Access*, vol. 12, pp. 69941–69980, 2024, doi: 10.1109/ACCESS.2024.3393548.

[56] X. Meng, "Conceptualizing and measuring privacy boundary turbulence in technological contexts: Constructing a measurement scale," *Information Processing and Management*, vol. 61, no. 3, 2024, doi: 10.1016/j.ipm.2024.103658.

[57] H. Qiu *et al.*, "Flash: fast model adaptation in Ml-centric cloud platforms," *Proceedings of Machine Learning and Systems 6 (MLSys 2024)*, 2024, pp. 524–544.

[58] D. D. N. Nguyen *et al.*, "Design and robust evaluation of next generation node authentication approach," *IEEE Transactions on Dependable and Secure Computing*, vol. 21, no. 6, pp. 5311–5323, 2024, doi: 10.1109/TDSC.2024.3373778.

BIOGRAPHIES OF AUTHORS

Blessing Ifeoluwa Omogbehin is an M.Sc. student in the Department Computing and Informatics at Botswana International University of Science and Technology, Palapye, Botswana focusing on applying machine learning models to enhance cybersecurity. Her research centers on improving data security and access control in cloud computing environments. With a strong foundation in programming, database management, and information security, she has developed practical tools for network analysis and encryption using Python. Passionate about cybersecurity education, blessing is dedicated to advancing AI-driven solutions that address real-world security challenges. She can be contacted at email: ob24019134@studentmail.biust.ac.bw.

Tshiamo Sigwele is currently a lecturer in the Department of Computer Science and Information Systems at Botswana International University of Science and Technology (BIUST) with research interests in cloud computing, machine learning, and wireless communication and digital healthcare. He graduated in 2017 with a Ph.D. in cloud computing and telecommunications from the University of Bradford, UK. He has published a lot of internationally recognized publications with a very active research profile. He worked as a researcher from 2017 to 2018 in a British Council-funded project, BLESS U: Bandar Lampung Enhanced Smart Health Services with Smart Ubiquity, with a grant total of €89,937 and published several high-quality publications. He is currently supervising Ph.D. and M.Sc. students in the areas of cloud computing and machine learning. He is involved in several research projects at BIUST. He can be contacted at email: sigwelet@biust.ac.bw.

Thabo Semong is a senior lecturer and researcher in the Department of Computing and Informatics at Botswana International University of Science and Technology, Palapye, Botswana. His research focuses on key areas including computer networks, software-defined networking (SDN), the internet of things (IoT), network security, and UAV optimization. With a robust academic profile, he has contributed significantly to these fields, boasting over 20 high-quality publications. His work is dedicated to advancing the understanding and application of cutting-edge technologies in networked systems and security. He can be contacted at email: semongt@biust.ac.bw.

Aone Maenge is an M.Sc. student in the Department of Computer Science and Information Systems at Botswana International University of Science and Technology. His research focuses on applying various machine learning models for diabetes prediction, with a particular emphasis on utilizing random forest algorithms to enhance predictive accuracy. His work aims to improve early diagnosis and treatment strategies for diabetes. He has been actively collaborating with fellow researchers to refine his models and contribute to advancements in health informatics. He is committed to leveraging machine learning to drive impactful changes in healthcare outcomes. He can be contacted at email: ma23018971@studentmail.biust.ac.bw.

Zhivko Nedev is a senior lecturer and researcher in the Department of Computing and Informatics at Botswana International University of Science and Technology, Palapye, Botswana. He is a seasoned researcher with a strong focus on theoretical computer science, combinatorial number theory, and algorithm design. His work spans various complex topics, including balanced sets, the Magnus-Derek game, and graph theory, where he has developed algorithms for finding paths in directed planar and outerplanar graphs. He has made significant contributions to the study of balanced sets in finite fields and has proposed efficient strategies for combinatorial games, particularly in reducing computational complexity. He continues to push the boundaries of computational theory and its real-world applications. He can be contacted at email: nedevz@biust.ac.bw.

Hlomani Hlomani is a senior lecturer at the Botswana International University of Science and Technology. He received his undergraduate degree in Information Technology from the Cape Peninsula University of Technology, Cape Town, South Africa. He also received both his M.Sc. and Ph.D. degrees in Computer Science from the University of Guelph, Guelph, Ontario, Canada, in 2009 and 2014, respectively. His research interests include artificial intelligence, the semantic web, ontologies, knowledge management, and knowledge engineering. He can be contacted at email: hlomanihb@biust.ac.bw.