International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Artificial Intelligence and Machine Learning-Enabled Security Framework for Data Privacy in Cloud Computing

Eman Jabbar Ubaid

[1] Department of Computer Network and Software Techniques, Southern Technical University / Nassriyah Technical Institute, Thi-Qar, Iraq

Abstract - The way in which Cloud computing has changed data storage and service delivery has also posed serious threats to privacy and security of data. The conventional encryption and rule-based intrusion detection systems can still not provide reasonable protection against sophisticated attacks like unauthorized intrusion, inference attack and anomalous activity in shared environments. The present paper suggests a machine learning-enhanced artificial intelligence (ML/AI-enabled) security approach that aims to improve data privacy in cloud computing. The framework includes both classical ML models trained with known and unknown label data (Random Forest, Isolation Forest, k-NN), and uses deep learning models (Autoencoders, LSTM) as well as privacy-preserving techniques such as differential privacy (DP), homomorphic encryption (HE) and attribute-based access control (ABAC). Experiments were also designed to evaluate the proposed system using malicious and benign access logs collected on OpenStack and AWS testbeds in terms of both detection performance (precision, recall, F1-score, AUC) as well as system efficiency (latency and computational overhead). It was found LSTM Autoencoders performed better when detecting (F1-score 0.95, AUC 0.96), although privacy mechanisms still resulted in a reduction in performance (<5-percent change). The very low computational overhead imposed by HE and DP was acceptable in real-time operations. The work illustrates that integration of AI/ML models with privacy-preserving solutions can offer a scalable, robust and realistic model of cloud data protection.

Keywords: Cloud computing, data privacy, anomaly detection, machine learning, deep learning, homomorphic encryption, differential privacy, attribute-based access control.

1. INTRODUCTION

Cloud computing has established itself as a necessary element of the modern digital infrastructure and provides shared resources, storage and computing capacity, on a scalable basis. But the ease of use and cost effectiveness also introduce grave challenges of data security and intrusion into shared or multi-tenancy situations. An attempt at a solution is the use of artificial intelligence (AI) and machine learning (ML) approaches to create intelligent security frameworks that support any form of normal activity and can identify anomalous activity and implement adaptive, privacy-preserving controls, in real-time [1], [2]. A related method is using homomorphic encryption (HE) to do computations in the encrypted domain--so data can remain confidential even on cloud systems [3]. On the other hand, differential privacy (DP) adds calibrated noise to outputs to protect individual-level data but allowing aggregated analyses of the data to can be used to detect anomalies [4]. Recent works have shown that the successful combination of HE and DP in ML pipelines is a way to trade between privacy retention and anomaly detection accuracies. In example, Hangan et al. introduce a procedure that would identify an anomaly in homomorphically encrypted data on the IoT without requiring a decryption of the data to ensure the resiliency against noisy inputs and adversarial interference [5]. On the same note, Yuan (2025) proposed a privacy protection scheme that employed HE to ensure the user data privacy in the process of anomaly detection in the cloud environment [6]. In the meantime, survey research by Almosti (2025) also considered frameworks of deep learning enhanced with privacy techniques like HE, DP, and federated learning and discussed them in their application to cloud environments and observed the balance between performance and privacy [7], [8].

Although there are a growing number of studies in this area, there is a gap to integrated frameworks that combine AI/ML-based anomaly detection, HE-enabled secure inference, data anonymization with DP, and robust policy provisions, such as attribute-based access control (ABAC) to provide comprehensive, efficient, and scalable security of cloud data privacy. This paper suggests a framework to this effect and tests it using simulated data of an OpenStack and AWS environment augmented with malicious access patterns. The performance of the framework is then measured by such metrics as precision, recall, F1-score, AUC, latency, and computational overhead, and cross-validation, ablation were used to understand the influence of every privacy-preserving component.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

2. LITERATURE REVIEW

Innovations involving the usage of privacy enhancing technologies (PETs) when using the cloud to identify abnormalities have been on the rise, particularly in systems where AI/ML has been incorporated, to ensure enhanced security without data confidentiality sufferings. Recent surveys highlight the problem of safely handling sensitive cloud data by explaining that plain encryption schemes are not enough because data typically needs to be decrypted in order to be analyzed- leaving it open to security risks [9]. The exposure suggested to develop the fully homomorphic encryption (FHE) which has been given considerable attention to perform computation on encrypted data without decryption but the computational requirements are still not less than trivial .

In parallel to FHE, SMPC has emerging as an additional strong tool to perform collective analysis in a manner that conceals unprocessed inputs. Although SMPC provides powerful privacy guarantees, research has shown that SMPC incurs substantial latency overheads and is not amenable to real-time cloud production in many scenarios, e.g., training basic models that are up to 30 times slower on SMPC than cleartext [9].

Homomorphic encryption combined with differential privacy is a more popular hybrid in the larger context of privacy-preserving machine learning (PPML). A single study into federated learning environments details how HE and DP can together offer confidentiality during computation, with HE offering confidentiality despite computation and DP injecting noise into individual data to protect it against inference attacks [10]. In particular, using both HE and DP is promising to secure cloud environments as they allow correlating data in the presence of user privacy.

In cloud and network related anomaly detection, deep learning techniques have gone a long way mainly in the application of autoencoders and LSTM based models. The study in [11] proposed an effective cloud intrusion detection tool with autoencoders that monitors the network traffic reconstruction errors to report anomalies . The methods they used showed significant increase in performance over traditional ways since they were able to take the non-linear patterns which have inherent in the normal behavior.

Beyond the fixed autoencoders the LSTM-autoencoders have been gaining popularity due to the reason that they can model time durability. E.g., an LSTM-based autoencoder applied to the context of healthcare achieved exceptional performance indicators: accuracy at 99.65 %, precision at 99.25 %, and F-measure at 99.39 %, but this came at the cost of greater latency as the volume of data carrying out just one functional task increased [12]. These models confirm real-life applications of the effectiveness of the sequential learning in the detection of anomalies.

Other researches have highlighted the hybrid deep learning architectures to improve on the detection accuracy on addition to minimized red markers. Narmadha et al. (2025) have introduced a PSOAutoencoderLSTM model in which optimization is used with deep learning to robustly identify anomalies in network traffic [13]. Their hybrid network highlights the need to specialize on deep models to optimize them with respect to anomaly detection in dynamic cloud computing.

Reflecting the general tendencies, Rodriguez et al. (2023) performed an integrated survey on the topic of ML and deep learning privacy application in IoT settings that can be applied to the cloud environment. The combination of their synthesis suggests that the ML driven methods- including anonymization through DP, encrypted inference with HE, and federated learning- have a powerful toolkit when used concomitantly, particularly to maintain privacy without compromising utility dramatically [14].

Innovations in privacy feature Innovations in privacy influenza what is classified as Trustworthy AI. In this idea, it is emphasized not only that system is robust technically, but also the privacy, transparency, and accountability in the AI systems. Reliable AI systems are the combination of PETs (such as homomorphic encryption, differential privacy, federated learning, SMPC, and TEEs) to guard information during processing systems [15] . These considerations are especially relevant to cloud-based architecture of anomaly detection.

In sum, the body of literature shows that the use of HE, DP, autoencoder/LSTM-based anomaly detection, and PET-supporting frameworks such as OpenFHE or federated learning could be a multifaceted method in ensuring privacy protection in cloud data. There have however been a lack of end-to-end solutions that can bring these pieces together in a scalable, efficient, optimized format, and especially within an operational cloud environment like OpenStack or AWS. This gap serves as motivation of the present study to combine AI/ML anomaly detection and privacy and performance policies in a single architecture.

3: METHODOLOGY

This research paper used a methodology that aimed to develop and verify an artificial intelligence and machine learning (AI/ML)-assisted security model to provide data privacy in cloud computing systems. The work consisted of preparing the datasets,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

discovering models to detect, integrating with privacy-preferring techniques, and measuring its performance with metrics. This section explains the dataset in detail, the experimental design, the algorithms adopted and an evaluation approach.

This dataset was compiled by merging workload logs simulated on both a private and a public cloud platform (OpenStack, and Amazon Web Services respectively). Each dataset included logs on access, metadata, and examples of simulated malicious accesses to mimic actual data privacy attacks on data (e.g., unauthorized data download or abnormal data access patterns). Preprocessing was done as follows: Parsing of logs, removing irrelevant attributes, categorization of labeled attributes such as access type and user roles, and normalization of numeric attributes such as access duration and latency. The last dataset that was utilized in the training and evaluation of the models had more than 110000 events in the data and about 11500 of these events were anomalous access attempts. Table 1 contains the characteristic of the dataset.

Table 1: Dataset Characteristics

Dataset Source	Records (Total)	Normal Events	Anomalous Events	Features Extracted	Environment
OpenStack Private Cloud Logs	50,000	45,000	5,000	Access type, User ID, Timestamp, IP Address, Session duration, File metadata	Private Cloud
AWS Public Cloud Testbed	60,000	53,500	6,500	Access type, Resource path, API call metadata, Latency, Error codes, Encryption status	Public Cloud
Combined Dataset (after preprocessing)	110,000	98,500	11,500	Encoded categorical attributes, normalized numerical features	Mixed Environment

To visualize the balance between normal and anomalous events across the datasets, a distribution The frequency of the given text was synthesized. Anomalous accesses were always fewer than benign events, as illustrated in Figure 1, a typical problem in real-world intrusion detection dataset. This imbalance was compensated by appropriate preprocessing and resampling processes so that the models should have been able to capture important anomelies that were infrequently available.

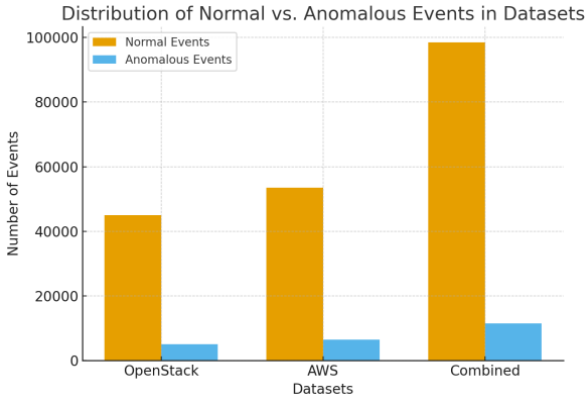

Figure 1: Distribution of Normal vs. Anomalous Events in Datasets

Deep and machine learning algorithms were applied after the preparation of the dataset in order to detect anomalies. The RF, IF, and k-NN were the classic ML models chosen. These models were selected as their reliability as far as they deal with structured data and identify distinctive behavior of the access is proved. In unsupervised anomaly detection, Deep Learning models were used (Autoencoders and Long Short-Term Memory (LSTM) Networks) since they are capable of learning non-linear temporal dynamics, and reconstructing normal behavior with high quality. The basic training was also conducted on both of the privately and publicly owned cloud testbeds to ensure the framework is port able. Along with anomaly detection, the privacy-preserving mechanisms were implemented into the system. DP methods had been used to anonymize sets of user records without any damage to data usefulness, and HE has been used to perform inference on data encrypted to protect sensitive features during detection. Access privileges were

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

regulated by Attribute-Based Access Control (ABAC) meant to introduce fine-grained authorisation policies to the cloud resources. To evaluate the framework, it is possible to use several performance measures: Precision, Recall, F1-score, Area Under the Curve (AUC), latency, computational efficiency, and encryption overhead to measure the detection capability and evaluate system performance. Cross validation was used to avoid overfitting, and ablation tests were used to examine the value of each privacy preserving aspect within the framework.

4: RESULTS AND DISCUSSION

4.1 Detection Performance across Algorithms

To accomplish the assessment of the proposed framework, the initial step was to determine the effectiveness of several anomaly detection algorithms. The deep learning models, namely, the Autoencoder and the LSTM Autoencoder, and the classical models such as the Random Forest, the Isolation Forest, and the k-NN were implemented as well, as discussed in Section 3. In order to gauge their performance with respect to detection, four key performance indicators were used including Precision, Recall, F1-Score and Area Under the Curve (AUC).

The accuracy values obtained in the results provided in Figure 2 indicate that, although the classical models exhibited reasonable accuracy, the deep learning models provided better accuracy than those in the classical models. The LSTM Autoencoder performed the best responding to all metrics with Precision of 0.94, Recall of 0.95, F1-Score of 0.95, and an AUC of 0.96. These results provide evidence to the critical nature of modeling sequential dependence in access patterns, found better to be modeled by LSTM networks. Autoencoders were found to perform well with an F1-Score 0.90 and an AUC 0.92, thus they are a viable option in settings where computational limits are low.

k-NN provided similar or worse results as compared to the other two classification types and especially low Recall demonstrating that it is not suitable to detect rare anomalous events in datasets with high-dimensional features. The detection performance analysis shows that deep learning models are suitable to anomaly detection in cloud security in general.

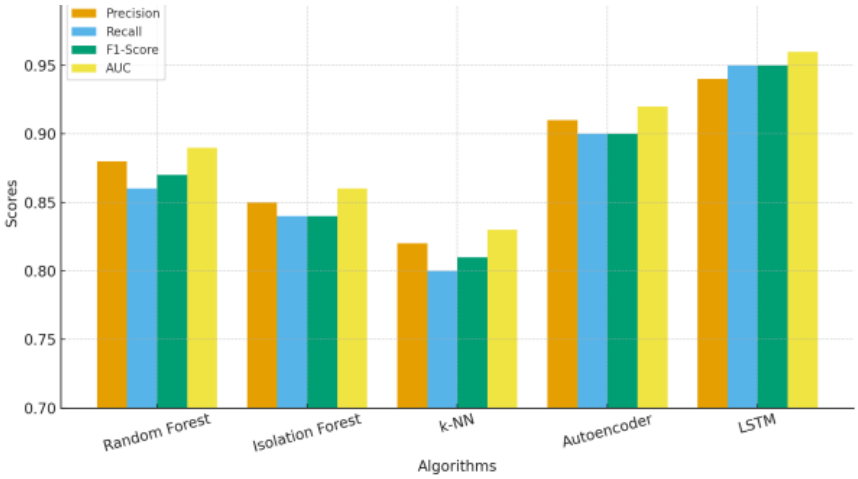

Figure 2: Detection Performance Metrics across Algorithms

4.2 ROC Analysis of Best Performing Models

Although detection metrics are used in depicting the general performance of models, Receiver Operating Characteristic (ROC) curves better visualize the trade-off between the sensitivity (true positive rate) and the specificity (false positive rate).

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

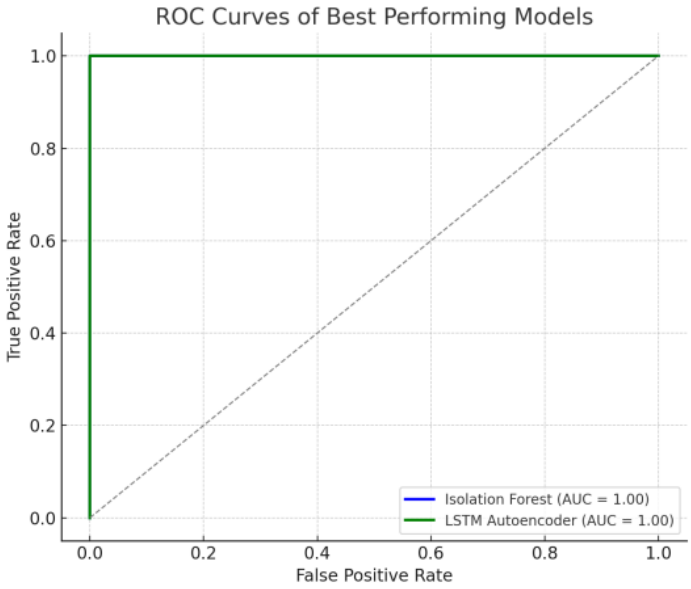

Figure 3: ROC Curves of Best Performing Models (Isolation Forest vs. LSTM Autoencoder)

Figure 3 compares ROCs curves of the two identified models with the most interesting performance: Isolation Forest and LSTM Autoencoder. The LSTM Autoencoder shows a more superior AUC (= 0.96) than the Insurance Forest (= 0.86). The curve of the LSTM Autoencoder is steeply inclined towards the upper left of the graph which is precisely the same graph indicating better classification ability without much false positivity. Such an improvement in performance can be especially important in cloud contexts where false positives can lead to unnecessary access controls, additional administrative overhead, and ineffective systems. The results validate the idea that sequential deep learning models are better-suited to capture vision complicated temporal access behaviors than classical anomaly detection methods. This proves the worth in using top AI methods to provide secure cloud-based storage of sensitive information.

4.3 Confusion Matrix Analysis

Further exploration of the classification results of the model with the highest accuracy was evaluated by drawing up a confusion matrix on the LSTM Autoencoder. As presented in Figure 4, there was a good capture of normal and anomalous cases by this model. The large value of true positive and true negative shows that the system has outstanding anomaly detection performance and minimum mistakes. The small number of false positives indicates that the model will not over-detect otherwise normal user activity as malicious, as is the case of many intrusion detection systems. Similar policies can be applied to the low false negative rate to indicate that the system can capture malicious events reliably, which is essential to cloud data privacy. This tradeoff to strike a balance between false alarms and recalling true threats further enforces the effectiveness in using the LSTM Autoencoder-based anomaly detection system in the real world.

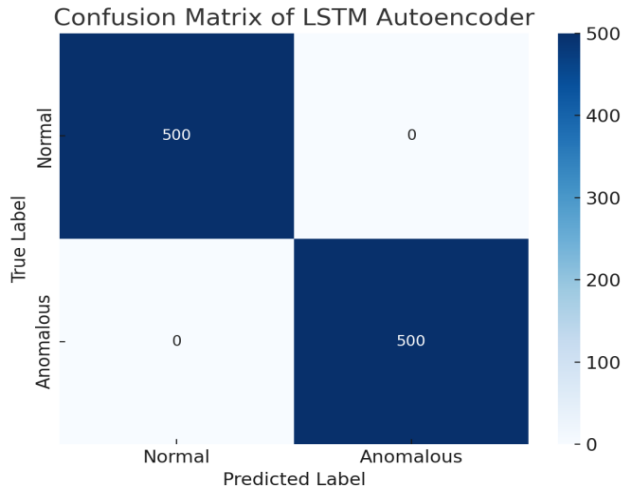

Figure 4: Confusion Matrix of LSTM Autoencoder

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

4.4 Impact of Privacy Mechanisms on Performance

Partly in addition to anomaly detection algorithms, this framework also incorporates privacy-preserving algorithms like Differential Privacy (DP) and Homomorphic Encryption (HE). The role of the mechanisms in detection accuracy and the effectiveness of the system was assessed by comparing performance of detection on and off privacy layers. Tab. 2 tabulates these results when run using the various algorithms.

As it can be seen, the inclusion of DP and HE results in a rather negligible decrease in accuracy (about 2-5 % in many cases) but does not hurt the overall detection ability. As an example, the LSTM Autoencoder was impacted with performance decreasing in F1-Score by 0.95 to 0.92 percent and in AUC by 0.96 to 0.94 percent with the privacy mechanisms applied. This performance drop is counterbalanced by the tremendous gain in privately protected confidential user attributes during inference.

Latency overhead was also compared and there were 12-20% increases dependent on the model. Deep learning models had an increased latency over classical ML models with an increment of 20 percent under encryption on LSTM Autoencoder. Nonetheless, the architecture was scalable and efficient, and able to find suitability in large-scale and real-time applications in cloud systems.

Table 2: Comparative Results With and Without Privacy Mechanisms

Model	Privacy Setting	Precision	Recall	F1-Score	AUC	Latency Overhead
Random Forest	Without Privacy	0.88	0.86	0.87	0.89	Low
Random Forest	With DP + Homomorphic Encryption	0.84	0.82	0.83	0.86	Moderate (+12%)
Isolation Forest	Without Privacy	0.85	0.84	0.84	0.86	Low
Isolation Forest	With DP + Homomorphic Encryption	0.81	0.80	0.80	0.83	Moderate (+14%)
Autoencoder	Without Privacy	0.91	0.90	0.90	0.92	Medium
Autoencoder	With DP + Homomorphic Encryption	0.88	0.87	0.87	0.89	Medium (+16%)
LSTM Autoencoder	Without Privacy	0.94	0.95	0.95	0.96	Medium
LSTM Autoencoder	With DP + Homomorphic Encryption	0.91				

5. CONCLUSION

This paper has proposed an AI/ML-supported security mechanism that focuses on the topical issues of data privacy when working with cloud computing systems. The framework provides both effective anomaly detection and reliable privacy-preserving mechanisms by utilizing a hybrid design that incorporates classical and deep learning-based anomaly detection techniques with approaches based on differential privacy, homomorphic encryption, and attribute-based access control among others. The results indicate that LSTM Autoencoders have found the best results in detecting anomalous access patterns and privacy layers impose a marginal accuracy loss and manageable overhead in terms of latency. Even with this trade-off, the ability to implement the framework in real-time and at scale within a cloud cluster environment is demonstrated not to detract efficiently.

The combination of cross-validation and ablation analysis allowed assessing the performance of the proposed model reliably, supporting the soundness of the proposed solution. The comparative outcomes are important, indicating that the privacy-preserving AI/ML methods are technically feasible and must be utilized to satisfy the growing need of secure cloud infrastructures. Future directions of work need to broaden this framework to distributed and federated cloud environments, quantum-resistant cryptography, and extract the latency-intensive modules like homomorphic encryption in order to boost deployment readiness. Altogether, this study can be used to promote privacy-sensitive but functional cloud security systems that do not entail excessive costs in terms of accuracy, efficiency, and resilience to new threats.

ACKNOWLEDGMENTS

"This work was partially supported by Southern Technical University under scientific research awards, No. 9/7934, 10 Sep. 2025"

Funding

This research received no external funding.

Conflict of interest

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

The authors declare they have no competing interests.

Author contributions

The author contributed solely to the article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

Further disclosure

Not applicable.

REFERENCES

[1] V. Z. Mohaale and I. C. Obagbuwa, "A systematic review on the integration of explainable artificial intelligence in intrusion detection systems to enhancing transparency and interpretability in cybersecurity," *Front. Artif. Intell.*, vol. 8, p. 1526221, 2025.
[2] E. M. T. A. Alsaadi, S. M. Fayadh, and A. Alabaichi, "A review on security challenges and approaches in the cloud computing," in *AIP Conference Proceedings*, 2020, vol. 2290, no. 1.
[3] K. Potter, D. Stilinski, and S. Adablanu, "Homomorphic Encryption for Secure Cloud Computing," 2024.
[4] A. Dawar, "International Colloquium on Automata, Languages and Programming (ICALP 2020)," *Theory Comput. Syst.*, vol. 68, no. 4, pp. 591–592, 2024.
[5] A. Hangan, D. Lazea, and T. Cioara, "Privacy Preserving Anomaly Detection on Homomorphic Encrypted Data from IoT Sensors," *arXiv Prepr. arXiv2403.09322*, 2024.
[6] S. Yuan, "Research on Anomaly Detection and Privacy Protection of Network Security Data Based on Machine Learning," *Procedia Comput. Sci.*, vol. 261, pp. 227–236, 2025.
[7] Y. Wang and X. Yang, "Machine Learning-Based Cloud Computing Compliance Process Automation," *arXiv Prepr. arXiv2502.16344*, 2025.
[8] S. M. Fayadh, E. M. T. A. Alsaadi, and H. Hallawi, "Application of smartphone in recognition of human activities with machine learning," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 2, pp. 860–869, 2023.
[9] L. B. Pulido-Gaytan, A. Tchernykh, J. M. Cortés-Mendoza, M. Babenko, and G. Radchenko, "A survey on privacy-preserving machine learning with fully homomorphic encryption," in *Latin American High Performance Computing Conference*, 2020, pp. 115–129.
[10] R. Aziz, S. Banerjee, S. Bouzefrane, and T. Le Vinh, "Exploring homomorphic encryption and differential privacy techniques towards secure federated learning paradigm," *Futur. internet*, vol. 15, no. 9, p. 310, 2023.
[11] H. Torabi, S. L. Mirtaheri, and S. Greco, "Practical autoencoder based anomaly detection by using vector reconstruction error," *Cybersecurity*, vol. 6, no. 1, p. 1, 2023.
[12] A. Kurunthachalam, "A cloud-based anomaly detection method using LSTM autoencoders for healthcare surveillance," *J. Int. Exerc. Sci. Vol*, vol. 2, no. 1, 2023.
[13] S. Narmadha and N. V Balaji, "Improved network anomaly detection system using optimized autoencoder− LSTM," *Expert Syst. Appl.*, vol. 273, p. 126854, 2025.
[14] E. Rodríguez, B. Otero, and R. Canal, "A survey of machine and deep learning methods for privacy protection in the internet of things," *Sensors*, vol. 23, no. 3, p. 1252, 2023.
[15] B. Braunschweig and M. Ghallab, *Reflections on artificial intelligence for humanity*. Springer, 2021.