LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 22, NO. S4(2024)

MACHINE LEARNING APPROACHES AND CHALLENGES IN ANOMALY DETECTION SYSTEMS WITHIN CLOUD ENVIRONMENT: A REVIEW

Manisha Milind Patil[1] , Dr. Dhanraj Tambuskar[2]

[1]Research Scholar, Sri Balaji University,Pune,
[2]Associate Professor , Sri Balaji University,Pune

Abstract: Data sharing, Consumption based pricing, Definite services, Fast adaptability ,joint resource access are some of the key areas of Cloud computing popularity. A abundant users can use the stashed framework of cloud and more number of transactions are generated. Due to this cloud security has significant importance with the help of cloud security users can design policies and control accesses to protect the cloud environment from different kinds of threats. A data which is generated during transaction which shows abnormal behaviour, pattern or operations is called as anomaly. The anomalous data can create the troubles like security threats, system malfunction, Efficiency decrease can be identified in the different ways like unauthorized access, network flow pattern, and unpredicted system performance, inappropriate use of resources or anomalous system activity. Finding anomalous activity is crucial because of the vast amount of data that is shared in cloud environments during different transactions. Strengthen security, offer protection against valuable infrastructure, maintain uninterrupted system monitoring, and improve cloud anomaly detection performance in order to guarantee seamless cloud operation. By employing anomaly detection methods, cloud users and providers can bolster the security of cloud infrastructures, thereby reducing the risk of vulnerabilities and attacks. .Three important methodological domains are highlighted by our analysis: statistical techniques, deep learning, and machine learning. and outlines how each model is specifically applied for anomaly detection. Additionally, We outline the typical application areas of anomaly detection within cloud computing systems, as well as the public datasets frequently utilized for assessment. Finally we. make recommendations for future study directions and talk about the consequences of our findings.

Keywords : Cloud Security,Deep learning ,Anomaly detection ,Machine learning, Cloud Computing

I. Introduction

Cloud technologies enable cloud service models to easily access a shared network, storage, and resources as needed. The three approaches that are utilized in private, public, and hybrid cloud deployment strategies are IaaS,PaaS,SaaS [8]. Essentially, cloud platform offers online access to servers, storage, databases, and a variety of application services. It increases cost-effectiveness, scalability, and flexibility for both people and enterprises.Instead of having the hardware or infrastructure on location, it is comparable to remotely accessing resources and services via the Internet.. 1.1 Benefits Cloud computing provides many benefits and considerable advantages, including: Cost-effectiveness: Cloud computing eliminates the need to invest in and maintain physical gear, resulting in lower capital and operational expenditures.

Dynamic: Depending on requirement, users can acquire resources and release the resources.Cloud service providers allow to customized the requirements to fulfill the service

Accessibility: Users have the ability to access files, applications, and systems from any device connected to the internet. This capability removes the requirement to remain in a fixed office space or rely on specific hardware.

Reliability: Cloud service providers typically provide comprehensive saving and restoration options, giving the assurance of data availability in the event of natural disasters or hardware malfunctions.

Security: Cloud providers put a lot of effort into security measures to keep data safe from unofficial access. This ensures that data is well-protected and that they meet industry standards.
Innovation: Cloud computing makes it easier to quickly launch new apps and services, encouraging innovation and helping businesses stay competitive in the fast-paced digital world.
Cloud infrastructure operates using common Internet protocols and relies on virtualization methods, which can make it susceptible to security threats.Cloud environments are exposed to various types of attacks like DoS, SQL Injection[24],XSS attacks , others[13] . These threats may also originate from traditional attack vectors, including Address Resolution Protocol (ARP) poisoning, IP spoofing, and DoS attacks. [29].Ultimately, cloud computing is important because it may help both consumers and businesses save money by streamlining processes, increasing productivity, and spurring innovation. Due to the following factors, cloud computing security is crucial[13][29].
Data protection: To avoid unwanted access, leaks, or breaches, it is essential to guard personal and organizational data saved on cloud.
Compliance: Strict laws protecting privacy and data are in place in many businesses. Cloud computing security features assist businesses in adhering to these laws.
Business Continuity: With security measures in place, data may be restored in the case of a cyber attack or data breach, reducing downtime and guaranteeing business operations continue.
Risk management: By assisting in the detection of possible threats, weaknesses, and dangers, cloud security solutions enable businesses to proactively manage and reduce these risks.
Trust and Reputation: Upholding strict security procedures fosters trust among clients and associates, safeguarding the credibility and reputation of the company.
Shared Responsibilities: Understanding the idea of shared responsibility in cloud computing is important. Companies need to protect their data and applications, while cloud providers take care of securing the infrastructure.

II. Types of attacks

Within cloud environments, security can be threatened by a range of risks, such as malicious insiders, phishing schemes, fraudulent activities, signature wrapping attacks, denial-of-service (DoS) incidents, insecure interfaces, and other emerging vulnerabilities [12].Finding deviations or uncertainties in data is known as anomaly detection. Time and resources can be saved by taking this action early on[16]

III. Literature Review

Abualhaj et al.,(2023)The research paper explores a tailored KNN was developed for enhancing Malicious detection by modifying the distance metric parameter. The main focus is on enhancing the fulfilment of the KNN algorithm by choosing the appropriate measure of proximity. for binary and multiclass classification. The study compares three distance metrics: Euclidean, Manhattan, and Minkowski distances[1].
Nurudeen M.I et al., (2019) in this paper the author has discussed the Intrusion Detection algorithm which is adaptable and approach is divided into four parts viz feature selection component. This section uses ant colony optimization in collaboration with correlation based methods for feature selection. The second part is the change point detection component, which applies binary segmentation to identify points where data patterns shift in time series data. The third component handles training and updating the classifier, specifically the SGD-SVM model,

in an adaptive way. The final part is the detection component, which looks for signs of attacks in new data samples based on the test data. [2].

T. Sreenivasulareddy et al.,(2022) discussed machine learning approaches for Detection of unauthorized access in networked systems. The main objective is to upgrade the accuracy of detecting and preventing cyber attacks compared to traditional ML approaches [7].

Promise R. Agbedanu1 et al.,(2022) an efficient IDS model for IoT that provides robust security while addressing the limitations of constrained devices. The incremental ensemble approach balances performance and resource efficiency, marking an advancement in cyber security solutions for IoT ecosystems. While the model achieves an optimal balance of accuracy, speed, and resource usage, further research is suggested to enhance performance across various IoT datasets and explore handling concept drift (updation in data pattern over time)[7].HanaaAttou et al.,(2023)discussed the anomaly identification in cloud infrastructures employing time series machine learning methods. Experiments demonstrate that A-CIDS improves detection accuracy for port scanning and DDoS attacks, reaching a high success rate in detection while keeping false positives to a minimum both before and during VM migrations. Compared with traditional and existing adaptive systems, A-CIDS shows superior adaptability and robustness in dynamic cloud environments. This paper examines anomaly detection in cloud networks, with particular emphasis on the security challenges introduced by large-scale and dynamic data environments.. It evaluates two ways first is One-Class Support Vector Machine and Autoencoders, emphasizing their ability to identify anomalous or outlier data without requiring labeled datasets[5] .

Ibrahim N.M. et al.,(2019) discussed an adaptable approach to detecting intrusions within cloud computing environments for addressing challenges in anomaly detection within cloud environments, especially during virtual machine (VM) migrations. VM migrations, essential for dynamic resource management, can disrupt traditional IDS as they cause frequent changes in network behaviour, potentially leading to false positives[10].M. M. Belal et al.,(2022)looked at the different techniques used to detect intrusions and the challenges involved in maintaining security in cloud environments. They also talked about a classification of cloud threats and security concerns organized by the layers defined by the Cloud Security Alliance (CSA), pointing out how conventional security approaches often fall short. The study provides an analysis of case studies that leverage machine learning and deep learning approaches, focusing on attack types, evaluation methods, datasets, and measures for effective attack prevention. The author also discussed that Distributed Denial of Service attack, Unauthorized access has been identified by Support vector machines. Fuzzy C-means and convolutional networks, recurrent networks. This algorithm provides high accuracy rate and lower false alarm rate.[13]..Xiangyu Liu1 et al.,(2022) discussed anomaly detection method by using estimation method using MemAe-gmm-ma which is the unsupervised method in the cloud environment.The different density estimation methods are also described like deep autoencoder combined with a Gaussian Mixture Model.The main objective of memory module within autoencode is to handle co Association among intrinsic data and generalization, allowing it to detect specific anomalies.Mem.Performance measure of Ae-gmm-ma enhanced accuracy by Perform dimensionality reduction on high-dimensional datasets and anomalies can be detected from values of reconstruction error Experiments conducted on the variety datasets demonstrate notable improvements in F1 scores compared to existing models, achieving a 4.47% increase on NSL-KDD and a 9.77% increase on CIC-IDS-2017. These results highlight the resilience of the suggested method, even in the presence of contaminated data, while

advancing unsupervised anomaly detection in cloud environments by improving sensitivity to subtle anomalies and reducing over fitting.[32]
Andrew M et al.[21](2020) has analyzed the application of CNN for malware detection. The author discussed various CNN architectures by using the DARPA dataset. The findings indicated that as the inten of the network layers increased, the false positive rate also rose, while the overall accuracy demonstrated an improvement.Chirbene et al.[14](2020) discussed the model for discovery of intrusion by using machine learning algorithm which is Utilizing the UNSW-NB15 dataset, this analysis is grounded in decision tree and random forest methodologies, employing five distinct features. and nine different attacks.The proposed model uses windows past information for selection of different features.

IV. Evaluation of Anomaly Detection Methods in Cloud

This section discusses recent research on identifying anomalies in cloud networks, employing different Practices involving machine learning, deep learning, and statistical analysis. An assessment of the articles concerning the methodologies applied and performance metrics like recall, precision, and accuracy is provided.3.1 Approaches used for Anomaly identification.

A Approach leveraging machine learning technologies

Machine learning encompasses all techniques that do not rely on predetermined rules but instead learn the patterns and relationships within data to generate guidelines. In the following, we refer to classical machine learning methods as those that do not depend on layer based architecture.Here some notable work done by researchers,represented by Fig.1.

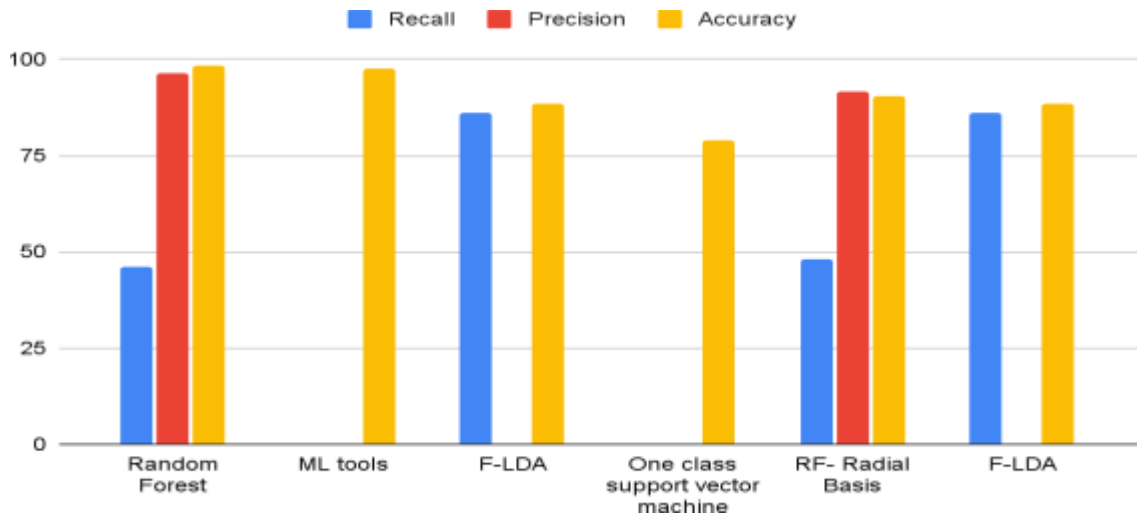


Fig 1. Comparison of performance measures of different models of Machine learning

Hanaattou et al.[8] used the Random Forest algorithm and support vector machines algorith in cloud environment used for discovery of anomalies.They also reduced the number of features and assumed the threshold value for features like duration variable >1500 ,src bytes variable > 0.,dst bytes is higher than 50 000 or detection of anomalies.Ranjithkumar et al [9],have discussed the anomalies in the network traffic.As LDA is one of the famous statistical models which uses

unstructured and unlabelled data.Fuzzy clustering is used with LDA to It eliminates any unnecessary words or corpora produced by the network.Here, the fuzzy model of the classifier output is created using information entropy.Yasarathna, T. L.[11] have discussed the SVM in OCSVM is designed to train data using a single class, which is thought to be the most discriminative class.It demonstrates a OCSVM overall accuracy of 79.17%,60.69% on various dataset Parameswarappa et al. in [18] presented an innovative firewall technique to improve secure cloud-based computing using machine learning tools. By combining past node assessments with the machine learning algorithm's current decision to forecast the ultimate classification of attacks. Using the UNSW-NB-15 dataset, the recommended method was assessed and obtained an F1-Score and accuracy. But the cost of computation has been enormous.Al-Ghuwairiet. al[13] have proposed model for intrusion detection by using time series model.The model outlined in this proposal described by time series quantization (TSQ) and utilization of the Granger causality test(GCT) .Through TSQ, The information is altered into a_collection of distinct values. This process is important for GCT which is useful to identify the association among time series data. The author additionally indicated that the model's performance might be optimized through a reduction in the quantity of input predictors, which would Contribute to minimizing time required for training Table 1 presents the strengths, limitations, and methods employed in the research study.

Ref.	Year	Methods	Strengths	Limitations	Datasets
[8]	2023	RF, DT, and SVM	reducing the number of explanatory variables reduces data collection time and execution time.	the potential of using a small number of features by contrasting the results with those of other classifiers. But recall is still not well enough using NSL-KDD,	KDD
[18]	2023	ML tools	Low-rate false alarms.	High computing cost. Requires parameters tuning	UNSW-NB-15 dataset
[9]	2022	F-LDA	It considers unlabelled data for detection of network anomalies .	Deals with network traffic flow only	KDD
[11]	2022	One Class Support Vector Machine(OCSVM)	support vector model is focused to train data that has only one class,	Class imbalance problem	YAHOO UNSW-NB15
[13]	2023	Time series analysis	The proposed method improved the forecast accuracy by reducing the number of input predictors and saving training time and resources.	Feature selection process has not given importance	CSE-CIC-IDS2018)

Table 1: A Summary of Approaches Based on Machine Learning

B Deep learning based Approach

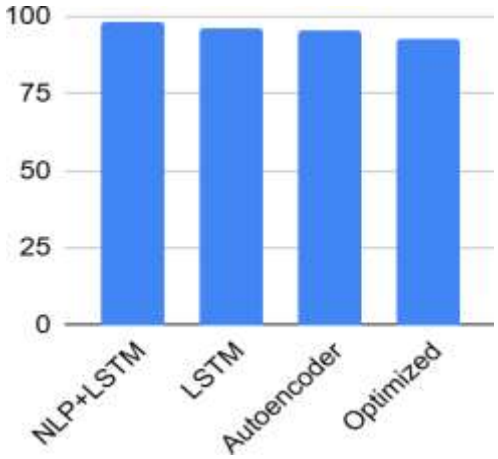
A segment of machine learning identified as "deep learning" makes use of neural networks that have multiple (≥ 2) hidden layers, which distinguishes it from classic methods.. The difficulty of obtaining pertinent high-level, abstract features from raw data is a major obstacle in many real-world applications.Researchers' contributions to the development of deep learning algorithms for

anomaly detection AutoEncoder, LSTM, NLP, are included in this section.Zhu et al.[17] discussed The LogNL strategy aims to find irregularities in the logs of cloud platforms. Long-term short memory is one of the component and other component used is Naatural language processing. The feature vector, which is formed via natural language processing, contains semantic information about the login.LSTM can identify the abnormalities by utilizing the feature vector's information. The LSTM method produces a 97.7% F1 Score and 97.7% accuracy on the HDFS and OpenStack datasets, but it needs some parameter tweaking.LSTM and a recurrent neural network are useful tools for detecting DDoS attacks.[19].The LSTM, represents a distinct category of RNN that effectively tackles the challenges of disappearing and Blasting gradients, which helps to optimize training duration and boost accuracy. This system was subjected to training and assessment on the CICIDS2017 dataset, which featured a broad spectrum of attack types. The research emphasized four critical domains: data acquistion, feature engineering, model training and the development of the classification model.The outlined training model comprises a solitary LSTM input layer with 77 neurons and a tanh activation function. Twelve hidden layers are therefore present, comprising five LSTM layers and seven fully Kunag Y et al.[20] talked about the deep learning model with Autoencoder, which uses grid search and random search algorithms to manage the automatic hyper parameter optimization process. The suggested model was evaluated using the CICIDS 2018 and NSL_KDD datasets, yielding detection rates of 95.79% and 83.33%, respectively. Interestingly, The output layer features two neurons that are entirely connected and use sigmoid activation functions.

This model is structured with 14 layers of recurrent neural networks. have a 96.25% accuracy rate. T. Thilagam et al.[22] have created a bespoke RC-NN network model that is optimized for intrusion and anomaly detection.For feature optimization, they have employed the ALO method, and for improved performance, they have employed the nlstm and bilstmlayers.

Table II:Deep Learning Approach for anomaly detection

Ref.	Year	Methods	Strengths	Limitations	Datasets
[17]	2022	NLP+LSTM	Detect various anomalies. Low-rate false alarms.	Requires parameter tuning. Not dependent on both physical and virtual features.	HDFS and OpenStack
[19]	2020	LSTM	Reduces the time required to predict the attack	Detects only DDos Attack	CICIDS2017
[20]	2021	AutoEncoder	Automatic HPO techniques help determine the most appropriate combination of hyperparameters to obtain the best performance	Struggles with complex and imbalance dataset	CICIDS2017
[22]	2020	Optimized custom RC-NN	Lower the false alarm rate	The position and orientation of objects are not encoded.	darpa,CSE-CIC-IDS2018



V. Problems and Important Standards for Upcoming Cloud Anomaly Detection Solutions

- For both researchers and network operators, identifying irregularities in cloud networks is a major difficulty.This is attributed to various factors, such as the high congestion of data on network, flow of uniterrupted data streams, the necessity for quick identification, and the occurrence of complex assult.
- Outlined below are several challenges faced by approaches to anomaly detection:
- It is crucial to improve both prediction accuracy and the ability to generalize across different datasets . [9]
- Basic traditional machine learning algorithms are inadequate for the accurate detection of intrusion attacks. Furthermore, deep learning methods do not always deliver superior results. In this context, a combined machine learning approach can provide better detection rates and enhance model development. [7]
- By dimensionality reduction researchers can improve the Efficacy of the model corresponding with respect to prediction accuracy and the incidence of false alarms .[15]
- There is a need for a model which can detect the newly identified anomalies in the cloud environment. [8]
- Future research may concentrate on investigating novel characteristics of network traffic, implementing flexible resilient identification of anomalies and leveraging approaches of machine learning techniques to improve the accuracy and efficiency of the model.[23]

VI. References

1. Abualhaj, M. M., Abu-Shareha, A. A., Shambour, Q. Y., Alsaaidah, A., Al-Khatib, S. N., & Anbar, M. (2023). Customized K-nearest neighbors' algorithm for malware detection. International Journal of Data and Network Science, 8(1), 431–438. https://doi.org/10.5267/j.ijdns.2023.9.012

2. Ibrahim, N. M., & Zainal, A. (2019). An Adaptive Intrusion Detection Scheme for Cloud Computing. International Journal of Swarm Intelligence Research, 10(4), 53–70. https://doi.org/10.4018/ijsir.2019100104
3. T, S. R., & R, S. (2022). Ensemble Machine Learning Techniques for Attack Prediction in NIDS Environment. Iraqi Journal for Computer Science and Mathematics, 78–82. https://doi.org/10.52866/ijcsm.2022.02.01.008
4. Agbedanu, P. R., Musabe, R., Rwigema, J., &Gatare, I. (2022). Using Incremental Ensemble Learning Techniques to Design Portable Intrusion Detection for Computationally Constraint Systems. International Journal of Advanced Computer Science and Applications, 13(11). https://doi.org/10.14569/ijacsa.2022.0131104
5. Al-Ghuwairi, A., Sharrab, Y., Al-Fraihat, D., AlElaimat, M., Alsarhan, A., &Algarni, A. (2023). Intrusion detection in cloud computing based on time series anomalies utilizing machine learning. Journal of Cloud Computing Advances Systems and Applications, 12(1). https://doi.org/10.1186/s13677-023-00491-x
6. Alhazmi, L. (2023). Enhancing Cloud Security: An optimization-based deep learning model for detecting Denial-of-Service attacks. International Journal of Advanced Computer Science and Applications, 14(7). https://doi.org/10.14569/ijacsa.2023.0140737
7. T, S. R., & R, S. (2022). Ensemble Machine Learning Techniques for Attack Prediction in NIDS Environment. Iraqi Journal for Computer Science and Mathematics, 78–82. https://doi.org/10.52866/ijcsm.2022.02.01.008
8. Attou, H., Guezzaz, A., Benkirane, S., Azrour, M., &Farhaoui, Y. (2023a). Cloud-Based Intrusion Detection Approach Using Machine Learning Techniques. Big Data Mining and Analytics, 6(3), 311–320. https://doi.org/10.26599/bdma.2022.9020038
9. Ranjithkumar, S., & Pandian, S. C. (2021). Fuzzy Based Latent Dirichlet Allocation for Intrusion Detection in Cloud Using ML. Computers, Materials & Continua/Computers, Materials & Continua (Print), 70(3), 4261–4277. https://doi.org/10.32604/cmc.2022.019031
10. Ibrahim, N. M., & Zainal, A. (2019). An Adaptive Intrusion Detection Scheme for Cloud Computing. *International Journal of Swarm Intelligence Research*, *10*(4), 53–70. https://doi.org/10.4018/ijsir.2019100104
11. Yasarathna, T. L., &Munasinghe, L. (2020). Anomaly detection in cloud network data. Smart Computing and Systems Engineering, 62–67. https://doi.org/10.1109/scse49731.2020.9313014.
12. Qayyum, A., Aneeqa, I., Usama, M., Iqbal, W., Qadir, J., Elkhatib, Y., & Al-Fuqaha, A. (2020). Securing Machine Learning (ML) in the Cloud:A Systematic Review of Cloud ML Security. Frontiers in Big Data. https://eprints.lancs.ac.uk/id/eprint/148524/
13. Al-Ghuwairi, A., Sharrab, Y., Al-Fraihat, D., AlElaimat, M., Alsarhan, A., &Algarni, A. (2023). Intrusion detection in cloud computing based on time series anomalies utilizing machine learning. *Journal of Cloud Computing Advances Systems and Applications*, *12*(1). https://doi.org/10.1186/s13677-023-00491-x
14. Chkirbene, Z., Erbad, A., Hamila, R., Mohamed, A., Guizani, M., &Hamdi, M. (2020). TIDCS: a Dynamic Intrusion Detection and Classification system based feature selection. IEEE Access, 8, 95864–95877. https://doi.org/10.1109/access.2020.2994931
15. Attou, H., Mohy-Eddine, M., Guezzaz, A., Benkirane, S., Azrour, M., Alabdultif, A., &Almusallam, N. (2023). Towards an intelligent intrusion detection system to detect

malicious activities in cloud computing. Applied Sciences, 13(17), 9588. https://doi.org/10.3390/app13179588

16. Shahzad, F., Mannan, A., Javed, A. R., Almadhor, A. S., Baker, T., &Obe, D. A. (2022). Cloud-based multiclass anomaly detection and categorization using ensemble learning. Journal of Cloud Computing Advances Systems and Applications, 11(1). https://doi.org/10.1186/s13677-022-00329-y
17. B. Zhu, J. Li, R. Gu, and L. Wang, "An approach to cloud platform log anomaly detection based on natural language processing and lstm," in Proceedings of the 2020 3rd International Conference on Algorithms, Computing and Artificial Intelligence, 2020, pp. 1–7.
18. P. Parameswarappa, T. Shah, and G. R. Lanke, "A Machine Learning-Based Approach for Anomaly Detection for Secure Cloud Computing Environments," in 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), IEEE, 2023, pp. 931–940.
19. Nayyar, S., Arora, S., & Singh, M. (2020). Recurrent Neural network based Intrusion Detection System. IEEE, 0136–0140. https://doi.org/10.1109/iccsp48568.2020.9182099
20. Kunang, Y. N., Nurmaini, S., Stiawan, D., &Suprapto, B. Y. (2021). Attack classification of an intrusion detection system using deep learning and hyperparameter optimization. Journal of Information Security and Applications, 58, 102804. doi:10.1016/j.jisa.2021.102804
21. McDole, A., Abdelsalam, M., Gupta, M., & Mittal, S. (2020). Analyzing CNN based Behavioural Malware Detection Techniques on Cloud IAAS. In *Lecture notes in computer science* (pp. 64–79). https://doi.org/10.1007/978-3-030-59635-4_5
22. Thilagam, T., &Aruna, R. (2021). Intrusion detection for network based cloud computing by custom RC-NN and optimization. ICT Express, 7(4), 512-520.
23. Alhazmi, L. (2023). Enhancing Cloud Security: An optimization-based deep learning model for detecting Denial-of-Service attacks. International Journal of Advanced Computer Science and Applications, 14(7). https://doi.org/10.14569/ijacsa.2023.0140737
24. www.fixrunner.com
25. Nasim, S. S., Pranav, P., & Dutta, S. (2025). A systematic literature review on intrusion detection techniques in cloud computing. Deleted Journal, 28(1). https://doi.org/10.1007/s10791-025-09641-y