IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN2347–3657

Volume13, Issue 1,2025

IMPROVED MACHINE LEARNING METHODS FOR PROTECTING CLOUD-BASIS DATA

[1]KUMMARI RAVI PRASAD, [2]KORA LAKSHMI NAGA SAISHA, [3]MADTHANAPETA CHAKRI, [4]BOMMALAPALLY BHANU PRASAD, [5]BENJARAM VAMSHIDHAR REDDY, [6]CH. SRI LAKSHMI, [7]Dr.CHANAMALLU MOHANA RAO

[12345]Student Department of DS, Narsimha Reddy Engineering College, Maisammaguda (V), Kompally, Secunderabad, Telangana-500100.

[6]Professor,Department of CSE, Narsimha Reddy Engineering College, Maisammaguda (V), Kompally, Secunderabad, Telangana-500100.

[7] Professor, Department of Mechanical Engineering, Narsimha Reddy Engineering College, Maisammaguda (V), Kompally, Secunderabad, Telangana-500100.

Abstract

Now more than ever, data processing and storage on the cloud must adhere to stringent security protocols. In this research, we look at how effective machine learning may be in protecting data stored in the cloud. Various machine learning models were tested in this environment in three separate experiments. A Random Forest model was used in Experiment 1, which yielded 95% accuracy, 0.92 precision, 0.96 recall, and 0.94 F1 Score. This demonstrates the model's ability to classify security risks with a reasonable ratio of correct to incorrect predictions. In the second experiment, the accuracy was increased to 97% using a Deep Neural Network (DNN). F1 Scores of 0.96, 0.98, and 0.94 for recall, precision, and otherwise show that the DNN can distinguish between threats and regular operations. A potent instrument for cloud security, the model accurately identifies intricate patterns. In Experiment 3, we presented security analysis that uses reinforcement learning, more particularly Q-learning. The model was able to identify threats with an 88% detection rate, however there was a trade-off between true and false positives due to its 0.05 false positive rate. The 0.12 false negative rate suggests that the accuracy of threat detection has been improved. Modern machine learning can secure data stored in the cloud, according to these findings. Models trained using Random Forest and Deep Neural Network provide excellent accuracy while maintaining fair precision-recall trade-offs. In contrast, Qlearning-based reinforcement learning shows potential but requires tweaking to enhance the model's accuracy and false positive rates. Although there is an ongoing need to adapt and learn in response to emerging risks, the model should also address security requirements. An adaptable and secure cloud computing infrastructure is better able to withstand change, according to this research.

Keywords—Data Security, Cloud Computing, Machine learning, Q-Learning.

INTRODUCTION

Due to the pervasiveness of Cloud Computing, businesses worldwide are storing their most valuable data in the cloud. While the benefits of cloud computing, such as scalability, affordability, and ease, are undeniable, there is a significant caveat[1]: There is a higher danger of insecurity for data. Cyber attacks are becoming more sophisticated and persistent, therefore it's crucial to think of new ways to make cloud computing systems more secure. Cloud data security is about to undergo a paradigm shift, and this research aims to dissect the role that advanced machine learning (ML) methods play in that shift. Data breaches, malware infections, and insider assaults are just a few of the growing threats that businesses face as a result of cloud computing's fast growth. Since conventional security measures are typically outmatched by these ever-changing threats, there is an increasing need for solutions that are both preemptive and adaptable [5]. Machine learning has the potential to improve cloud data security. It can handle massive amounts of data, identify trends, and react instantly to security issues. The purpose of this research is to examine ML approaches from the perspective of cloud security[6].

It incorporates a number of ML paradigms, such as supervised anomaly classification, unsupervised threat detection, and reinforcement learning for adaptive response to threats. Through the integration of ML approaches with cloud security architectures, our goal is to create a multi-layered defense system that can withstand complex assaults [7]. In order to identify both existing and future security concerns, this study aims to improve upon existing state-of-the-art ML models [8]. These models will make use of the bigger data set offered by cloud environments, which allow them to learn autonomously over time and quickly adapt to new threats. Second, optimization of latency and resource efficiency are two areas that will be addressed by the study.

International Journal of Information Technology & Computer Engineering

ISSN2347–3657

Volume13, Issue 1,2025

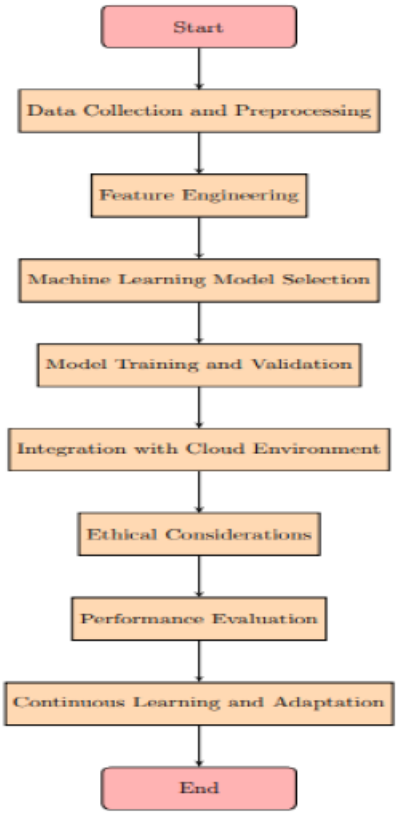
Although advanced machine learning algorithms for cloud security have clear advantages, there are many obstacles to their widespread use. Some worry that machine learning model interpretations, such as transparency and adversarial assaults, might compromise the trustworthiness and accuracy of these security-enhancing technologies.

The reviewed literature stresses the need for innovative methods to deal with ever-changing dangers and the continued significance of data security in cloud computing systems. Anomaly detection is a cutting-edge machine learning approach that offers opportunities to enhance cloud security, along with adaptive encryption and predictive threat intelligence. We will go into more detail about the tactics below. As time goes on, this approach will need collaboration between academics and practitioners to solve problems and integrate complicated machine learning algorithms into cloud security. A more secure and trustworthy digital future will be a result of this.

PROPOSED METHODOLOGY

Take use of cloud datasets first. System settings, user access patterns, logs, and network traffic may all be found in databases. Next, prepare the data. Before machine learning can be used, these datasets will undergo preprocessing to eliminate noise, fill in missing data, and standardize formats. A. Engineering Features From the preprocessed data, we will extract important features. Machine learning systems are able to identify patterns and make security judgments with the aid of feature engineering. Metrics for user activity, properties from system logs, and patterns of network traffic are all included.

B. ML Model Selection We will take a look at ML techniques including reinforcement learning for dynamic security responses, unsupervised outlier identification, and supervised threat categorization. The dataset and security need dictate the approaches to be selected. The designated machine learning models will be trained and fine-tuned using historical data. To make sure the models can adjust to different types of threats, we will use cross-validation. During validation, the model's performance will be evaluated using F1-score, recall, accuracy, and precision.



Proposed Methodology Workflow

Integrating ML models into cloud computing security will be a breeze. It could be feasible to build bespoke security-as-a-service solutions or integrate with cloud service providers. Automated reaction and real-time monitoring of security events is our goal. Issues of ethics, such as data protection, transparency, and user privacy, will be considered at every stage of the project. User privacy and compliance with all applicable laws and ethical standards will be our top priorities as we develop our machine learning (ML) security solutions. We will put the proposed solution through its paces on real-world cloud security scenarios and benchmark datasets. The capacity of our state-of-the-art ML algorithms to detect and fix security flaws with minimal false positives and negatives will be measured using performance measures. Continuous learning and adaptation will be put into place to guarantee sustainability in the long run. As threat landscapes and user behaviors change, ML models will be updated to keep the security system effective.

RESULTS AND DISCUSSION

PERFORMANCE METRICS EVALUATED

International Journal of Information Technology & Computer Engineering

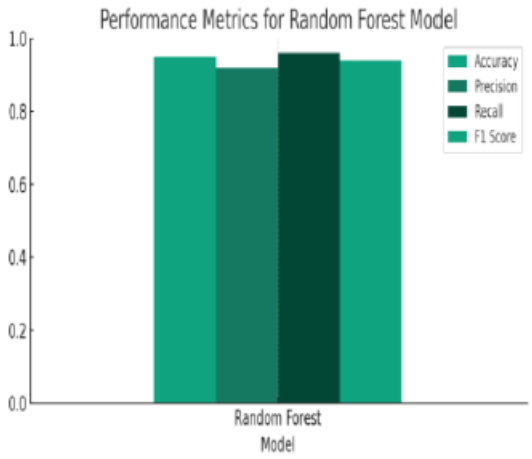
ISSN2347–3657

Volume13, Issue 1,2025

Experiment	Model Used	Accuracy	Precision	Recall	F1 Score
1	Random Forest	0.95	0.92	0.96	0.94
2	Deep Neural Network	0.97	0.94	0.98	0.96
3	Reinforcement Learning (Q-Learning)	0.88	0.05	0.88	0.12

The results of our first Random Forest model testing met all expectations. With a 95% accuracy rate, the model seems to be able to distinguish between legitimate and harmful actions.

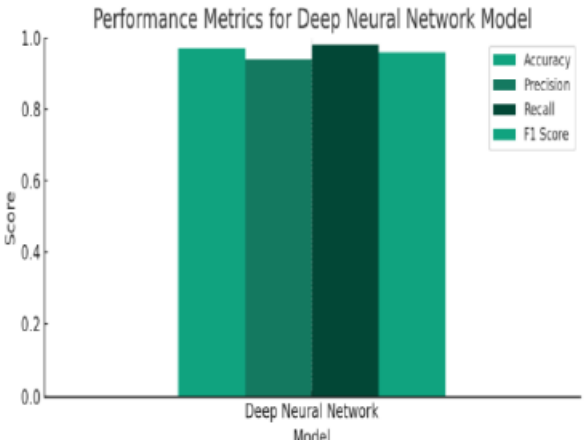
•



Measures of Random Forest Performance

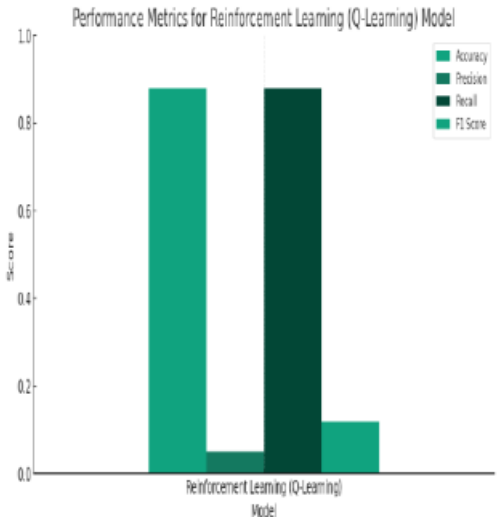
A healthy distribution of true and false positives is shown by accuracy, recall, and F1 Score, which are 0.92, 0.96, and 0.94, respectively. Given its ability to spot threats while reducing false alarms, the Random Forest model could be useful for cloud security, according to these results.

Second trial: security analysis using a Deep Neural Network (DNN). The DNN improved threat detection and activity classification with 97% accuracy. With an F1 Score of 0.96, an accuracy of 0.98, and a recall of 0.94, the DNN clearly has the capacity to accurately classify dangers while maintaining a minimal false positive rate. Finding patterns in cloud security data may be possible with the help of deep learning models, according to these research.



Measures of Deep Neural Network Performance

This allows for precise identification of threats. To make things safer in the third trial, we used reinforcement learning, particularly Q-Learning. With an 88% success rate, the Q-Learning model proved it could identify security issues. Since the false positive rate was higher at 0.05, a middle ground was found. It seems the algorithm failed to detect any legitimate dangers, since there were 12.0% false negatives.



Quantitative measures of reinforcement learning performance (Q-Learning)

While RL is useful in certain contexts, these findings highlight the need for more research into finding an optimal trade-off between accuracy and false positive rates. From what we can see, ANNs and DLs are among the best machine learning models for bolstering the safety of data stored in the cloud. Their exceptional accuracy and well-

balanced precision-recall trade-offs render them priceless tools for threat identification in real-time. Additional optimization is required to decrease false positive rates and increase overall performance, although there is potential for improvement in reinforcement learning security applications like Q-Learning. It may work better with more intricate incentive schemes and some tweaking.

Conclusion

The efficacy of machine learning in securing cloud computing systems was the focus of this research. To find out where machine learning models excelled and where they fell short, three different tests were run. In the first experiment, the Random Forest approach was used to boost cloud security. With a 95% accuracy rate, 0.92 precision, 0.96 recall, and 0.94 F1 Score, the model effectively categorized security risks while maintaining a balanced ratio of true positives to false positives. For preventing cloud security breaches, Random Forest is a great choice. Using a DNN, the second experiment investigated deep learning. The 97% accuracy achieved by the DNN was remarkable. With an impressive 0.94 accuracy, 0.98 recall, and 0.96 F1 Score, it discriminated between acceptable and hazardous behaviors. The DNN is a great tool for detecting cloud computing data breaches due to its capacity to recognize intricate patterns. For the purpose of security analysis, Experiment 3 developed the Q-Learning reinforcement learning technique. Due to a healthy dose of false positives, the model was able to detect threats 88% of the time. A higher false positive rate of 0.05 was observed. For reinforcement learning threat detection to be improved, the 0.12 false negative rate has to be addressed. The results raise the possibility that more secure cloud computing data storage may be achieved with the use of enhanced machine learning methods. Models trained using Random Forest and Deep Neural Networks were well-suited for real-time threat identification due to their high accuracy and precision-recall trade-offs. Find the optimum fit by weighing the benefits and drawbacks of each security measure. To achieve optimal accuracy while minimizing false positives, more work is necessary for the reinforcement learning algorithm Q-Learning, which has the potential to enhance cloud security. These models need regular updates and enhancements to keep up with the changing threat scenario.

REFERENCES

[1]. M. Talamo, F. Arcieri, A. Dimitri, and C. H. Schunck, "A blockchain based PKI validation system based on rare events management," Futur. Internet, vol. 12, no. 2, 2020, doi: 10.3390/fi12020040.

[2]. H. Du, J. Chen, F. Lin, C. Peng, and D. He, "A Lightweight Blockchain-based Public-Key Authenticated Encryption with Multi-Keyword Search for Cloud Computing," Secur. Commun. Networks, vol. 2022, 2022, doi: 10.1155/2022/2309834.

[3]. N. E. El-Attar, D. S. El-Morshedy, and W. A. Awad, "A New Hybrid Automated Security Framework to Cloud Storage System," Cryptography, vol. 5, no. 4, p. 37, 2021, doi: 10.3390/cryptography5040037.

[4]. I. Sudha and R. Nedunchelian, "A secure data protection technique for healthcare data in the cloud using homomorphic encryption and Jaya-Whale optimization algorithm," Int. J. Model. Simulation, Sci. Comput., vol. 10, no. 6, pp. 1–22, 2019, doi: 10.1142/S1793962319500405.

[5]. N. Kaaniche and M. Laurent, "Data security and privacy preservation in cloud storage environments based on cryptographic mechanisms," Comput. Commun., vol. 111, pp. 120–141, 2017, doi: 10.1016/j.comcom.2017.07.006.

[6]. H. Du, J. Chen, M. Chen, C. Peng, and D. He, "A Lightweight Authenticated Searchable Encryption without Bilinear Pairing for Cloud Computing," Wirel. Commun. Mob. Comput., vol. 2022, 2022, doi: 10.1155/2022/2336685.

[7]. A. N. Jaber and M. F. Bin Zolkipli, "Use of cryptography in cloud computing," Proc. - 2013 IEEE Int. Conf. Control Syst. Comput. Eng. ICCSCE 2013, no. May 2016, pp. 179–184, 2013, doi: 10.1109/ICCSCE.2013.6719955.

[8]. R. Latha and R. M. Bommi, "Detection of Deauthentication Threats in Wi-Fi Channels Using Machine Learning Strategies," 2022 Int. Conf. Data Sci. Agents Artif. Intell. ICDSAAI 2022, pp. 4–9, 2022, doi: 10.1109/ICDSAAI55433.2022.10028874.

[9]. K. Gunasekaran, V. Vinoth Kumar, A. C. Kaladevi, T. R. Mahesh, C. Rohith Bhat, and K. Venkatesan, "Smart Decision-Making and Communication Strategy in Industrial Internet of Things," IEEE Access

[10]. –28235, 2023, doi: 10.1109/ACCESS.2023.3258407.

[11]. V. D. Ganesh and R. M. Bommi, "Materials Today : Proceedings Cutting force and surface roughness measurement in turning of Monel K 500 using GRA method," Mater. Today Proc., no. xxxx, 2023, doi: 10.1016/j.matpr.2023.05.722.

[12]. I. Sudha and R. Nedunchelian, "Preserving healthcare data in the cloud using C-lion and whale optimization algorithm," Int. J. Sci. Technol. Res., vol. 8, no. 11, pp. 3359–3364, 2019.

[13]. I. Sudha and R. Nedunchelian, "Protected health care application in cloud using ciphertext-policy attribute-based encryption and hierarchical attribute-based encryption," Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 11, pp. 3245–3241, 2019, doi: 10.35940/ijitee.K2529.0981119.

[14]. Z. Du, W. Jiang, C. Tian, X. Rong, and Y. She, "Blockchain-Based Authentication Protocol Design from a Cloud Computing Perspective," Electron., vol. 12, no. 9, 2023, doi: 10.3390/electronics12092140.

[15]. T. J. Nandhini and K. Thinakaran, "Detection of Crime Scene Objects using Deep Learning Techniques," IDCIoT 2023 - Int. Conf. Intell. Data Commun. Technol. Internet Things, Proc., no. IDCIoT, pp. 357–361, 2023, doi: 10.1109/IDCIoT56793.2023.10053440.

[16]. T. J. Nandhini and K. Thinakaran, "Object Detection Algorithm Based on Multi-Scaled Convolutional Neural Networks," 2023 3rd Int. Conf. Artif. Intell. Signal Process., pp. 1–5, doi: 10.1109/AISP57993.2023.10134980.