IJARST
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

APPLICATIONS OF MACHINE LEARNING ALGORITHMS ON CLOUD COMPUTING SECURITY

[1]VANGALA KALYANI

Faculty, at JNTUH Univercity

Email Id: echo.kalyani@gmail.com

ABSTRACT

With the advances in machine learning (ML) and deep learning (DL) techniques, and the potency of cloud computing in offering services efficiently and cost-effectively, Machine Learning as a Service (MLaaS) cloud platforms have become popular. In addition, there is increasing adoption of third-party cloud services for outsourcing training of DL models, which requires substantial costly computational resources (e.g., high-performance graphics processing units (GPUs)). Such widespread usage of cloud-hosted ML/DL services opens a wide range of attack surfaces for adversaries to exploit the ML/DL system to achieve malicious goals. The popularity and usage of Cloud computing is increasing rapidly. Several companies are investing in this field either for their own use or to provide it as a service for others. One of the results of Cloud development is the emergence of various security problems for both industry and consumer. One of the ways to secure Cloud is by using Machine Learning (ML). Machine learning (ML) is the investigation of computer algorithms that improve naturally through experience. In this work, we use the UNSW dataset to train the supervised machine learning models. We then test these models with ISOT dataset. We present our results and argue that more research in the field of machine learning is still required for its applicability to the cloud security.

Keywords — Cloud, Machine Learning, Cloud Security, Privacy, cloud-hosted machine learning models.

1. INTRODUCTION

Recently cloud computing is gaining significant traction and virtualized data centers are becoming popular as a costeffective infrastructure and solution for enterprise applications. Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) are being widely deployed and utilized by the end users. In this way, users neither require knowledge, control, and ownership in the computing infrastructure nor they need to host, control or own an infrastructure in order to deploy their applications. Instead, they simply access or rent the hardware or software paying only for what they use. The possibility of paying-as-you-go along with on-demand elastic operations by cloud hosting providers is gaining popularity in enterprise computing model [1]. Regardless of its advantages, the transition to this computing paradigm is hampered by major security issues, which are the subject of many recent studies. Recently there has been much interest in Machine Learning (ML) techniques for network and cloud security.

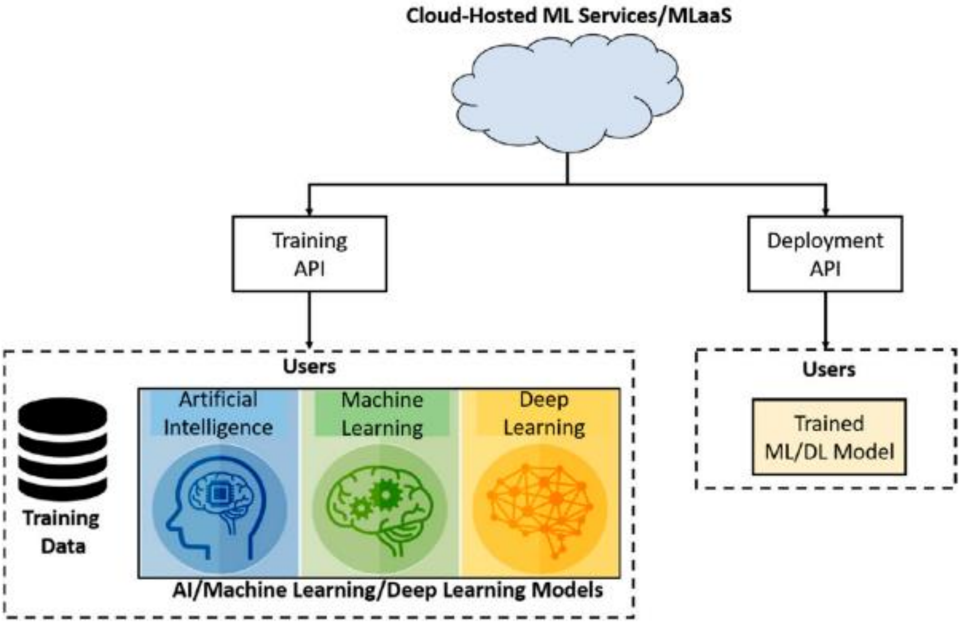


FIGURE 1 | Taxonomy of different defenses proposed for defending attacks on the third-party cloud-hosted machine learning (ML) or deep learning (DL) models.

MLaaS refers to different ML services that are offered as a component of a cloud computing services, for example, predictive analytics, face recognition, natural language services, and data modeling APIs. MLaaS allows users to upload their data and model for training at the cloud. In addition to training, cloudhosted ML services can also be used for inference purposes, that is, models can be deployed on the cloud environments; the system architecture of a typical MLaaS is shown in Figure 1.

ML algorithms are used to solve security issues and manage data more efficiently [11]. ML is the use of man-made consciousness that enables frameworks to normally take in and improve truly without being expressly customized [12]. ML focuses on the advancement of computer programs that can find a suitable pace use it to learn for themselves [13]. The approach toward learning starts with perceptions or data, such as models, direct understanding, or heading, to channel for structures in information and pick better decisions later on the subject to the models that are given.

2. LITERATURE REVIEW

Omar Abdel Wahab et.al (2021) Cloud-based systems are subject to various attack types launched by Virtual Machines (VMs) manipulated by attackers having different goals and skills. The existing detection and defense mechanisms might be suitable for simple attack environments but become ineffective when the system faces advanced attack scenarios wherein simultaneous attacks of different types are involved. This is because these mechanisms overlook the attackers' strategies in the detection systems design, ignore the system's resource constraints, and lack sufficient knowledge about the attackers' types and abilities.

Phuc Trinh Dinh et.al (2021) Cloud computing is now considered to be the most cost-effective platform for offering business and consumer IT services over the Internet. However, it is prone to new vulnerabilities. Specifically, a newly discovered type of attack, called an economic-denial-of-sustainability attack known as EDoS, exploits the pay-per-use model to scale up the resource usage over time to the degree that the cloud user has to pay for the unexpected usage charge. To prevent EDoS attacks, we propose an effective solution in the SDN-based cloud computing environment. We first introduce a machine-learning-based approach adopting a framework called MAD-GAN which applies an unsupervised multivariate anomaly detection technique based on Generative Adversarial Networks (GANs), using the Long-Short-Term-Memory Recurrent Neural Networks (LSTM-RNN) to detect EDoS attacks.

Phuc Trinh Dinh et.al (2021) Software-defined networking (SDN) nowadays is extensively being used in a variety of practical settings, provides a new way to manage networks by separating the data plane from its control plane. However, SDN is particularly vulnerable to Distributed Denial of Service (DDoS) attacks because of its centralized control logic. Many studies have been proposed to tackle DDoS attacks in an SDN design using machine-learning-based schemes; however, these feature-based detection schemes are highly resource-intensive and they are unable to perform reliably in such a large-scale SDN network where a massive amount of traffic data is generated from both control and data planes. This can deplete computing resources, degrade network performance, or even shut down the network systems owing to being exhausting resources.

PradoshPriyadarshan et.al (2021) in recent years, cyber security has become a challenging task to protect the networks and computing systems from various types of digital attacks. Therefore, to preserve these systems, various innovative methods have been reported and implemented in practice. However, still more research work needs to be carried out to have malware free computing system. In this paper, an attempt has been made to develop simple but reliable ML based malware detection systems which can be implemented in practice. Keeping this in view, the present paper has proposed and compared the performance of three ML based malware detection systems applicable for computer systems. The proposed methods include k-NN, RF and LR for detection purpose and the features extracted comprise of Byte and ASM.

Cheng Cheng et.al (2021) As exploitation of low and medium airspace for air traffic management (ATM) is gaining more attention, aerial vehicles' security issues pose a major challenge to the AirGround Integrated Vehicle Networks (AGIVN). Traditional surveillance technology lacks the capacity to support the intensive air traffic management (ATM) of the future. Therefore, an advanced automatic dependent surveillance-broadcast (ADS-B) technique is applied to track and monitor aerial vehicles in a more effective manner. In this paper, we propose a grouping-based conflict detection algorithm based on the preprocessed ADS-B dataset, and analyze the experimental results and visualize the detected conflicts.

3. SECURITY ANALYSIS

3.1 Cloud Threats

CC is a developing model that has significant potential to grow and is becoming widely popular. However, even with its unique characteristics, it has various security threats and protection challenges, as discussed in this section [30]. The categorization is performed based on the CIA Triad and attacks on cloud components.

3.2 Cloud Security Threats

The major security threats in CC are classified under confidentiality, integrity and availability. These issues are discussed briefly here.

1. Confidentiality threats involves an insider threat to client information, risk of external attack, and data issues. First, insider risk to client information is related to unapproved or illegal access to customer information from an insider of a cloud service provider is a significant security challenge. Second, the risk of outside attack is increasingly relevant for cloud applications in unsecured area. This risk includes remote software or hardware hits on cloud clients and applications. Third, information leakage is an unlimited risk to cloud bargain data because of human mistake, lack of instruments, secured access failures, after which anything is possible.

2. Integrity threats involve the threats of information separation, poor client access control, and risk to information quality. First is the risk of information isolation, which inaccurately joins the meanings of security parameters, ill-advised design of VMs, and off base client-side hypervisors. This is complicated issue inside the cloud, which offers assets connecting the clients; if assets change, that could affect information trustworthiness. Next is poor client access control, which because of inefficient access and character control has various issues and threats that enable assailants harm information assets.

3. Availability threats include the effect of progress on the board, organization non-accessibility, physical interruption of assets, and inefficient recovery strategies. First is the effect of progress on the board that incorporates the effect of the testing client entrance for different clients, and the effect of foundation changes. Both equipment and application change inside the cloud condition negatively affect the accessibility of cloud organizations. Next is the non-accessibility of services that incorporate the non-accessibility of system data transfer capacity, domain name system (DNS) organization registering software, and assets.

3.3 Attacks on the Cloud

Four relative analysis attacks are classified by their segments: network-based, VM-based, storage-based, and application-based as depicted in Figure 2.

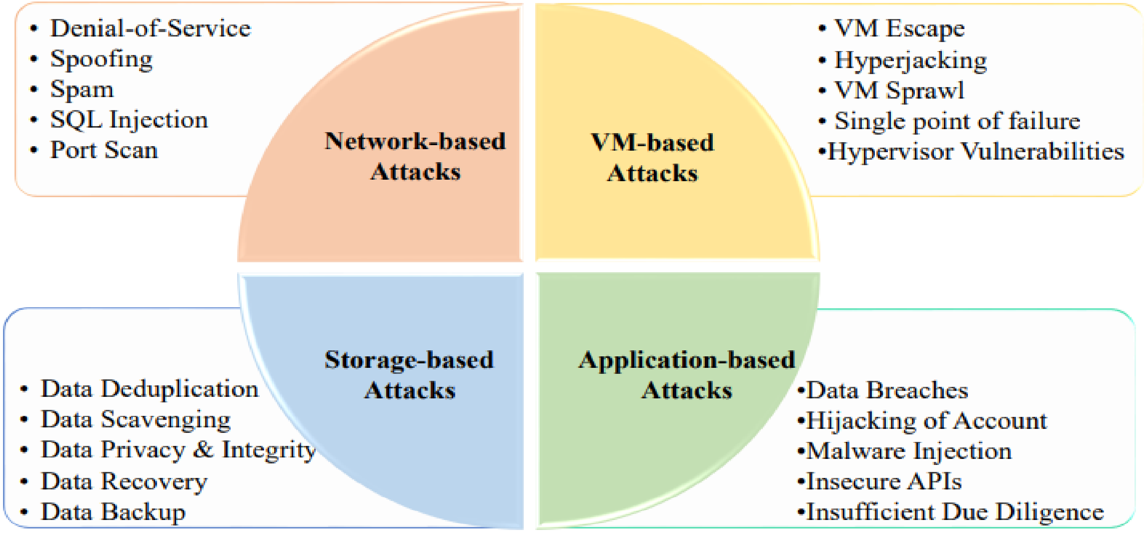


Figure 2. Attacks on cloud components.

1. Network-based attacks: Three types of system attacks discussed here are port checking, botnets, and spoofing attacks. A port scan is useful and of considerable interest to hackers in assessing the attacker to collect relevant information to launch a successful attack. Based on whether a network's defense routinely searches ports, the defenders usually do not hide their identity, whereas the attackers do so during port scanning. A botnet is a progression of malware-contaminated web associated devices that can be penetrated by hackers. A spoofing assault is when a hacker or malicious software effectively operates on behalf of another user (or system) by impersonating data. It occurs when the intruder pretends to be someone else (or another machine, such as a phone) on a network to manipulate other machines, devices, or people into real activities or giving up sensitive data.

2. VM-based attacks: Different VMs facilitated on a frameworks cause multiple security issues. A side-channel assault is any intrusion based on computer process implementation data rather than flaws in the code itself. Malicious code that is placed inside the VM image will be replicated during the creation of the VM. VMs picture the executive's framework offers separating and filtering for recognizing and recovering from the security threats.

3. Storage-based attacks: A strict monitoring mechanism is not considered then the attackers steal the important data stored on some storage devices. Data scavenging refers to the inability to completely remove data from storage devices, in which the attacker may access or recover this data. Data de-duplication refers to duplicate copies of the repeating data. This attack is mitigated by ensuring the duplication occurs when the precise number of file copies is specified.

4. Application-based attacks: The application running on the cloud may face many attacks that affect its performance and cause information leakage for malicious purposes. The three primary applications-based attacks are malware infusion and stenography attacks, shared designs, web services, and convention-based attacks.

4. RESEARCH METHODOLOGY

4.1 Design of the Cloud

As indicated by CC architecture, the five most significant elements affect and are affected by CC, alongside its security suggestions. Figure 3 illustrates the design of CC that includes a start-to-finish reference design that represents the layers of the Open Systems Interconnection (OSI) Model. CC is a complicated design with multiple zones of vulnerability. The components of CC are as follows:

• Cloud Consumer: An individual or association that maintains career, relationship, and utilization administrations from the cloud providers.

• Cloud Provider: An individual or organization for manufacturing, or administration, available to invested individuals.

• Cloud Auditor: A gathering that can direct the self-sufficient examination of cloud organizations, information system activities, implementation, and security of cloud users.

• Cloud Broker: A substance that manages the usage, implementation, and conveyance of cloud benefits and arranges links between cloud purchasers and cloud suppliers.

• Cloud Carrier: A medium that offers a system of cloud administrations from cloud suppliers to the cloud consumers.

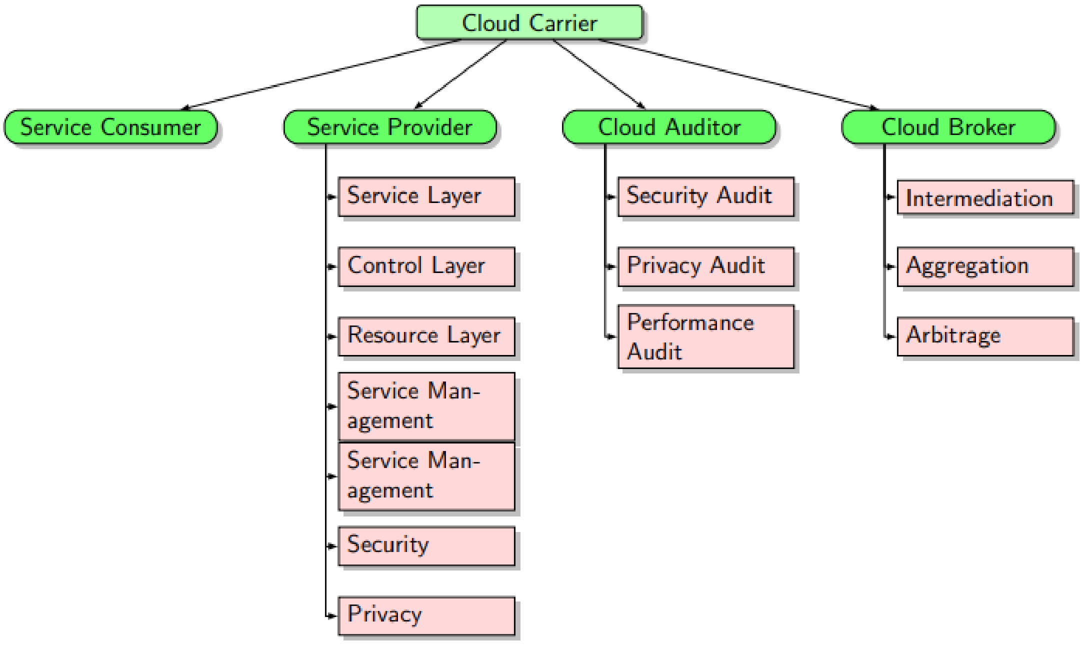


Figure 3. Cloud computing architecture.

4.2 Cloud Deployment Models

CC has four deployment models: private, public, hybrid, and community cloud [28]. Each deployment model has different costs and value propositions. Therefore, deciding the deployment model is a difficult and critical decision. Figure 4 illustrates the cloud deployment models. Table 1 presents a comparative analysis of the benefits and issues of cloud models.

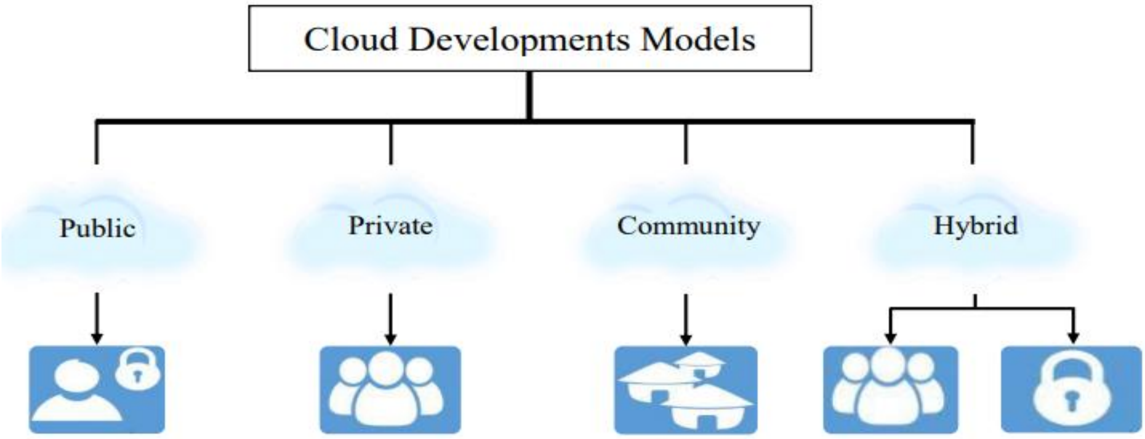
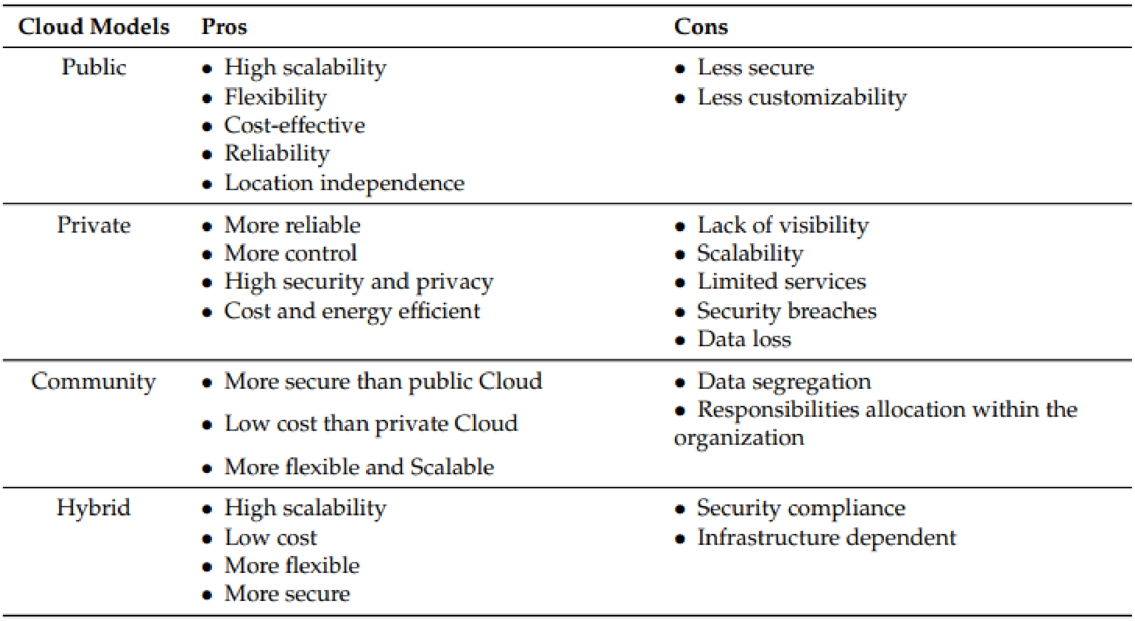


Figure 4. Cloud deployment models.

Table 1. Comparative analysis of the cloud deployment models.



Cloud Models	Pros	Cons
Public	• High scalability • Flexibility • Cost-effective • Reliability • Location independence	• Less secure • Less customizability
Private	• More reliable • More control • High security and privacy • Cost and energy efficient	• Lack of visibility • Scalability • Limited services • Security breaches • Data loss
Community	• More secure than public Cloud • Low cost than private Cloud • More flexible and Scalable	• Data segregation • Responsibilities allocation within the organization
Hybrid	• High scalability • Low cost • More flexible • More secure	• Security compliance • Infrastructure dependent

5. ML ALGORITHMS FOR THE CLOUD SECURITY

In this section, we study different ML algorithms that have been used to overcome the security issues in CC.

Supervised Learning

Supervised learning is the ML task of learning a limit that maps a commitment to a yield based on model data yield sets. It infers a limit from data involving many planning models. The supervised ML algorithms are those algorithms that require outside help.

Supervised ANNs

ANNs are the bits of a computing framework intended to recreate how the human mind analyzes and processes data. They are the establishments of ML that solve issues otherwise impossible or troublesome for humans or statistical principles. Hussin et al. [59] predicted basic distributed computing security issues using ANN algorithms. An ANN algorithm was used to determine security issues in a banking organization. ANNs were used for improving the execution and learning neural capacities. Levenberg-Marquardt (LMBP) algorithms were used to predict the presentation for the cloud security level. LMBP is a nonlinear improvement model used to measure the exactness of the forecasts present and decrease the error between genuine yields and focus for the preparation procedure; the mean square error (MSE) is estimated to decide the presentation.

K-NN

K-NN is likely the simplest algorithm among the ML algorithms for relapse and classification issues. The K-NN algorithm uses information and characterizes new information based on similarity measures (e.g., distance). Classification is finished by a larger part vote to its neighbors. The security of information in the cloud remains challenge. Different frameworks are being used to enhance cloud data security, such as data encryption. The methodologies of information security cannot be applied. The comprehension of the necessities of security is fundamental to the legitimate use of these measures. Zardari et al.

[70] proposed a data classification approach based on data confidentiality. The authors described a methodology of information grouping that depends on the security and protection of information. The K-NN method of information arrangement was executed in the cloud administrations and virtual conditions. The target of using K-NN incorporates the grouping of information based on their security prerequisites. The information was grouped into two classes: touchy and non-delicate (or open) information.

Naive Bayes

In ML, Navies Bayes classifiers are a group of basic "probabilistic classifiers" that apply Bayes' hypothesis with solid (naive) freedom suppositions between the highlights. They are among the least complex Bayesian system models. Designed a distributed denial-of-service (DDoS) detection system based on the C4.5 algorithm to mitigate the DDoS threat. The hidden innovations and legacy conventions contain bugs and vulnerabilities that can enable interruption by the attackers. Assaults, such as DDoS, cause serious harm and influence the performance of the cloud. DDoS assaults have become one of the fundamental dangers to security. A DDoS attack executes an assault by permitting an interloper to interact with a computerized PC organization. Infected with malware, PCs and different machines (e.g., IoT devices) transform into bots (or zombie). Then, the assailant has remote control over the bots, which is known as a botnet. The traditional intrusion detection techniques have limitations such as large false alarms, noise that reduces the capabilities of the IDS by generating the rate of a false alarm, and constant updating of software to track the new threats. ML methods are acquainted with call attention to the dangers more productively than traditional IDS. Distinctive ML algorithms are used to identify the threat in a DDoS.

SVM

SVM is an ML algorithm that investigates information for grouping and regression analysis. SVM is a supervised learning technique that analyzes at information and sorts it into one of two classes. An SVM outputs a guide of the arranged information with the edges between the two as far separated as could reasonably be expected.

6. RESULTS AND DISCUSSION

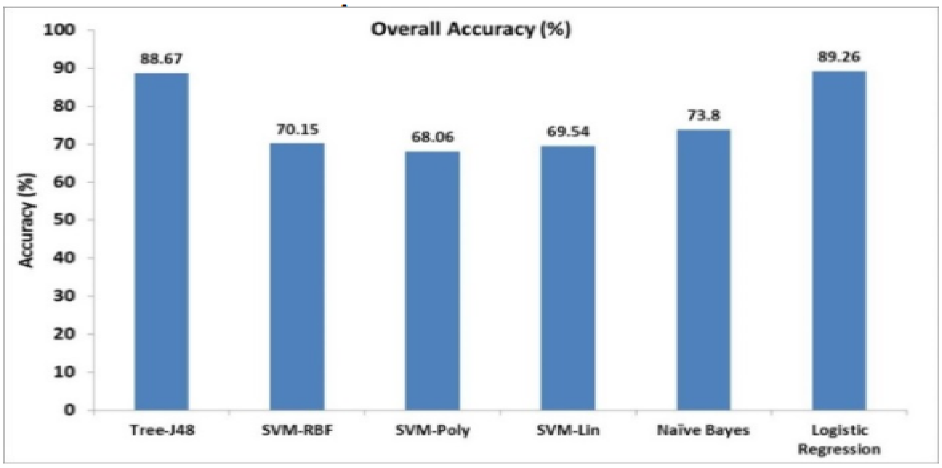


Fig. 5: Overall accuracy with UNSW Dataset

Fig. 5 displays the overall performance of the machine learning techniques mentioned earlier. We observe that with the UNSW training and testing datasets, logistic regression performs the best with 89.26% accuracy, followed by J48 algorithm, with 88.67% accuracy. The polynomial SVM algorithm has the least accuracy with 68.06% accuracy. However, overall accuracy is not enough to compare the performance of these algorithms.

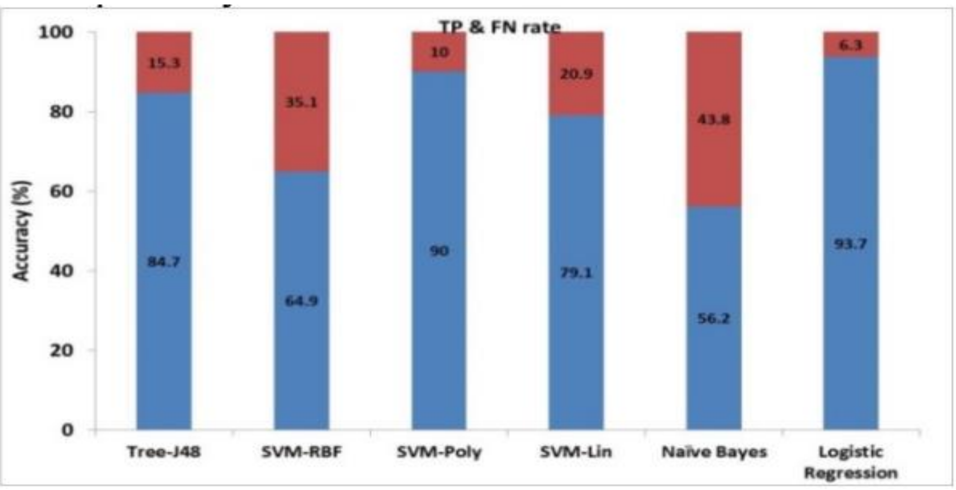


Fig. 2: TP and FN rate with UNSW dataset

Fig. 6 below shows the TP and FN percentages for these algorithms. We observe that logistic regression performs the best, with a 93.7 % rate for identifying anomalous packet correctly (TP), while it identifies 6.3% of anomalous packets as normal (FN), which is an error case. J48 follows with the TP and FN percentages as 84.7 and 15.3 respectively. Naïve Bayes performs the worst with the percentages as 56.2 and 43.8 respectively.

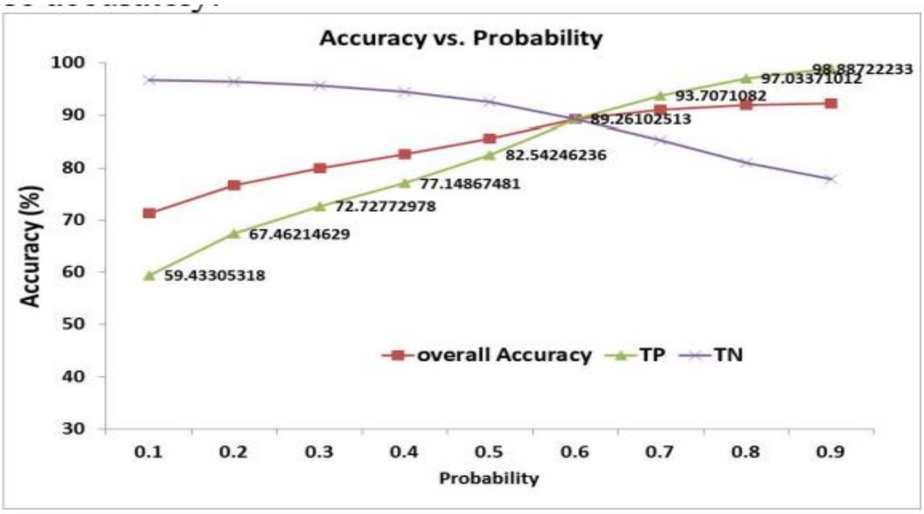


Fig. 7: Varying Frequency threshold with UNSW dataset

Fig. 7 shows the overall accuracy of these models with the ISOT dataset. As we can see, the overall accuracy is still impressive, that is around 95% with J48 and LR and around 90% with SVM.

CONCLUSIONS

In this study, security threats and attacks as the most challenging issues in CC were analyzed. Different types of ML algorithms e.g., ANNs, K-NN, Naïve Bayes, SVM, K-Means, and SVD were investigated as solutions to address the security issues in CC. We reviewed several proposed techniques that used ML algorithms for cloud security. Our study also found very few surveys based on ML techniques in Cloud security form, with no usage of their feature selection/extraction strategy. Therefore, we recommend more thorough research and more empirical experiments to address the need for ML in Cloud security. In addition, research papers should present their results using multiple evaluation metrics when considering imbalanced datasets. Moreover, we noticed that little work has been done using deep learning techniques in cloud security. We encourage researchers to take advantage of the deep learning in this regard

REFERENCES

[1] M. T. Khorshed, A. B. M. S. Ali, and S. A. Wasimi, ‘‘Trust issues that create threats for cyber attacks in cloud computing,’’ in Proc. IEEE 17th Int. Conf. Parallel Distrib. Syst., Dec. 2011, pp. 900–905, doi: 10.1109/ICPADS.2011.156.

[2] A. P. Achilleos, K. Kritikos, A. Rossini, G. M. Kapitsaki, J. Domaschka, M. Orzechowski, D. Seybold, F. Griesinger, N. Nikolov, D. Romero, and G. A. Papadopoulos, ‘‘The cloud application modelling and execution language,’’ J. Cloud Comput., vol. 8, no. 1, p. 20, Dec. 2019, doi: 10.1186/s13677-019-0138-7.

[3] P. Kumar and P. J. A. Alphonse, ‘‘Attribute based encryption in cloud computing: A survey, gap analysis, and future directions,’’ J. Netw. Comput. Appl., vol. 108, pp. 37–52, Apr. 2018, doi: 10.1016/j.jnca.2018.02.009.

[4] T. Halabi and M. Bellaiche, ‘‘Towards quantification and evaluation of security of cloud service providers,’’ J. Inf. Secur. Appl., vol. 33, pp. 55–65, Apr. 2017, doi: 10.1016/j.jisa.2017.01.007.

[5] R. Kumar, S. P. Lal, and A. Sharma, ‘‘Detecting denial of service attacks in the cloud,’’ in Proc. IEEE 14th Int. Conf. Dependable, Autonomic Secure Comput., 14th Int. Conf. Pervas. Intell. Comput., 2nd Int. Conf. Big Data Intell. Comput. Cyber Sci. Technol. Congr. (DASC/PiCom/DataCom/CyberSciTech), Aug. 2016, doi: 10.1109/DASCPICom-DataCom-CyberSciTec.2016.70.

[6] T. Halabi and M. Bellaiche, ‘‘A broker-based framework for standardization and management of cloud security-SLAs,’’ Comput. Secur., vol. 75, pp. 59–71, Jun. 2018, doi: 10.1016/j.cose.2018.01.019.

[7] M. R. Chinnaiah and N. Niranjan, ‘‘Fault tolerant software systems using software configurations for cloud computing,’’ J. Cloud Comput., vol. 7, p. 3, Jan. 2018, doi: 10.1186/s13677-018-0104-9.

[8] B. Xu, S. Chen, H. Zhang, and T. Wu, ‘‘Incremental k-NN SVM method in intrusion detection,’’ in Proc. 8th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS), Nov. 2017, pp. 712–717, doi: 10.1109/ICSESS.2017.8343013.

[9] R. Moreno-Vozmediano, R. S. Montero, E. Huedo, and I. M. Llorente, ‘‘Efficient resource provisioning for elastic cloud services based on machine learning techniques,’’ J. Cloud

Comput., vol. 8, no. 1, p. 5, Dec. 2019, doi: 10.1186/s13677-019-0128-9.

[10] A. AlEroud and G. Karabatis, ‘‘A contextual anomaly detection approach to discover zero-day attacks,’’ in Proc. Int. Conf. Cyber Secur., Dec. 2012, pp. 40–45, doi: 10.1109/CyberSecurity.2012.12.

[11] Omar Abdel Wahab;JamalBentahar;HadiOtrok;AzzamMourad Resource-Aware Detection and Defense System against Multi-Type Attacks in the Cloud: Repeated Bayesian Stackelberg Game IEEE Transactions on Dependable and Secure Computing Year: 2021 DOI: 10.1109/TDSC.2019.2907946

[12] Phuc Trinh Dinh;Minho Park Economic Denial of Sustainability (EDoS) Detection using GANs in SDN-based Cloud 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE) Year: 2021 DOI: 10.1109/ IEEE PhuQuoc Island, Vietnam

[13] Phuc Trinh Dinh;Minho Park BDF-SDN: A Big Data Framework for DDoS Attack Detection in Large-Scale SDN-Based Cloud 2021 IEEE Conference on Dependable and Secure Computing (DSC) Year: 2021

[14]PradoshPriyadarshan;PrateekSarangi;AdyashaRath;Ganapati Panela Machine Learning Based Improved Malware Detection Schemes 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) : 2021

[15] Cheng Cheng;LiangGuo;TongWu;JinlongSun;GuanGui;BamideleAdebisi; HarisGacanin;Hikmet Sari Machine Learning-Aided Trajectory Prediction and Conflict Detection for Internet of Aerial Vehicles IEEE Internet of Things Journal Year: 2021