*Journal of Science and Technology*
*ISSN: 2456-5660 Volume 06,Special Issue 01, August 2021,*

www.jst.org.in *DOI: https://doi.org/10.46243/jst.2021.v6.i04.pp01-05*

# Cloud Security using Machine Learning

Divya Chhaprwal[1], Prof. Mohan Yelpale[2]

*[1,2]Dept. of Computer Engineering, NBN Sinhgad School of Engineering, Pune, Maharashtra, India*

***To Cite this Article***
*Divya Chhaprwal, Prof. Mohan Yelpale , "Cloud Security using Machine Learning", Journal of Science and Technology, Vol. 06,Special Issue 01, August 2021, pp01-05.*

***Article Info***
*Received: 15.07.2021 Revised: 24.07.2021 Accepted:10.08.2021 Published: 16.08.2021*

***Abstract:*** *Cloud computing is quickly gaining in popularity and utilisation. Several businesses are investing in this subject, either for their own benefit or as a service to others. The creation of numerous security risks for both industry and consumers is one of the consequences of Cloud development. Machine Learning is one of the methods for securing the cloud (ML). On the Cloud, machine learning techniques have been employed in a variety of methods to avoid or attack detection and security flaws. We conduct a SLR (Systematic Literature Review) of Machine Learning and Cloud Security methodology and strategies in this study. The results of the SLR were divided into three key study categories after we reviewed 63 relevant studies: (i) the many forms of Cloud security threats, (ii) the machine learning approaches employed, and (iii) the performance results. Eleven cloud security areas have been identified. Furthermore, with 16 percent and 14 percent usage, The most common Cloud security threats are distributed denial-of-service (DDoS) and data privacy topics. On the other hand, we discovered 30 ML approaches, Some of them were hybrids, while others were standalones. In both hybrid and standalone models, SVM is the most extensively used machine learning method. In addition, In order to illustrate the efficacy of their suggested model, 60 percent of the papers compared it to other models. There were also 13 other evaluation indicators specified.The rate of true positives is the highest widely used statistic, whereas training time is the least widely utilised. Finally, The most popular are KDD and KDD CUP'99widely utilised datasets research that are relevant, out of a total of 20.*

***Key Word**: Cloud security, machine learning, DDoS, privacy, and security are all topics that come up frequently.*

## I. INTRODUCTION

Cloud computing is a new type of technology advancement that provides information technology facilities, platforms, and software. In addition to Internet services It's said to signify the realisation of "Computing for Use," a long-held dream. Private, public, and hybrid Clouds are increasingly being used by businesses. Its main purpose is to enable users to pay for and use what they want, including on demand services. So that they can meet their software or infrastructure needs. Despite the fact that cloud computing is generally viewed as a significant and good IT infrastructure transformation, there is still a lot of work to be done in terms of security to offset its weaknesses. Because cloud data centres handle a vast amount of personal and corporate information, cloud security problems and vulnerabilities must be recognised and addressed. Cloud infrastructure may be vulnerable to assaults since it runs on conventional Internet protocols and employs virtualization techniques. Traditional sources of assault ARP (Address Resolution Protocol), IP spoofing, and Denial of Service (DoS) are only a few examples (DoS)attacks, among others. They could possibly originate from somewhere else. In the cyber security realm, zeroday attacks, also known as unknown assaults, are viewed as a major challenge. Traditional detection and prevention strategies are ineffective in dealing with these threats while simultaneously dealing with a large amount of data. Machine Learning (ML) techniques are particularly beneficial for detecting both traditional and zero-day assaults. ML mixes computer science and statistics to enhance prediction. In machine learning, there are three types of learning: supervised, unsupervised, and semi-supervised.

The classification model for supervised machine learning is built using categorised data that has been trained. Unsupervised learning techniques allow you to train a model without using any instructions. Naive Bayes, Decision Trees, Linear Regression, Support Vector Machines (SVM), and Nearest Neighbor, more methods are available for each. Unsupervised algorithms, such as K-means clustering, provide one example. Multi-layered computing models can learn data using Deep Learning (DL). representations at different levels of abstraction. It has made substantial progress in a variety of areas, including Image analysis, speech recognition, and text recognition are all examples of image analysis. The primary goal of this research is to perform a thorough examination of the machine learning approaches Cloud security risks and vulnerabilities are solved, detected, and prevented using this technology. Despite the vast number of research articles on Cloud security using machine learning, we are aware of only a handful Systematic Reviews in this field. The research papers were carefully gathered and chosenfor our analysis based on the following criteria: (I) the machine learning approaches utilised for Cloud security, (II) the security areas for which machine learning techniques are applied, and (III) the estimates and accuracy of the machine learning techniques employed.

The remainder of this study is divided into five sections: The review of the literatureis included in Section II. In Section III, the technique for carrying out this investigation is detailed. In Section IV, you'll find a list of the findings and outcomes. Section V discusses the review's limitations, while Section VI contains Discussion and research recommendations for the future.

## II. LITERATURE REVIEW

This section discusses the security and privacy issues that currently exist in Cloud computing. Cloud computing is a very wide field because it transmits and hosts its services through the Internet. It offers services that are tailored to the demands of its customers and charges appropriately. As a result, the Cloud becomes increasingly important as consumers come to rely on it, and businesses can now simply engage Cloud services.

Clients dread policies that are hidden from them, According to Khorshed and colleagues (Khorshedet al. ), Trust issues between customers and Cloud providers are referred to as gaps in Cloud computing. Cloud providers, on the other hand, are concerned that users would exploit their services and launch attacks utilising them. Vulnerabilities are described as Cloud security vulnerabilities that an adversary can exploit to gain network and other infrastructure resources access, according to Modi et al. The assaults that can be conducted must be discovered and understood in order to keep the Cloud safe from these dangers and avoid any harm. The following are the most commonly discussed attacks in Cloud computing:

**1.A DoS (Distributed Denial of Service) attack:** is an attempt to prevent a service from being available to its users. DDoS (Distributed Denial of Service) is a type of DoS assault that uses a large number of computers.

**2.Attack of the zombies:** when an attacker sends a flood of requests from unrelated hosts to the victim's network.As a result of this attack, Cloud's expected behaviour is disrupted, affecting Cloud service accessibility.

**3.Attack by a phishing website:** a scheme to deceive and steal personal information from unwary consumers by diverting them to a fake website An attacker may be using a Cloud service to hide the accounts and services of other Cloud users via a phishing attack site.

**Table No. 1:** Literature survey of papers

| **Sr No.** | Survey | Year | **Description** | Difference |
|---|---|---|---|---|
| 1. | A survey of security issues, threats, and solutions in cloud computing. | 2020 | This study addresses the myriad Cloud platforms have created security and privacy concerns. In addition, they provide a new classification of current security technologies in this field. | It discusses Cloud computing's security challenges and risks, as well as potential remedies. |
| 2. | Cloud computing security is the subject of a survey. | 2019 | This study examines the most serious Cloud computing is under threat. For comparison study, it also contained solutions and prospective countermeasures. | It discusses the threats to cloud security and how to defend against them. Machine learning methods are not covered. |
| 3. | Issues and concerns with cloud security | 2018 | This study examines Cloud computing deployment approaches, as well as concerns with service models. | It discusses the concerns and concerns associated with cloud security. |

| | | | | |
|---|---|---|---|---|
| 4. | A survey of network detection based on deep learning. | 2017 | This paper discusses deep learning approaches with an emphasis on network detection. | It discusses the use of deep learning to detect clouds. |
| 5. | A study of how supervised learning techniques can be used to detect cloud-based assaults. | 2016 | The architecture, types, hazards, and threats of cloud computing are all discussed in this survey. | Cloud computing and its security are the only things that matter. risks are discussed. |

In contrast to the previous reviews, We offer a comprehensive cloud security research project that employs machine learning techniques. Furthermore, there is no systematic literature review that covers the same topics that we are aware oftopics as ours. Furthermore, our research differs from the linked work in various ways in a table, for example:

1. Learn about machine learning methodologies, model types, and whether it's a hybrid or stand-alone model.
2. A precise evaluation of each technique's benefits and drawbacks.
3. A thorough examination of the challenges surrounding cloud security.
4. Showcase the highest levels of precision in terms of security.
5. From 2004 to 2019, it covers the most recent decade.

## III. METHODOLOGY

In this study, we used Kitchenham and Charters' methodology to carry out a thorough investigation. Their methods divide the procedure into numerous parts, each of which contains multiple stages. Organizing, carrying out, and finally reporting are the three processes.
The sections that follow show how this paper followed the review methodology.

### A.RESEARCH QUESTIONS

From 2004 to 2019, this SLR seeks to outline as well as clarifying Machine Learning (ML)approaches and implementations utilised in Cloud security. To that goal, the following three research questions (RQs) are posed:

**Q1: Which aspects of cloud security are covered in this review?**
→Identifies the Cloud security topics covered in the gathered papers, as well as the categories, number of studies, and if the articles were presented at a conference or published in a journal.

**Q2: What machine learning algorithms are employed in cloud computing security?**
→The type of machine learning model, the style of analysis, and the characteristics employed in the gathered papers are all addressed. This RQs examines the similarities and differences between the research.

**Q3: How accurate are machine learning models in terms of total estimation?**
→Focuses on four areas of estimating accuracy that are mentioned in the papers: the model validation procedures, the accuracy metric, the accuracy value, the construction data set, and the accuracy metric. It contrasts these features with those of the other publications.

### B.DATA EXTRACTION STRATEGY

The goal of The goal of this step is to create a semi-structured document. response to each article's research questions. Every article has the following information: paper number, title of paper, year of publication, type of publication, domain, Q1, Q2, and Q3. It's worth noting that not every paper addressed every research question.

### C. SYNTHESIS OF EXTRACTED DATA

We employ several ways to extract evidence to address the Qs in order to synthesise the data collected from the articles chosen. This explains the synthesis technique we used in detail:

A. Q1 and Q2: Organize the data you've gathered into a table based on what you've Q1 and Q2, the narrative synthesis approach is utilised.

B. Q3: To calculate the findings of Q3, we use binary outcomes for quantitative data extraction.

D.CLOUD SECURITY AREAS

Papers on the security side of anomaly detection in the cloud.

**Table No. 2:** Summary of Papers

| ID | Summary of the Paper on Detection |
|---|---|
| A1 | Creates a method for increasing the accuracy of cloud-based detection systems. |
| A2 | Proposes a method for detecting and classifying network traffic attacks. |
| A3 | Proposes a framework for detecting user activity and profiling it. |

## IV. LIMITATIONS OF THIS REVIEW

This project is limited to journal and conference publications on machine learning in cloud security. We were able to exclude a huge number of nonrelevant research papers by using our search approach strategy. Because the papers we chose perfectly meet our research goal, we only collected a small number of them. We also used quality evaluation criteria to choose papers that present synthesised results.

## V. CONCLUSION

To examine ML approaches utilised in Cloud security, we conducted a thorough literature study. The study looked into studies that answered three questions: cloud security, the sort of machine learning techniques employed, and the ML model's accuracy estimation. After applying our selection criteria, we ended up with 60 research papers. The following is a summary of our findings:

A. The 11 Cloud security areas revealed in Q1 findings are Anomaly detection, attack detection, privacy preservation, security, vulnerability detection, data confidentiality, data privacy, DDoS, DoS, and intrusion detection are all terms used to describe the detection of anomalies, attacks, and intrusions (ID). The most important issues are DDoS and data privacyoften studied topics, with 16 percent and 14 percent usage rates, respectively.

B. A total of 30 machine learning approaches were deployed in Q2, some as hybrids and others as standalones. In both hybrid and standalone models, SVM is the most extensively used machine learning method. Sixty percent of the studies compared their designs to those of others machine learning models in order to acquire the best rating and either verify or improve their model's correctness.

C. TPR, Accuracy, FPR Precision, TNR, FAR, Detection, F-measure, FNR, and Training Time C. TPR, Accuracy, FPR Precision, TNR, FAR, Detection, F-measure, FNR, and Training Time were among the 13 evaluation measures listed by A. Q3. TPR was the most popular measure, while Training time was the least popular. Furthermore, datasets have been utilised to assess the performance of models. The most popular are KDD and KDD CUP '99 popular of the 20 datasets discovered.

## REFERENCES

[1]. 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.

[2]. R. Kumar, S. P. Lal, and A. Sharma, ``Detecting denial of serviceattacks 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

[3]. Makrand M Jadhav, Gajanan H. Chavan, and Altaf O. Mulani, "Machine Learning based Autonomous Fire Combat Turret", Turkish Journal of Computer and Mathematics Education, Vol.12 No.2 (2021), 2372-2381, https://doi.org/10.17762/turcomat.v12i2.2025