Artificial Intelligence Review (2024) 57:132
https://doi.org/10.1007/s10462-024-10776-5

# Research trends in deep learning and machine learning for cloud computing security

**Yehia Ibrahim Alzoubi[1] · Alok Mishra[2] · Ahmet Ercan Topcu[3]**

Accepted: 24 April 2024 / Published online: 2 May 2024


**Abstract**
Deep learning and machine learning show effectiveness in identifying and addressing cloud security threats. Despite the large number of articles published in this field, there remains a dearth of comprehensive reviews that synthesize the techniques, trends, and challenges of using deep learning and machine learning for cloud computing security. Accordingly, this paper aims to provide the most updated statistics on the development and research in cloud computing security utilizing deep learning and machine learning. Up to the middle of December 2023, 4051 publications were identified after we searched the Scopus database. This paper highlights key trend solutions for cloud computing security utilizing machine learning and deep learning, such as anomaly detection, security automation, and emerging technology's role. However, challenges such as data privacy, scalability, and explainability, among others, are also identified as challenges of using machine learning and deep learning for cloud security. The findings of this paper reveal that deep learning and machine learning for cloud computing security are emerging research areas. Future research directions may include addressing these challenges when utilizing machine learning and deep learning for cloud security. Additionally, exploring the development of algorithms and techniques that comply with relevant laws and regulations is essential for effective implementation in this domain.

**Keywords** Cloud security · Deep learning · Machine learning · Cybersecurity · Trends

✉ Alok Mishra
alok.mishra@ntnu.no

Yehia Ibrahim Alzoubi
yehia.alzoubi@aum.edu.kw

Ahmet Ercan Topcu
ahmet.topcu@aum.edu.kw

[1] College of Business Administration, American University of the Middle East, Eqaila, Kuwait

[2] Faculty of Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

[3] College of Engineering and Technology, American University of the Middle East, Eqaila, Kuwait

## 1 Introduction

The SANS Institute found that 84% of effective cyberattacks exploit human behaviors. The SANS Institute is a collaborative research and educational institution that specializes in cybersecurity education and training (SANS 2024). SANS stands for "SysAdmin, Audit, Network, Security," reflecting its initial emphasis on system administrators and network security professionals. Additionally, a Ponemon Institute study found that the average cost of a data leak brought on by network failures is $4.1 million. The Ponemon Institute is a research group that focuses on information security policy, privacy, and data protection. They conduct independent research, surveys, and studies on topics such as data breaches, privacy practices, cybersecurity trends, and risk management (PonemonInstitute. 2024). Cybersecurity Ventures projects that the yearly cybercrime cost will exceed USD 7 trillion globally in 2022 and $10.5 trillion by 2025 (Morgan 2022). Cybersecurity Ventures is a leading researcher and publisher covering the global cybersecurity market. They provide industry insights, cybersecurity statistics, market forecasts, and trends through their reports, newsletters, and articles (CybercrimeMagazine 2024). Furthermore, research by the Information Systems Security Association found that using various data sources could improve the precision of vulnerability management identification by up to 50% (Oltsik 2023). The Information Systems Security Association is a nonprofit organization that provides educational forums, publications, and networking opportunities to cybersecurity professionals (ISSA 2024). The amount of sensitive information is anticipated to increase by 50% annually, according to a report by the International Association of Computer Science and Information Technology (IACSIT). Also, according to the same report, anomaly detection can identify up to 85% of breaches. Unfortunately, the same report claimed that adversarial cyberattacks can deceive Deep Learning (DL) models up to 90% of the time. The IACSIT is a professional organization that brings together researchers, scholars, and practitioners in the fields of computer science and information technology (IACSIT 2024). On the other hand, using security intelligence may decrease the time necessary to find a security breach by up to 50%, according to research by the SANS Institute. According to the same research, the SANS Institute found that finding a security event takes an average of 200 days (Spitzner 2022). These statistics provide a clear picture of the scope of the security issue and the role that Machine Learning (ML) and DL can play in addressing it in the future.

Contemporary malware presents a significant challenge for traditional detection systems due to its sophisticated and deceptive nature (Rao and Jain 2024). In cloud environments, antivirus programs often struggle to detect complex malware, such as encrypted or metamorphic variants, leading to an increased risk of undetected attacks (Abbas and Myeong 2023). Despite their widespread use, traditional security methods like firewalls and Intrusion Detection Systems (IDS) have limitations in cloud settings. They cannot effectively identify novel threats, zero-day attacks, or malicious mining programs, nor can they handle large volumes of data (Belal and Sundaram 2022). Consequently, there is a pressing need to ensure high detection rates with accuracy to reduce false positives and bolster security measures. One promising approach to addressing these challenges is adopting ML-based and DL-based IDS (Amiri et al. 2024). Powerful methods like ML and DL can be used to increase the security of cloud computing systems (Heidari et al. 2023). They have been demonstrated to be efficient in recognizing and preventing security threats and can be utilized for tasks including intrusion detection, anomaly detection, and classification. In order to find patterns and anomalies, categorize data, and generate predictions, these techniques

may analyze and learn from massive amounts of data. This can aid in spotting malicious activity, detecting intrusions, and identifying other security concerns.

However, while several literature reviews have explored ML and DL role in cloud computing security, none of these studies included bibliographic analyses. Additionally, most previous review papers on ML and DL usage for cloud computing security have focused either on different ML and DL techniques or algorithms, or on specific techniques such as Convolutional Neural Networks (CNN), or types of attacks like denial-of-service attacks, as elaborated in Sect. 2.4. Moreover, the field of ML and DL for cloud computing security is continuously evolving. Thus, this paper aims to fill this gap by offering a comprehensive review of the trends and challenges associated with using ML and DL for cloud computing security. This investigation is crucial for understanding such a dynamic field. Identifying trends, patterns, and connections between various research projects can also aid in synthesizing existing information. It can help identify potential biases and limitations in earlier research, thereby preventing the repetition of these errors and ensuring the study's thoroughness and validity (Herrera-Franco et al. 2020). Furthermore, conducting a bibliographic analysis helps in understanding previous research, identifying addressed problems, and pinpointing gaps in the literature that require further investigation.

The paper contributes to the field of cloud computing security by utilizing DL and ML in several ways. First, the paper addresses the need of comprehensive reviews by offering updated statistics on the development and research landscape of cloud computing security using DL and ML techniques. We conducted a bibliographical analysis before analyzing the field's recent and upcoming developments. Second, the paper highlights trending ML and DL solutions to address cloud security issues, including anomaly detection, security automation, cloud-native security, image-based detection, network traffic analysis, and the role of emerging technologies, providing insights into current approaches for addressing cloud security threats. Third, the paper identifies challenges including data privacy, scalability, explainability, generalizability, and label bias associated with using ML and DL for cloud security, bringing attention to areas that require further exploration and refinement. Fourth, the paper identifies several new algorithms and approaches being developed. Thus, both ML and DL are continuously growing in the context of cloud computing security. Finally, based on the findings, the paper draws future research directions focused on addressing challenges and developing algorithms and techniques that comply with relevant regulations and laws, emphasizing the importance of advancing the field in a responsible and compliant manner.

The remainder of this paper is structured as follows: The background of ML, DL, and cloud computing is covered in Section 2. The research methodology employed in this study is explained in Section 3. The bibliographical analysis of the relevant literature is discussed in Section 4. The current and future trends of deploying ML and DL for cloud computing security are discussed in Section 5. Section 6 concludes this paper. The definitions of the abbreviations used in this study are summarized in Table 1.

## 2 Background and related work

### 2.1 Cloud computing

Cloud computing is an information services delivery model that makes resources available to users through the Internet as needed and on a pay-as-you-go basis (Alzoubi

**Table 1** Table of abbreviations used in the paper

| Abbreviation | Definition | Abbreviation | Definition |
|---|---|---|---|
| CASB | Cloud Access Security Brokers | LSTM | Long Short-Term Memory |
| CNN | Convolutional Neural Networks | ML | Machine Learning |
| DL | Deep Learning | NLP | Natural Language Processing |
| DBN | Deep Belief Networks | PaaS | Platform as a Service |
| GAN | Generative Adversarial Networks | PCIDSS | Payment Card Industry Data Security Standard |
| GDPP | General Data Protection Regulation | ResNets | Deep Residual Networks |
| HMM | Hidden Markov Models | RF | Random Forest |
| HIPAA | Insurance Portability and Accountability Act | RNN | Recurrent Neural Networks |
| IaaS | Infrastructure as a Service | SaaS | Software as a Service |
| IDS | Intrusion Detection System | SVM | Support Vector Machine |
| KNN | K-Nearest Neighbor | VAE | Variational Autoencoders |

et al. 2022a). It makes it possible for users to acquire and use pooled computational resources, like storage, servers, and applications, without worrying about maintaining and managing those resources' infrastructure (Alzoubi et al. 2021). In cloud computing, there are three primary services: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) (Abdel-Basset et al. 2018). While SaaS offers consumers subscription-based access to software services like customer relationship management and email, PaaS provides a platform for users to design, execute, and administer applications. On the other hand, IaaS gives consumers access to hardware-related resources like CPUs, storage, and memory.

In addition, other deployment options are available with cloud computing, including private, public, and hybrid clouds. Third-party suppliers run public clouds, accessible to everyone, whereas a single enterprise runs private clouds. Enterprises may exploit the advantages of both private and public clouds by combining them in hybrid clouds (Abdel-Basset et al. 2018). Cloud computing services are offered by several major companies, such as (Gartner. 2023):

- Amazon Web Services: Amazon Web Services owned 32% of the cloud computing industry in 2021, making it the leading provider. It has a worldwide network of data centers and offers several services, including SaaS, PaaS, and IaaS.
- Microsoft Azure: With a 20% market share, Azure is the second-largest cloud computing service. It provides IaaS, PaaS, and SaaS services comparable to those of AWS and has a worldwide network of data centers.
- Google Cloud Platform: With a 7.7% share of the market, Google Cloud Platform is the third-largest cloud computing service. It offers a wide variety of services, such as SaaS, PaaS, and IaaS, and has a worldwide network of data centers.
- Alibaba Cloud: With a 4.6% share of the market, Alibaba Cloud is the fourth-largest provider of cloud computing services. It is a division of Alibaba Group that provides several services, including SaaS, PaaS, and IaaS, and a worldwide network of data centers. Still, it is mainly concentrated on the Asian market.

### 2.2 Machine learning and deep learning

With ML, a type of artificial intelligence, machines may automatically pick up new skills and improve over time (Mishra and Tyagi 2022). It entails employing algorithms to examine data, gain knowledge, and determine a forecast or course of action without involving humans (Lin et al. 2022). The size of the worldwide ML market is anticipated to increase from $8.41 billion in 2020 to $39.09 billion by 2025 at a CAGR of 36.5% throughout the forecast period (MarketsandMarkets. 2023). The three main types of ML are supervised learning, reinforcement learning, and unsupervised learning (Belal and Sundaram 2022; Gupta et al. 2017; Topcu et al. 2023).

- Supervised learning: Voice recognition, picture categorization, and Natural Language Processing (NLP) are examples of tasks that require supervised learning. Labeled data is used in supervised learning to train a model to forecast outcomes based on fresh data.
- Unsupervised learning: It is employed for activities like grouping and anomaly detection. It entails discovering structures or trends in unlabeled data without a clear prediction objective.
- Reinforcement learning: It is employed in activities like robotics and gaming. It entails preparing an agent to choose between incentives and punishments when making decisions.

DL is a subclass of ML that models and resolves complex issues, including decision-making, NLP, and voice and picture recognition, using artificial neural networks (Amiri et al. 2024; Heidari et al. 2022). These neural networks are composed of multiple layers, so they are called "deep" learning. The DL market size is anticipated to increase from $1.81 billion in 2020 to $10.95 billion by 2025 at a CAGR of 44.1% throughout the forecast period (MarketsandMarkets. 2023). Rather than relying on human feature extraction, DL algorithms can enhance the performance of ML models by automatically extracting features from raw data (Aldallal 1916). Since the data is complicated and unorganized, it is beneficial for voice and picture detection tasks.

### 2.3 Machine learning and deep learning for cloud computing security

As effective techniques for improving cloud computing security, ML and DL have come to light. Traditional methods frequently fail to handle the dynamic and complex nature of cyber threats in cloud security (Ge et al. 2021). In cloud environments, ML and its subset, DL, provide flexible and intelligent solutions to strengthen security and reduce risks. Massive volumes of data created in the cloud may be analyzed by ML algorithms, which can then find patterns and abnormalities that could point to security lapses. Proactive threat identification is made possible by ML models that utilize previous data to learn how to differentiate between harmful and regular activity (Gupta et al. 2017). This is furthered by DL, which uses neural network architecture to automatically extract complex aspects and representations from data, enabling more precise and nuanced threat detection.

One noteworthy usage is intrusion detection, where ML and DL algorithms are used to continually watch user behavior and network activity (Kasongo 2023). These systems offer real-time protection against possible attacks by triggering warnings or automatic replies in the event that abnormalities or suspicious patterns are discovered (Ma et al.

2023). Furthermore, cloud security systems may foresee and proactively handle new vulnerabilities using ML and DL approaches in predictive analysis (Kumar et al. 2022a). ML and DL technologies are increasingly essential in strengthening the resilience of cloud infrastructure against various cyber threats as cloud computing advances. The following explains the critical ML and DL milestones for cloud security (Belal and Sundaram 2022; Yıldız 2024; Dasgupta et al. 2022; Saran et al. 2022).

- The cloud computing sector grew significantly after 2010, coinciding with the understanding that more advanced security measures were required. Security systems for virus analysis, anomaly detection, and user behavior analytics started to use ML algorithms. The goal was to improve the detection and reaction to threats capabilities by utilizing ML.
- The introduction of DL, especially with the rise in popularity of neural networks and deep neural architectures, was a significant turning point in 2012. DL methods, such as Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNNs), have a remarkable ability to process intricate data structures.
- Concerning cloud security, DL has made significant progress in pattern and picture recognition, increasing threat detection accuracy. After 2015, as cloud environments advanced in sophistication, so did cyber threats. ML and DL played key roles when creating IDS for cloud-based infrastructures. These systems used ML algorithms to continuously monitor user activity and network traffic to spot behavioral anomalies that could be signs of impending security problems.
- Predictive analytics has received more attention in cloud security since 2019. These days, historical data analysis, trend identification, and possible security risk prediction are all done using ML and DL models. Moreover, real-time reactions to security issues are made possible by integrating automation with ML and DL, which shortens the time between detection and mitigation. One of the innovations is the creation of adversarial ML, which improves resistance to complex assaults. There are still issues to be resolved, such as the requirement for interpretability in intricate DL models and handling moral questions related to security and artificial intelligence.

### 2.4 Related survey literature

To contextualize our study, we surveyed pertinent review articles regarding the utilization of ML and DL in cloud security. We specifically focused on papers published between 2022 and 2024, examining titles and abstracts in the Scopus database to select comprehensive literature reviews. Papers exclusively addressing singular techniques (e.g., (Ravinder and Kulkarni 2023) and (Pandey et al. 2023)) or specific attacks (e.g., (Nair et al. 2023) and (Gupta et al. 2022)) were excluded. It was essential to focus only on articles discussing ML and DL techniques for cloud security purposes, in general. Table 2 outlines the key insights gleaned from these reviews, revealing three overarching themes: the application, challenges, and emerging trends of ML and DL in cloud security.

- Application of ML and DL for cloud computing security: The literature provides insights into the application of ML and DL techniques for enhancing cybersecurity across various domains. In (Abdullahi et al. 2022), the authors categorized AI techniques for IoT cybersecurity, highlighting SVMs and RF as popular choices. The authors in Belal and Sundaram (2022) emphasized SVM, RF, and KNN for cloud

**Table 2** Related recent literature

| Study | Focus | ML & DL security trends |
|---|---|---|
| Abdullahi et al. 2022) | Detecting cybersecurity attacks in the IoT using AI methods | • RF and Support Vector Machine (SVM) are the most used methods<br>• RNN, Neural Networks, and Extreme Gradient Boosting outperform other methods |
| Belal and Sundaram 2022) | ML and DL for security threat detection in cloud computing | • SVM is the most used method |
| Dasgupta et al. 2022) | ML in cybersecurity | • CNN and RNN are promising techniques<br>• Transfer learning is another promising technique |
| Kornaros 2022) | ML and for IoT device security | • ML and DL are promising for IoT device security, particularly in intrusion and anomaly detection<br>• ML scalable as it allows a system to evolve to become more efficient and effective over time |
| Aldhaheri et al. 2023) | DL for cyber threat detection in IoT | • Trade-off between the advanced structure of DL techniques and the complexity of the structure<br>• DL-based intrusion detection is still a popular field of future research |
| Bhuvaneshwari and Kaythry 2023) | DL strategies to enhance cybersecurity | • CNN, Deep Reinforcement Learning, and RNN, among other DL techniques, enhance cybersecurity |
| Hernandez-Jaimes et al. 2023) | AI as an intrusion detection for internet of medical things threats | • Several AI techniques were reported, including unsupervised, semi-supervised, and supervised ML/DL, online learning, adaptive incremental classifiers, and ML classifiers |
| Roy et al. 2023) | ML-based security systems for edge devices | • ML produces different outcomes with different datasets<br>• Extreme Gradient Boosting performs well in a huge number of observation datasets<br>• Random Forest (RF) is an exceptional classifier |
| Salem and Al-Saedi 2023) | DL and data mining for enhancing cloud computing security | • Several techniques were evaluated, including CNN, RNN, Deep Belief Networks (DBN), SVM, and clustering algorithms<br>• The performance of these techniques requires further improvement |
| Sarker et al. 2023) | ML for IoT security | • Many ML and DL techniques were argued to be suitable for IoT security<br>• ML and DL modeling should be included in a successful IoT security framework |
| Vinolia et al. 2023) | ML and DL for cloud computing IDS | • Unsupervised DL outperformed other ML and DL techniques |

**Table 2** (continued)

| Study | Focus | ML & DL security trends |
|---|---|---|
| Yi et al. 2023) | DL application in network threat detection | • Evaluated several models, including Self-Autoencoder, RNN, DBN, and Constrained Boltzmann Machine models |
| Balobaid et al. 2024) | ML and DL for IoT security | • ML and DL techniques like the K-Nearest Neighbor (KNN) method and Deep Stack Encoder Neural Networks can improve IoT security |
| Mishra 2024) | ML and DL for cloud computing security | • Anomaly detection, threat intelligence, and analytics of user and entity behavior<br>• Network security, data security, and endpoint security<br>• Identity and access management, incident response, security information and event management, continuous monitoring, and adaptation |
| This review | | • Bibliometric analysis of all papers on ML and DL for cloud security within the Scopus database<br>• Latest trends in this domain: anomaly detection, security automation, cloud-native security, emerging technology's role, insider threats, Cloud Access Security Brokers (CASB), image-based malware detection, and network traffic analysis<br>• Most prevalent ML and DL methods utilized in cloud security and the associated challenges<br>• The practical implications of these findings for industry applications |

security. The authors in Hernandez-Jaimes et al. (2023) analyzed AI methods for securing medical IoT devices, and the authors in Roy et al. (2023) evaluated ML methods for edge device security, focusing on Extreme Gradient Boosting and RF. Similarly, the authors of Bhuvaneshwari and Kaythry (2023) highlighted DL techniques like RNN and CNN for cybersecurity. Furthermore, the authors in Salem and Al-Saedi (2023) discussed DL and data mining in cloud security, emphasizing CNN and SVM.

- Challenges and trade-offs of using ML and DL for cloud computing security: In (Kornaros 2022), the authors examined ML and DL techniques for IoT device security, addressing integration trade-offs. The authors of Aldhaheri et al. (2023) discussed DL's advanced structure and trade-offs, proposing edge computing and transfer learning as solutions. Moreover, the authors of Yi et al. (2023) examined DL's role in network threat detection, categorizing models like Self-Autoencoder and RNN. Additionally, the authors of Vinolia et al. 2023 assessed DL-based unsupervised techniques for intrusion detection.
- Emerging trends and future directions for using ML and DL for cloud computing security: The authors of Sarker et al. (2023) explored ML and DL methodologies for IoT security, covering various classification and clustering methods. Also, authors in Dasgupta et al. (2022) identified DL methods like RNN for network traffic analysis. Moreover, the authors of Balobaid et al. (2024) discussed ML and DL techniques for IoT-based cloud computing security, highlighting the importance of updated security measures and methods like KNN and Deep Stack Encoder Neural Networks.

Finally, the authors in Mishra (2024) discussed several uses of ML and DL that were documented. This included anomaly detection, threat intelligence, analytics of user and entity behavior, network security, data security, endpoint security, identity and access management, incident response, security information and event management, continuous monitoring, and adaptation. This study is still under review but was added as a preprint to the SSRN database. Our paper offers several advancements compared to recent literature in the field. It conducts a comprehensive bibliometric analysis of all papers on ML and DL for cloud security within the Scopus database. Additionally, it delineates the latest trends in this domain, covering topics such as anomaly detection, security automation, cloud-native security, emerging technology's role, insider threats, CASB, image-based malware detection, and network traffic analysis. These trends are examined individually for both ML and DL techniques. Furthermore, the paper identifies the most prevalent ML and DL methods utilized in cloud security and addresses the associated challenges. Finally, it discusses the practical implications of these findings for industry applications.

## 3 Research method

A systematic review is highly relevant for investigations since it delves into the topic of study and assists the investigator in creating research questions that will expand their knowledge base. In a similar vein, bibliometric investigation features a strict and clear methodology that enables the selection and assessment of the literature using a transparent, repeatable process that allows understanding of a field of study (Herrera-Franco et al. 2020).

### 3.1 Study design

The bibliometric technique was used as our primary research method, with current and upcoming trends as secondary. The research covered the publications until the middle of December 2023. The study's approach was predicated on the notion that employing DL and ML for cloud security has garnered considerable scientific interest. The bibliometric investigation, centered on analyzing the literature of published articles assembled in a database, is considered a branch of science since it displays a thorough map of the knowledge structure and its assessment. This type of analysis enables a detailed investigation of the development of the phenomenon over time in a specific academic topic (Khudzari et al. 2018).

The sources or repositories' identities must be accurate and trustworthy. The Scopus database was selected primarily for its high-quality standards, broad coverage, simplicity of fetching data, and the most comprehensive library of computer and engineering journals (Herrera-Franco et al. 2020). Moreover, the Scopus database enables bibliography analysis through operational features like document type, source name, author names and affiliations, publication year, h-index performance measures for documents, and the number of citations, among others (Sweileh 2018).

### 3.2 Search strategy

The choice of search terms is crucial in the bibliometric analysis since they significantly affect the results. The search technique was based on title or abstract searches to reduce false-positive findings. Search terms were collected from recently reported ML and DL literature. The search approach was not language- or document-type-specific (i.e., article, conference proceedings, or book chapter) (Sweileh 2018). To search for the Scopus document, we employed the following Boolean operators: "machine learning" OR "deep learning" AND "cloud" OR "fog" OR "edge" AND "security" AND "attack" OR "protection" OR "detection" OR "response" OR "prediction." By including these keywords, the search is tailored to capture publications that discuss the utilization of ML and DL in cloud security, as well as related topics such as anomaly detection, security automation, and emerging technologies. This focused approach ensures that the retrieved publications are more likely to provide valuable insights into the trends, solutions, and challenges associated with ML and DL in cloud computing security, which is the primary focus of the study. Additionally, using specific keywords helps narrow down the search results and improves the relevance of the retrieved publications to the research topic.

The acquired data were checked for accuracy and looked at for their relevance to the study question. As a result, 2820 (i.e., 2438 ML and 1613 DL) publications were identified. We were able to count the number of ML and DL papers in cloud security by using built-in Scopus search features like "year," "author name," "subject area," and other fields (Sweileh 2018). The findings of this paper focused only on the highest 15 publications regarding the bibliographic characteristics (i.e., highest citations, affiliations, and so on). 15 was randomly chosen as the criterion to list the findings; this option has been employed in other bibliometric studies (Sweileh 2018).

We utilized the advanced search interface provided by Scopus, enabling us to specify search terms, publication types, date ranges, and other filters. Inclusion criteria likely specified that publications must pertain to cloud computing security and involve ML or DL

techniques. Additionally, we restricted our selection to peer-reviewed articles, book sections, or conference papers, and only articles written in English were considered during the analysis stage due to the authors' proficiency in English. Publications needing clear relevance to cloud computing, ML, DL, and security were excluded.

Following the retrieval of the initial set of search results, manual screening processes were employed to refine the selection further. This involved reviewing the publications' titles, abstracts, and keywords to assess their relevance to the topic of interest. Publications failing to meet the inclusion criteria were excluded at this stage. Subsequently, data extraction was carried out to collect pertinent information from each publication, including publication title, authors, publication year, journal/conference, abstract, and keywords. This data was compiled and analyzed to generate the statistics presented in the paper.

### 3.3 Thematic analysis

We employed thematic analysis methodologies (Alzoubi and Gill 2021) to distill the ML and DL trends for cloud computing security and delineate associated challenges, as expounded in Sect. 5. Thematic analysis served as the methodological approach for dissecting the gathered data, aiming to unearth commonalities and significant insights prevalent across the selected articles. Following the structured guidelines delineated by Miles and Huberman (1994), the analysis unfolded in four distinct phases: data collection, data reduction, data display, and conclusion drawing. During the data reduction phase, the collected information underwent meticulous preparation and familiarization to render it amenable for analysis (Miles and Huberman 1994). Preliminary coding, crucial for transforming raw data into a manageable format, employed keywords extracted from Scopus search results as a foundational framework (Alzoubi and Gill 2021). Subsequently, major themes emerged as the data were scrutinized to unveil pertinent issues, concepts, and challenges pertaining to the objectives of the study, encompassing trends, techniques, and hurdles encountered in the realms of ML and DL within cloud computing security. Coded data were systematically grouped into cohesive categories, fostering thematic exploration and understanding.

Data display, a pivotal facet of the analytical process, involves organizing and synthesizing information into a more coherent and accessible format, facilitating comprehension, and facilitating the transition to subsequent analysis stages (Alzoubi and Gill 2021). This stage was recurrent throughout the analysis, aligning with the progress made and laying the groundwork for subsequent analytical endeavors. Finally, the synthesis of tables and charts marked the culmination of the analysis, synthesizing the myriad data points into a cohesive and meaningful narrative that encapsulated the identified ML and DL trends within cloud computing security (Miles and Huberman 1994). This concluding phase unfolded concurrently with other analytical stages, underscoring its integral role in shaping the coherent portrayal of the study's focal points.

## 4 Bibliographic findings

The vast majority of the documents identified were written in English, with 2420 (99.2%) documents on ML and 1586 (98.3%) documents on DL. A small number of documents were written in Chinese, with 27 (1%) on ML and 27 (1.7%) on DL, while only one ML document was written in Turkish and one in Portuguese. Also, only one DL document was written in Turkish, and the other one was written in Korean. As

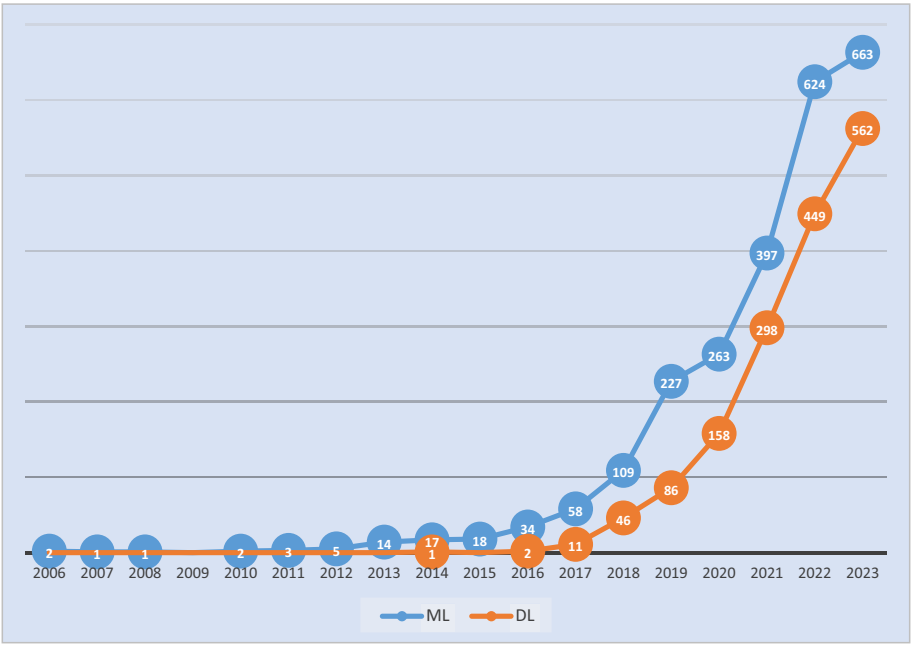


**Fig. 1** Published document number until middle of December 2023

shown in Fig. 1, over the past five or six years, there has been significant growth in both the volume of ML and DL articles. While ML articles first appeared in 2006, DL literature began in 2014. By the end of 2023, we anticipate that over 600 articles will have been published if the current trend continues. As shown in Table 3, the documents were published in nine different categories: articles (1715, or 42.3%), conferences (1515, or 37.4%), review conferences (503, or 12.4%), book chapters (145, or 3.5%), reviews articles (79, or 2.8%), and books (39, or 1.0%).

**Table 3** Published document channel

| Year | Number (ML) | Number (DL) |
|---|---|---|
| Conference | 1003 | 512 |
| Article | 937 | 778 |
| Review conference | 279 | 224 |
| Book chapter | 102 | 43 |
| Review article | 79 | 36 |
| Book | 25 | 14 |
| Erratum | 5 | 3 |
| Retraced | 4 | 1 |
| Editorial | 2 | 1 |
| Short survey | 1 | 0 |
| Data paper | 1 | 1 |
| Total number | 2438 | 1613 |

### 4.1 Top 15 search keywords used

Figure 2 depicts the visualization of the 15 most commonly used keywords. The term "machine learning" was the most frequently utilized keyword in ML, appearing 1212 times, while "computer crime" was the least frequently used, with only 247 appearances. The ML keywords were classified into four broad categories: learning, detection, security, and computing. The first category, learning, included keywords such as "machine learning," "machine-learning," "learning systems," "learning algorithms", "artificial intelligence", and "deep learning," which had a combined total of 3424 occurrences. The second category, computing, included "edge computing", "Internet of things", "cloud-computing", and "cloud computing", with a total of 1825 occurrences. The third category, security, consisted of "network security," "computer crime," and "security," with a total of 1454 occurrences. The fourth and final category, detection, comprised "intrusion detection" and "intrusion detection systems" with 744 occurrences.

On the other hand, in the field of DL, the most frequently used keyword was "deep learning" with 1170 appearances, while "cybersecurity" was the least frequently used with

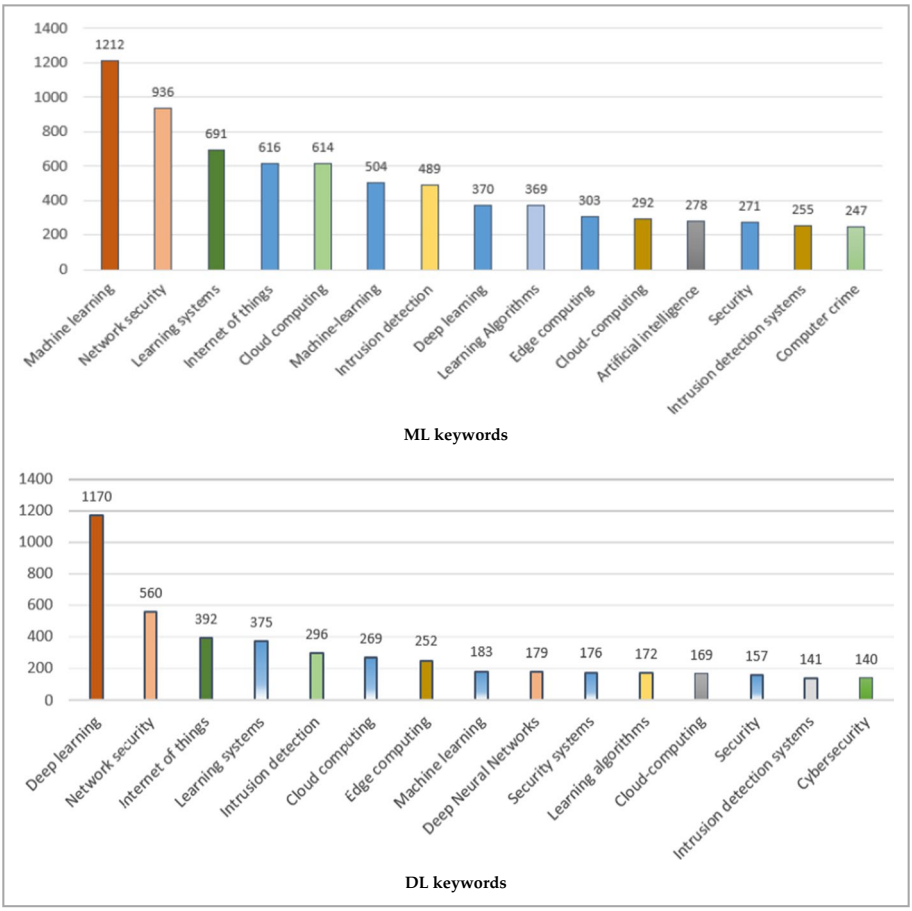


**Fig. 2** Top 15 used keywords

only 140 appearances. Like ML, DL keywords could also be broadly classified into four categories: learning, detection, security, and computing. The learning category included keywords such as "deep learning," "learning systems," "learning algorithms," "machine learning", and "deep neural networks", with a combined total of 2079 occurrences. The second category, the computing category, included "edge computing", "Internet of things", "cloud computing", and "cloud-computing", with a total of 1082 occurrences. The third category, security, consisted of "network security", "computer crime", "cybersecurity", and "security", with a total of 1033 occurrences. Finally, the detection was comprised of "intrusion detection" and "intrusion detection systems" with a total of 437 occurrences.

### 4.2 Top 15 research domains

Table 4 summarizes the top 15 research domains of the published article identified in the study. Out of the total 2438 documents, 2137 (87.6%) of the ML documents were categorized under computer science. In contrast, only 11 (0.56%) belonged to medicine, and biological sciences and other fields, including neuroscience, health, multidisciplinary, economics, arts, psychology, and immunology, accounted for only 87 (3.5%) ML documents. Similarly, out of the 1613 DL documents, 1427 (88.4%) were categorized under computer science, while only 17 (1.0%) belonged to the field of environmental science, and the remaining 44 (2.7%) were classified under other fields, including chemical engineering, multidisciplinary, health, nursing, and pharmacology.

### 4.3 Top 15 active countries

Figure 3 depicts the visualization of the 15 most active countries. The study highlights the global nature of ML and DL research for cloud security, with researchers from many countries contributing to the field. Authors from 98 different countries have contributed to the field of ML and DL research for cloud security. The top 15 active countries published 89.6% (2185) of ML documents and 98.2% (1584) of DL documents. China, India, and the USA are the leading nations in both ML and DL research. Regarding ML research, with 26.6% (649), India had the most publications, followed by China and the USA with 15.7% (348) each. The countries with the least number of ML publications were France, with 1.7% (42), and Taiwan, with 1.68% (41) studies of the total. Of the top 15 active countries for ML research, six were in Asia, four in Europe, two in the Middle East, and two in North America. In DL research, China led with 24.4% (394) of the published documents, followed by India with 23.8% (384), and the USA with 10.4% (169). Of the top 15 active countries for DL research, seven were in Asia, three in the Middle East, two in Europe, and two in North America.

### 4.4 Top 15 active institutions

Figure 4 depicts the visualization of the 15 most active institutions. ML and DL research for cloud security has received contributions from 98 countries and 149 institutions worldwide. In ML, *Vellore Institute of Technology* is the most active organization in this subject in India, accounting for 1.1% (27 articles) of the documents. *SRM Institute of Science and Technology* (India), *Chinese Academy of Science* (China), and *Qatar University* (Qatar) followed closely behind, each contributing 1% (18 articles) of the documents. Among the

**Table 4** Top 15 research domains

| Technique | Publication type | Number | Technique | Publication type | Number |
|---|---|---|---|---|---|
| ML | Computer science | 2137 | DL | Computer science | 1427 |
| | Engineering | 1210 | | Engineering | 822 |
| | Mathematics | 485 | | Mathematics | 311 |
| | Decision sciences | 334 | | Decision sciences | 199 |
| | Physics and astronomy | 209 | | Physics and astronomy | 144 |
| | Materials science | 154 | | Material science | 131 |
| | Energy | 120 | | Energy | 77 |
| | Social science | 107 | | Social sciences | 49 |
| | Biochemistry | 68 | | Medicine | 41 |
| | Chemistry | 56 | | Biochemistry, genetics and molecular biology | 35 |
| | Business, management and accounting | 52 | | Chemistry | 34 |
| | Environment science | 39 | | Business, management and accounting | 33 |
| | Chemical engineering | 25 | | Neuroscience | 21 |
| | Agricultural and biological sciences | 24 | | Earth and Planetary Sciences | 18 |
| | Medicine | 11 | | Environmental Science | 17 |
| | Others (Neuroscience, health, multidisciplinary, economics, arts, nursing, pharmacology, psychology, immunology) | 87 | | Others (chemical engineering, multidisciplinary, health, nursing, pharmacology) | 44 |

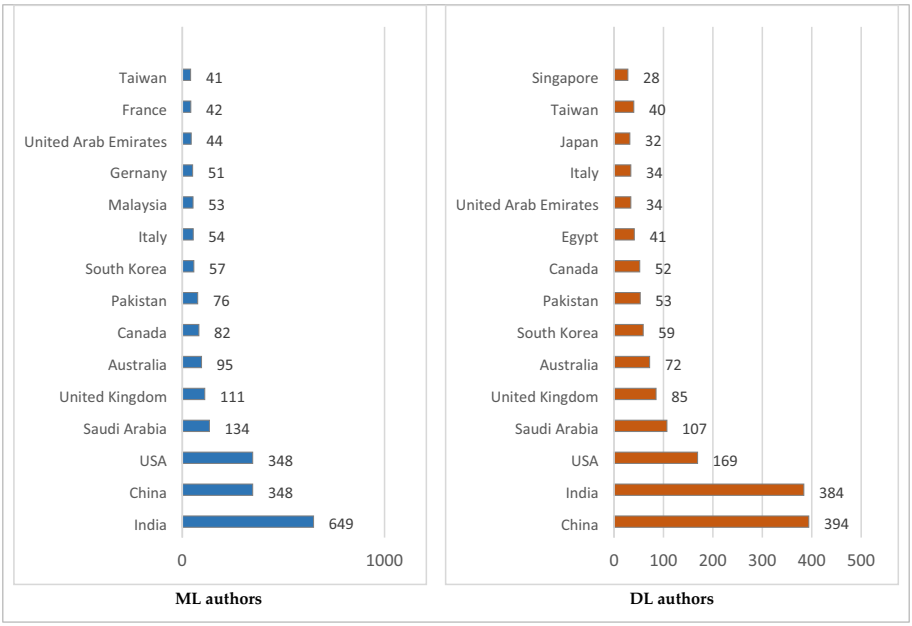


**Fig. 3** Top 15 active countries

top 15 active institutions, five were in China, four in Saudi Arabia, three in India, and one in Australia, Singapore, and Qatar. These statistics demonstrate the significant global presence of research in this field. In DL literature, the *Chinese Academy of Sciences* in China was the most active institution, contributing 1.8% (30 articles) of the documents. The *University of Electronic Science and Technology of China* (China) published 1.36% (23 articles), and Nanyang Technological University (Singapore) published 1.3% (21 articles). Among the top 15 active institutions, eight were in China, two in Saudi Arabia and India, and one in Singapore and Qatar. These figures highlight the substantial contributions of Chinese and Indian institutions to DL research.

### 4.5 Top 15 active authors

Figure 5 displays the authors who are most active in the fields of ML and DL regarding cloud security. 175 authors authored the ML-retrieved documents, with 224 papers needing to be specified. Of these documents, roughly 36 were written by a single individual, 120 by two authors, and the remainder by at least three, except those with undefined authors. The most active authors are Professors Azidine Guezzaz and Said Benkirane from Université Cadi Ayyad in Morocco, Professor Mourade Azrour from Université Moulay Ismaïl in Morocco, and Professor Mohsen Mokhtar Guizani from Mohamed Bin Zayed University of Artificial Intelligence in the United Arab Emirates, who published 10 (0.4%) of the documents.

On the other hand, 162 authors contributed to publishing the documents retrieved by DL, with 224 authors not specified. Of these documents, 11 were authored by a single individual, 70 by two authors, and the rest by at least three, except those with undefined

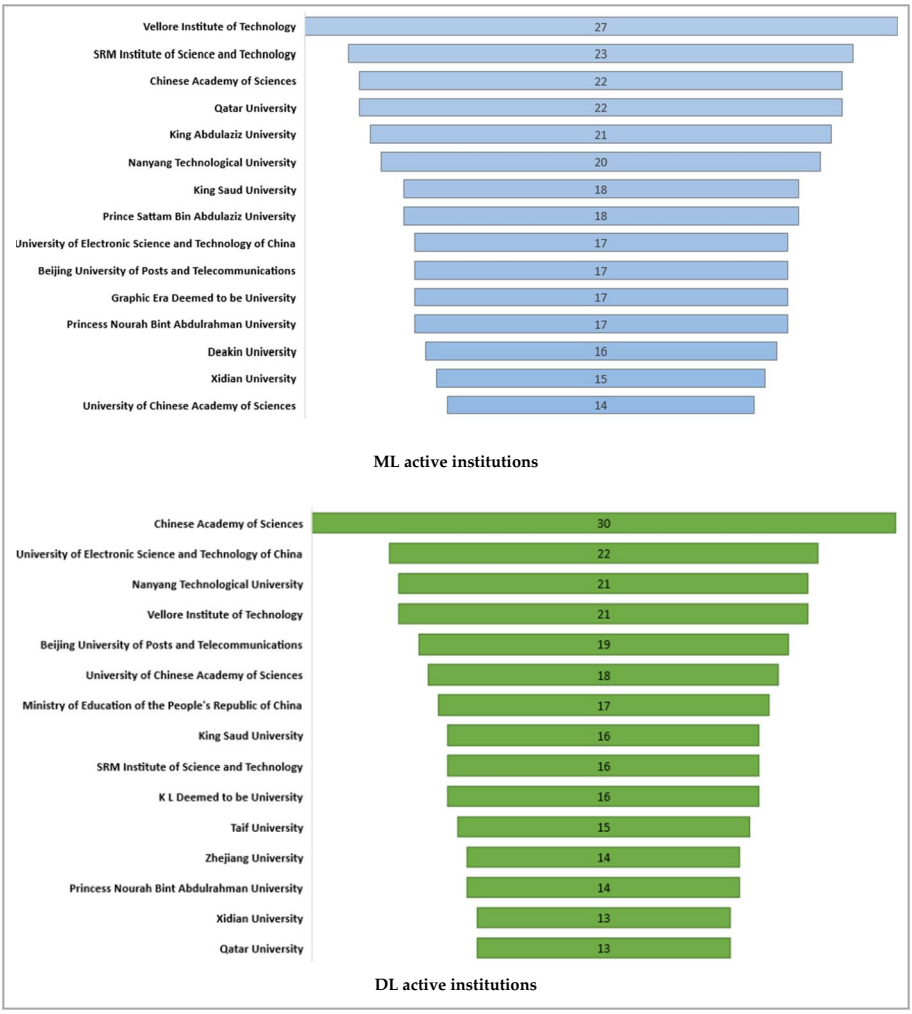


**Fig. 4** Top 15 active institutions

authors. The most active authors were Professor Mohsen Guizani from Qatar University and Professor Prabhat Kumar from LUT University in Finland, who published eight (0.5%) documents each. Professor Jong Hyuk Park from Seoul National University of Science and Technology in South Korea and Professor Mohamed Amine Ferrag from Guelma University in Algeria tied for second place, having published seven (0.4%) documents each.

### 4.6 Top 15 publication channels

Table 5 depicts the 15 most preferred publication channels. The ML documents that were retrieved were published on a total of 161 different channels. Among these, the majority were conference articles, with 1,003 (41.1%) publications, followed by journal articles, with 937 (38.4%) proceedings and workshops. The most active publication channel in this

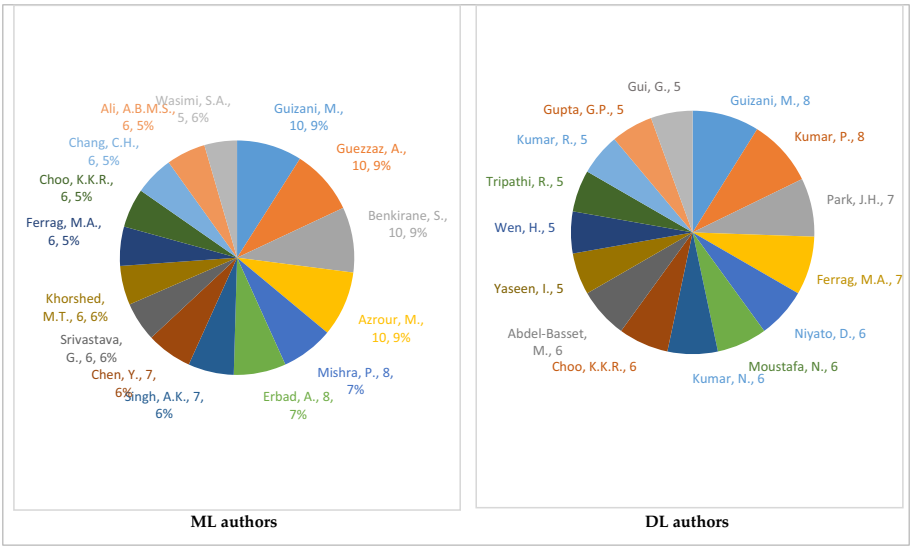


**Fig. 5** Top 15 active authors

field was the Lecture Notes in Computer Science, including the *Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics* series, which accounted for 97 (3.9%) documents. The *Lecture Notes in Networks and Systems* series, also published by Springer, had 94 (3.8%) documents. Among individual scientific journals, *IEEE Access* had the highest number of publications with 84 (3.4%) documents, followed by *Communications in Computer and Information Science Journal* with 63 (2.6%) documents.

The DL documents that were retrieved were published on a total of 163 different channels. Journal articles accounted for the majority of publications, with 778 (48.2%), followed by conference proceedings, with 512 (31.7%) proceedings and workshops. The most active publication channel in this field was *IEEE Access* with 72 (4.4%) documents, followed closely by the *Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics* series, with 61 (3.7%) documents.

### 4.7 Top 15 cited document

The retrieved ML documents received more than 14,000 citations, averaging about 5.7 citations per document. Similarly, the retrieved DL documents received 15,000 citations, averaging 9.6 citations per document. Table 6 provides an overview of the 15 most cited documents published since 2019. The ML document by Hassija et al. (Hassija et al. 2019) (published in *IEEE Access* in 2019) received the highest number of citations at 867, while the second highest cited ML document was by Mothukuri et al. (Mothukuri et al. 2021a) (published by *Future Generation Computer Systems* in 2021) with 545 citations. The highest-cited DL document (Tuli et al. 2020), with 413 citations, was published by *Future Generation Computer Systems* in 2020, and the second highest-cited DL document, cited 296 times, was authored by Priya et al. (Priya et al. 2020) (published by *Computer Communications* in 2020).

**Table 5** Top 15 publication channels

| ML | | DL | |
|---|---|---|---|
| Journal | Number | Journal | Number |
| Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics | 97 | IEEE Access | 72 |
| Lecture Notes in Networks and Systems | 94 | Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics | 61 |
| IEEE Access | 84 | Lecture Notes in Networks and Systems | 57 |
| Communications in Computer and Information Science | 63 | IEEE Internet of Things Journal | 45 |
| Advances in Intelligent Systems and Computing | 43 | Communications in Computer and Information Science | 38 |
| ACM International Conference Proceeding Series | 38 | Lecture Notes in Electrical Engineering | 33 |
| Sensors | 36 | Sensors | 23 |
| Lecture Notes in Electrical Engineering | 34 | Electronics | 22 |
| IEEE Internet of Things Journal | 33 | ACM International Conference Proceeding Series | 22 |
| Electronics | 23 | IEEE Transactions on Industrial Informatics | 20 |
| International Journal of Advanced Computer Science and Applications | 22 | Advances in Intelligent Systems and Computing | 19 |
| Smart Innovation Systems and Technologies | 17 | Lecture Notes of the Institute for Computer Sciences Social Informatics and Telecommunications Engineering LnIcst | 15 |
| Procedia Computer Science | 17 | IEEE Transactions on Intelligent Transportation Systems | 14 |
| Lecture Notes on Data Engineering and Communications Technologies | 17 | Computers Materials and Continua | 14 |
| Lecture Notes of The Institute for Computer Sciences Social Informatics and Telecommunications Engineering Lnicst | 17 | Proceedings Of SPIE The International Society for Optical Engineering | 13 |

**Table 6** Top 15 cited documents

| ML | | DL | |
|---|---|---|---|
| Study | Citation number | Study | Citation number |
| Hassija et al. 2019) | 867 | Tuli et al. 2020) | 413 |
| Mothukuri et al. 2021a) | 545 | Priya et al. 2020) | 296 |
| Gu et al. 2019) | 446 | Tian et al. 2020) | 239 |
| Goh et al. 2021) | 300 | Hossain et al. 2020) | 237 |
| Priya et al. 2020) | 296 | Chen et al. 2020) | 227 |
| Aloqaily et al. 2019) | 296 | Moustafa et al. 2019) | 224 |
| Gupta et al. 2020) | 263 | Mahdavifar and Ghorbani 2019) | 201 |
| Mothukuri et al. 2021b) | 212 | Alkadi et al. 2020) | 195 |
| Alrashdi et al. 2019) | 202 | Rathore et al. 2019) | 189 |
| Kumar et al. 2021) | 190 | Dai et al. 2020) | 173 |
| Moustafa 2021) | 183 | Yu et al. 2021) | 172 |
| Ravi and Shalinie 2020) | 176 | Parra et al. 2020) | 165 |
| Angelopoulos et al. 2020) | 176 | Ferrag et al. 2022) | 162 |
| Jia et al. 2020) | 171 | Zhou et al. 2021) | 158 |
| Saranya et al. 2020) | 171 | Bhattacharya et al. 2019) | 150 |
| Dovom et al. 2019) | 156 | Thakkar and Lohiya 2021) | 144 |

## 5 Trends of machine learning and deep learning for cloud security

We should analyze the majority of the identified documents on Scopus in order to give a thorough picture of present and future trends in ML and DL for cloud security. However, we used systematic sampling due to the large number of papers (4051) (Mostafa and Ahmad 2018). Every $k^{th}$ case in the population is chosen for the sample using the systematic sampling technique, where k equals the size of the population divided by the size of the sample. If the population is arranged in a certain way, such as by year or alphabetically, like in the case of this study, this strategy works well (Mostafa and Ahmad 2018). Systematic sampling has the advantages of being reasonably simple to apply and lowering the likelihood of sample bias since every $k^{th}$ member is chosen, guaranteeing a reasonably equitable representation of the population (Alzoubi et al. 2022a). The following two considerations were applied in the systematic sampling:

1. All the top 15 cited articles (reported in Table 5) were included in the review for both ML and DL.
2. To select the five studies for each of the past five years (2019–2023), as shown in Table 7, we employed a systematic approach. Given the substantial increase in research on cloud security utilizing ML and DL since 2019, we aimed to ensure comprehensive coverage of relevant studies. Initially, we conducted a systematic sampling process based on the articles deemed most relevant to our topic. This relevance was determined by carefully examining the titles and keywords of the 4051 articles retrieved from our search in the Scopus database, excluding the top 15 cited articles. This exclusion was to ensure that our selection process remained focused on identifying lesser-known but potentially valuable contributions to the field. By using this method, we aimed to provide

**Table 7** The selected papers (2019–2023)

| Technique | Year | Document |
|---|---|---|
| ML | 2019 | Dey et al. 2019; Rahila and Khonde 2019; Nissim et al. 2019; Balamurugan and Saravanan 2019; Yao et al. 2019) |
| | 2020 | Butt et al. 2020; Chkirbene et al. 2020; Mishra et al. 2020; Sugi and Ratna 2020; Jaber and Rehman 2020) |
| | 2021 | Ghazal et al. 2021; Rjoub et al. 2021; Tian et al. 2021; Rosero et al. 2021; Hameed et al. 2021) |
| | 2022 | Belal and Sundaram 2022; Dasgupta et al. 2022; Ullah et al. 2022; Lei et al. 2022; Prasad and Bharathi 2022; Brown et al. 2022) |
| | 2023 | Kasongo 2023; Nair et al. 2023; Dalal et al. 2023; Abosata et al. 2023; Zhou et al. 2023) |
| DL | 2019 | Yao et al. 2019; Alasmary et al. 2019; Abusitta et al. 2019; Krishnan et al. 2019; Bhattacherjee et al. 2019) |
| | 2020 | Gu et al. 2019; Wu et al. 2020; Ramchandran and Sangaiah 2020; Liu et al. 2020; Ferrag et al. 2020) |
| | 2021 | Ge et al. 2021; Tian et al. 2021; Thilagam and Aruna 2021; Makkar et al. 2021; Landman and Nissim 2021) |
| | 2022 | Gupta et al. 2017; Aldallal 1916; Kumar et al. 2022a; Zhang et al. 2022; Abou El Houda et al. 2022) |
| | 2023 | Kasongo 2023; Pandey et al. 2023; Abosata et al. 2023; Ahmed et al. 2023; Vu et al. 2022) |

a well-rounded representation of research developments, particularly considering publications from 2023 that may not yet have garnered citations. Some articles were found to be relevant to both ML and DL. To address this overlap, we adopted a meticulous approach to ensure consistency. We carefully reviewed each manuscript in its entirety to determine whether its primary focus leaned more towards ML or DL. Accordingly, we were able to categorize each paper according to its predominant emphasis, thus ensuring clarity and coherence in our analysis.

We had to establish a set of themes or categories for the common trends before we could begin examining the chosen articles. In order to do this, we systematically looked into the Scopus database's keywords for both ML and DL searches. A cross-analysis of the terms in both ML and DL was also done. The terms "Intrusion Detection", "Anomaly Detection", "Blockchain", "Feature Extraction", "Forecasting", "Attack Detection", "Feature Selection", "Authentication", "Information Management", "Automation", "Performance", and "optimization" were found to be frequently occurring keywords for both ML and DL. These keywords enabled us to generate the first set of themes and then organize the findings from the chosen articles according to how closely each topic was linked to the others. "Anomaly Detection", "Security Automation", "Native Security", and "Emerging Technologies" are the four common themes that we found. Apart from the shared themes between DL and ML, we observed that DL concentrates more on image identification and network traffic analysis, whereas ML search terms are more focused on insider threats and security brokers. These foci were consequently added to the DL and ML themes, respectively. Table 8 summarizes these trends.

### 5.1 Machine learning trends

This review revealed several trends and objectives for applying ML to cloud security. Insider threats, anomaly detection, security automation and intelligence, cloud-native security, and emerging technology impact are the six general categories that we used in this article to group them.

#### 5.1.1 Anomaly detection

Due to its effectiveness in identifying and responding to security incidents, ML-based anomaly detection is increasingly employed in cloud systems (Hassija et al. 2019; Jaber and Rehman 2020). Compared to conventional rule-based systems, ML-based anomaly detection is generally considered more accurate and effective (Chkirbene et al. 2020; Heidari and Jabraeil Jamali 2023). It is important to remember that anomaly detection is a typical application for ML and has been thoroughly investigated and used in various fields, including fraud detection, intrusion detection, and system monitoring (Dey et al. 2019; Rahila and Khonde 2019). Consequently, ML-based anomaly detection is expected to remain a common method for protecting cloud systems (Ghazal et al. 2021; Tian et al. 2021; Hameed et al. 2021; Abidi et al. 2021).

Anomaly detection involves finding patterns or occurrences inside a system that differ from typical behavior. It encompasses various techniques such as monitoring and classification of data to distinguish between regular and irregular activities, as well as real-time detection and response mechanisms to promptly address detected anomalies (Belal and

**Table 8** ML and DL research trends

| ML | | DL | |
|---|---|---|---|
| Trend | Focus | Trend | Focus |
| Anomaly detection (Belal and Sundaram 2022; Saran et al. 2022; Mothukuri et al. 2021a; Angelopoulos et al. 2020; Wu et al. 2019) | • Monitoring, classification, detection, and response<br>• Data and resource management<br>• Various ML algorithms and models | Anomaly detection<br>Aldallal 1916; Ge et al. 2021; Kasongo 2023; Kumar et al. 2022a; Mahdavifar and Ghorbani 2019; Zhou et al. 2023; Lahande and Kaveri 2022) | • Cloud anomaly monitoring<br>• Fraud Email analysis<br>• Technique evaluation<br>• Imbalanced datasets analysis<br>• Multi-model solutions<br>• Real-time analysis<br>• Reinforcement learning techniques |
| Security automation & intelligence<br>Kumar et al. 2021; Rosero et al. 2021; Quraishi 2022; Narayanan and Muthukumar 2022) | • Intelligence visualization<br>• Automated incident response, policy enforcement, security orchestration, and threat feedback<br>• Predictive modeling | Security automation (Tuli et al. 2020; Rjoub et al. 2021; Ferrag et al. 2020; Zhang et al. 2022; Ahmad et al. 2022) | • Automated detection and incident response<br>• Self-learning and root cause analysis |
| Native security (Narayanan and Muthukumar 2022; Kumar et al. 2022b; Mishra et al. 2022a) | • Network security solutions<br>• Management platforms<br>• Policy management and specific considerations | Native security<br>Dalal et al. 2023; Landman and Nissim 2021) (Parra et al. 2020; Thilagam and Aruna 2021) | • Security analysis and hunting<br>• Security incident management<br>• Cloud-agnostic and compliance management |
| Emerging technology's role<br>Abbas and Myeong 2023; Gupta et al. 2017; Dasgupta et al. 2022; Saran et al. 2022; Huang et al. 2022; Amin et al. 2023) | • Blockchain<br>• Quantum-resistant cryptography<br>• Secure multi-party computation<br>• Homomorphic encryption<br>• Zero trust network access | Emerging technology's role<br>Gupta et al. 2017; Pandey et al. 2023; Kumar et al. 2021; Dai et al. 2020; Wu et al. 2020; Ahmed et al. 2023) | • Blockchain<br>• Quantum-resistant cryptography<br>• Secure multi-party computation<br>• Homomorphic encryption<br>• Zero trust network access |
| Insider threats (Dasgupta et al. 2022; Mothukuri et al. 2021a; Gupta et al. 2020; Tian et al. 2021; Narayanan and Muthukumar 2022) | • Behavioral detection<br>• User profiling<br>• NLP and graph analysis | Image-based detection (Ullah et al. 2022; Ahmed et al. 2023; Wu et al. 2022) | • Malware behavior analysis<br>• Malware classification and visualization<br>• Malware variant detection<br>• Malware image compression |
| Security broker (Saran et al. 2022; Goh et al. 2021; Abosata et al. 2023) | • Identity and access management<br>• Data loss prevention<br>• Threat protection<br>• Compliance management | Network traffic analysis (Pandey et al. 2023; Gu et al. 2019; Chen et al. 2020; Dai et al. 2020; Zhang et al. 2022) | • Network traffic classification<br>• Traffic flow analysis<br>• Network traffic manipulation |

Sundaram 2022; Saran et al. 2022). This area also emphasizes efficient cloud data management practices to handle the large volumes of data generated during anomaly detection processes (Mothukuri et al. 2021a). Additionally, resource management strategies are essential for optimizing computational resources and ensuring timely anomaly detection (Wu et al. 2019). Finally, the integration of multiple ML algorithms and models enhances the accuracy and robustness of anomaly detection systems by leveraging diverse perspectives and approaches (Angelopoulos et al. 2020).

### 5.1.2 Security automation and intelligence

ML models constantly get better based on the information they gather over time. ML-based systems can self-learn and modify their detection and response to evolving threat environments. According to research, more people are using security automation. According to a 2019 SANS Institute survey, 42% of respondents claimed they were adopting security automation to enhance incident response (Spitzner 2022). Moreover, Gartner forecasted that 30% of all security incident response processes will be improved by artificial intelligence and ML by 2021 in order to strengthen incident detection and response (Gartner. 2023). On the other hand, the SANS survey revealed that 53% of respondents claimed they used threat intelligence to enhance incident response and cloud security. Automation of security-related tasks and security inelegance have been suggested as ML benefits.

Security automation and intelligence aim to streamline security operations and enhance threat detection and response capabilities through automated processes and intelligent decision-making (Rosero et al. 2021). Automated incident response mechanisms enable rapid identification and containment of security incidents, minimizing their impact on the organization (Quraishi 2022). Automatic threat intelligence visualization facilitates the interpretation of complex security data, enabling security professionals to make informed decisions effectively (Narayanan and Muthukumar 2022). Automated policy enforcement ensures compliance with security policies and regulations across various systems and applications, reducing the risk of security breaches. Security orchestration automates the coordination of security tools and processes, improving efficiency and effectiveness in managing security incidents (Kumar et al. 2021). Predictive modeling techniques leverage historical security data to forecast potential threats and vulnerabilities, enabling proactive risk mitigation measures (Narayanan and Muthukumar 2022).

### 5.1.3 Cloud-native security

In cloud-native security, security should be built into the cloud-based platforms, infrastructure, and applications. Multiple layers of protection are implemented, and security is continuously monitored for emerging vulnerabilities from the beginning of the development process to production. As more businesses use cloud-based technology, cloud-native security becomes more crucial, and ML is frequently used as a significant component of these solutions. There is an increase in the use of cloud-native security solutions. 55% of enterprises are employing cloud-native security solutions, according to a survey by IDC 2020, and that percentage is anticipated to reach 70% by 2022 (IDC. Idc marketscape worldwide managed security services 2020 vendor assessment. 2020). Additionally, according to a 2019 SANS Institute survey, 41% of participants claimed they were employing cloud-native security to enhance incident response and cloud security (Spitzner 2022).

Cloud-native security focuses on integrating security measures directly into cloud environments to protect cloud-native applications and data (Mishra et al. 2022a). This includes implementing cloud-native network security solutions to safeguard network communications and data transmission within cloud environments. Cloud-native management platforms offer centralized security management capabilities, enabling organizations to monitor and enforce security policies consistently across cloud resources (Narayanan and Muthukumar 2022). Cloud-native security policy management involves defining and enforcing security policies tailored to the specific requirements and configurations of cloud-native applications and services. Considerations such as multi-tenancy and dynamic resource allocation are essential for addressing the unique challenges of securing cloud-native environments effectively (Kumar et al. 2022b).

#### 5.1.4 Role of emerging technologies

In the ML and cloud security frameworks, plenty of investigation has been conducted to fully explore the possibilities of new technologies like blockchain and quantum computing. Although integration is not always possible immediately, the ideas present exciting possibilities for breakthroughs in the future (Gupta et al. 2017). Blockchain technology ensures the integrity and provenance of security data, enhancing trust and transparency in security operations (Heidari et al. 2023). Quantum-resistant cryptography protects sensitive data and ML models against emerging threats posed by quantum computing (Amin et al. 2023). Secure multi-party computation enables collaborative model building and inference without compromising data privacy and confidentiality (Mishra et al. 2023). Homomorphic encryption allows ML algorithms to operate directly on encrypted data, preserving privacy while enabling data analysis (Dasgupta et al. 2022; Huang et al. 2022). Zero trust network access models enhance security by enforcing strict access controls based on user and device attributes, minimizing the risk of data breaches and unauthorized access (Saran et al. 2022; Gupta et al. 2022). These emerging technologies complement ML approaches to address evolving security challenges effectively.

#### 5.1.5 Identifying insider threats

ML is being utilized increasingly to detect and address insider threats, as they are among the most significant concerns for enterprises. According to studies, ML-based systems are more precise and effective than conventional rule-based systems for detecting insider threats (Gupta et al. 2017). Additionally, studies have demonstrated that companies use behavioral analytics to identify and stop insider threats. 68% of respondents to a 2019 SANS Institute survey stated they used behavioral analytics to identify insider threats (Spitzner 2022). Identifying insider threats involves detecting and mitigating security risks posed by individuals within an organization who have access to sensitive information and resources (Dasgupta et al. 2022). ML techniques are applied to analyze user behavior, profile users, and process natural language to identify anomalous activities indicative of insider threats (Narayanan and Muthukumar 2022). Behavioral detection models monitor user actions and interactions with IT systems to detect deviations from normal behavior patterns (Tian et al. 2020). User profiling techniques leverage ML algorithms to create profiles of individual users based on their behavior, preferences, and access privileges, enabling organizations to identify suspicious activities and potential insider threats (Tian et al. 2021). NLP technologies analyze textual data, such

as emails and chat logs, to identify language patterns associated with insider threats (Mothukuri et al. 2021b). Graph analysis techniques model relationships between users, devices, and data to uncover hidden connections and potential security risks within the organization's network (Amiri et al. 2024).

#### 5.1.6 Cloud access security broker

CASB is software deployed in the cloud or installed on-premises and serves as a bridge between customers and cloud service providers (SANS 2024). The ability of CASB to automatically detect and prohibit malicious activities has been improved by applying ML algorithms. CASB offers security features, including encryption, data loss prevention, and intrusion prevention. ML is frequently employed as a crucial component of CASB solutions, which are becoming increasingly popular. According to an IDC survey from 2020, 45% of businesses are utilizing CASB solutions, and by 2022, that percentage is predicted to reach 60% (IDC. Idc marketscape worldwide managed security services 2020 vendor assessment. 2020). In addition, a 2019 SANS Institute survey revealed that 37% of participants claimed they used CASB to enhance incident response and cloud security (Spitzner 2022).

CASB solutions provide visibility and control over cloud services and applications to ensure data security and compliance in cloud environments. ML-powered CASB platforms offer identity and access management capabilities, enabling organizations to enforce granular access controls and authentication mechanisms to protect against unauthorized access (Yu et al. 2021). Data loss prevention features leverage ML algorithms to detect and prevent data leaks or unauthorized sharing of sensitive information on cloud storage and collaboration platforms (Narayanan and Muthukumar 2022). Threat protection modules utilize ML-driven threat intelligence to identify and block malicious activities, such as malware infections and phishing attacks, targeting cloud services and applications (Balobaid et al. 2024). Compliance management functionalities enable organizations to assess and enforce compliance with industry regulations and standards, such as the Insurance Portability and Accountability Act (HIPAA), PCIDSS, General Data Protection Regulation (GDPR), and CCPA, by analyzing cloud usage and data handling practices against regulatory requirements (PonemonInstitute. 2024).

### 5.2 Deep learning trends

Although DL for cloud security is an emerging research area, several focuses and aims were identified in this review. This review revealed several trends and objectives for applying DL to cloud security. Anomaly detection, image-based malware detection, cloud-native security, and security automation are the six general categories used in this article to group them. The endeavors in this context have concentrated on utilizing a range of neural network architectures, including CNNs, RNN, and Long Short-Term Memory (LSTM) (Kumar et al. 202a; Rjoub et al. 2021; Zhang et al. 2022; Abou El Houda et al. 2022). Furthermore, the focus has been on assessing the effectiveness of the network traffic analysis system through performance metrics such as F1-score, precision, and accuracy (Zhou et al. 2021; Thilagam and Aruna 2021; Makkar et al. 2021; Landman and Nissim 2021).

### 5.2.1 Anomaly detection

An important field of research and development, the use of DL for anomaly detection is growing in acceptance (Heidari and Jabraeil Jamali 2023). The application of DL in anomaly detection is a burgeoning area of research and development, gaining popularity due to its potential for learning intricate non-linear relationships, automatic feature extraction, and improved performance over time (Doriguzzi-Corin et al. 2020). DL has been demonstrated to be effective in several anomaly detection tasks across various domains, including spotting cyberattacks in network systems, intrusion detection, fraudulent financial transactions, and machinery breakdowns in industrial systems. According to a 2019 SANS Institute poll, the usage of DL in anomaly detection is anticipated to rise over the next two years, with 37% of respondents stating they intended to do so (Spitzner 2022). Here, the focus has been on generative models, such as Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN), which are employed to create synthetic data that can be used to train models for detecting anomalies. Moreover, the focus has been on assessing the anomaly detection system's performance using various metrics such as precision, recall, F1-score, and the Receiver Operating Characteristic Curve (AUC-ROC), which evaluates the system's ability to identify anomalies.

Anomaly detection involves identifying irregular patterns or events within a system that deviate from normal behavior (Kasongo 2023). DL techniques, such as deep neural networks, are employed to analyze complex and high-dimensional data to detect anomalies across various domains, including cybersecurity, finance, and industrial systems (Aldallal 1916). DL-based anomaly detection systems leverage advanced architectures such as CNNs and RNN to capture intricate non-linear relationships in data and automatically extract relevant features for anomaly detection (Mahdavifar and Ghorbani 2019). Evaluation metrics such as recall, precision, accuracy, and F1-score are used to assess the effectiveness of DL-based anomaly detection systems in detecting and mitigating security threats (Zhou et al. 2023; Lahande and Kaveri 2022).

### 5.2.2 Security automation

Based on the works cited in this paper, the following sections provide a summary of the previous work on utilizing DL for security automation (Tuli et al. 2020; Rjoub et al. 2021; Rosero et al. 2021; Zhang et al. 2022; Quraishi 2022). First, automation of anomaly detection has been demonstrated to be very accurate and have low false-positive rates when detecting unexpected patterns or behaviors that might point to a security problem. Second, automated incident response, in which DL has been used to automate responding to and mitigating issues, as well as prioritizing and escalating incidents depending on their severity or possible effect. DL has also been used to recover from incidents, such as by automatically rolling back modifications or restoring impacted resources. Third, self-learning: in incident response circumstances, when the cause of an occurrence is automated and rapidly determined, DL models that can explain how they came to their judgments might be very crucial. Finally, root cause analysis, where the underlying cause of events, such as a particular vulnerability or attack vector, has been found using DL. Moreover, DL has been used to predict and prevent future incidents.

### 5.2.3 Cloud-native security

Another prominent area of study and development is the application of DL for cloud-native security, which is growing in acceptance. This entails employing DL algorithms to integrate native security capabilities such as firewalls, IDS, and security information and event management systems into cloud-based services. Cloud-native security involves leveraging DL algorithms to integrate native security capabilities into cloud environments, such as firewalls, IDS, and SIEM systems (Dalal et al. 2023). DL-based native security solutions enable organizations to monitor and analyze security data in real-time, identify threats and vulnerabilities, and automate incident response processes (Landman and Nissim 2021). DL algorithms are used to evaluate security data in cloud systems, such as by detecting threat trends, identifying abnormal activities, and automating security tool orchestration. Additionally, DL-based incident management systems enable organizations to handle security events, determine incident causes, and remediate security issues effectively. Cloud-agnostic DL models are developed to ensure compatibility with various cloud platforms and address unique challenges in securing cloud environments, such as multi-tenancy and dynamic resource allocation (Parra et al. 2020; Thilagam and Aruna 2021). Compliance and regulatory management solutions leverage DL techniques to ensure cloud environments comply with industry regulations and standards, such as HIPAA and PCI-DSS, thereby enhancing data security and privacy in cloud deployments (Parra et al. 2020; Thilagam and Aruna 2021).

### 5.2.4 Role of emerging technologies

Cloud security may benefit significantly from the mix of emerging technologies and DL. The role of emerging technologies in DL encompasses various advancements aimed at enhancing the security and efficiency of DL applications. Blockchain technology ensures the integrity and reliability of training data by providing tamper-proof data provenance, thereby improving the trustworthiness of DL models (Alzoubi et al. 2022a, 2022b, 2022c; Rathore et al. 2019; Kumar et al. 2021; Dai et al. 2020; Ahmed et al. 2023). Quantum-resistant cryptography safeguards DL models and sensitive data against future quantum computing threats, ensuring long-term security (Amin et al. 2023). Secure multi-party computation enables collaborative DL model training and inference without compromising data privacy, facilitating secure knowledge sharing and insight generation. Homomorphic encryption allows DL computations to be performed directly on encrypted data, preserving data privacy and enabling privacy-preserving analytics in cloud environments (Gupta et al. 2017; Pandey et al. 2023; Wu et al. 2020). Zero-trust network access enhances DL infrastructure security by dynamically controlling access based on user and device profiles, reducing the risk of insider threats and unauthorized access (Landman and Nissim 2021).

### 5.2.5 Image-based malware detection

The usage of this technology is growing in popularity and is an active field of study and development. This entails applying DL algorithms to examine images and find malware concealed in images or other multimedia files, such as Trojans and ransomware (Rao and Jain 2024). Image-based malware detection leverages DL algorithms to analyze images and multimedia files for the presence of malware, such as Trojans and ransomware (Nahmias

et al. 2020). Malware behavior analysis involves using DL models to analyze the behavior of malware, such as its propagation methods and actions, to detect and mitigate threats effectively [100. Malware classification employs DL techniques to categorize malware into different types, aiding in identifying and addressing specific threats. Malware variant detection utilizes DL models to identify new iterations of existing malware, enhancing threat detection capabilities {Wu et al. 2022 #253]. Visualization techniques are used to understand how DL models make decisions and identify features used for malware detection, improving model interpretability and detection accuracy. Malware image compression employs DL to condense malware images, enhancing detection model efficiency while reducing storage and processing requirements (Ahmed et al. 2023; Wu et al. 2022).

#### 5.2.6 Network traffic analysis

As an active field of research and development, the use of DL for network traffic analysis is growing in acceptance. To do this, DL algorithms are used to examine network data and spot trends that might point to a possible security problem, such as malware or a network incursion (Dai et al. 2020). DL-based network traffic analysis focuses on using DL algorithms to analyze network data for security threat detection and performance optimization (Chen et al. 2020). Network traffic classification involves categorizing network traffic into different groups, such as legitimate or malicious traffic, using DL models to enhance threat detection capabilities (Pandey et al. 2023). Traffic flow analysis utilizes DL to evaluate network traffic flows, identify fraudulent activities, and improve network performance through anomaly detection. Network traffic manipulation employs DL techniques to compress, generate, or standardize network traffic data, enhancing the efficiency and effectiveness of detection models while reducing resource requirements and data complexity (Gu et al. 2019; Dai et al. 2020).

### 5.3 Major machine learning and deep learning techniques and algorithms

It's critical to remember that the distinctions between ML and DL are not always apparent, and specific approaches may fall somewhere in the middle. For instance, gradient boosting and RF are ML techniques that may also be applied in DL settings.

#### 5.3.1 Machine learning techniques

It is also essential to remember that applying ML approaches to cloud computing security is still a very young field of study, and new methods and algorithms are always being created. Identifying the optimum strategy is challenging since it will change based on the particular use case and dataset (Nissim et al. 2019; Balamurugan and Saravanan 2019; Yao et al. 2019). The proper settings, characteristics, and data pre-processing should be employed with the technique of choice in order to get good results. The following summarizes the major ML techniques (Mishra and Tyagi 2022; Saran et al. 2022; Nair et al. 2023; Butt et al. 2020; Chkirbene et al. 2020; Alzoubi et al. 2023). It is critical to select the appropriate methods and settings for a given problem since the performance of various strategies might vary based on the particular use case and dataset.

1. RF: This method is an ensemble method that can be used for intrusion detection and classification tasks.

2. SVM: This method is a potent algorithm that may be applied to problems requiring classification and regression.
3. K-means: This method uses a clustering algorithm to find patterns and irregularities in huge datasets.
4. Gradient boosting: This ensemble method may be applied to challenges, including classification and regression.
5. Genetic algorithm: The ML models' parameters are optimized using this method.
6. Transfer learning: This method applies the information gained from one task to another to enhance the performance of the ML models.
7. Federated learning: With the help of this method, several parties may train ML models without sharing any data.
8. Multi-task learning: Teaching the ML models to carry out several tasks at once is a strategy that helps them perform better.
9. Hyperparameter tuning: By adjusting the hyperparameters, this method helps ML models perform as well as possible.
10. Decision trees: One type of ML method that may be applied to classification and regression applications is decision trees. They may be used to find patterns and correlations in data, which helps detect intrusions and other security-related activities.
11. Bayesian networks: An example of a probabilistic graphic model is a Bayesian network, which may be applied to tasks like intrusion detection, anomaly detection, and classification.
12. Hidden Markov Models (HMM): Intrusion detection, anomaly detection, and sequence prediction are just a few examples of the activities that may be performed with HMMs, a category of probabilistic models.
13. Logistic regression: A statistical technique used for classification tasks is logistic regression. It may be utilized for data pattern recognition and intrusion detection.
14. Naive Bayes: Naive Bayes can be referred to as a probabilistic classifier. It may be utilized for data pattern recognition and intrusion detection.

#### 5.3.2 Deep learning techniques

DL techniques are thus named because they feature several layers, enabling them to learn more abstract and complicated input representations. The essential DL techniques revealed in this work are listed below (Gupta et al. 2017; Tuli et al. 2020; Dai et al. 2020; Wu et al. 2020; Ramchandran and Sangaiah 2020).

1. Reinforcement learning: This technique has recently been utilized to strengthen cloud computing security by teaching the agent how to behave securely.
2. RNN: These are a class of DL techniques used in applications, including intrusion detection, anomaly detection, and sequence prediction in cloud computing settings. They work especially well with sequential data, such as time series.
3. LSTM networks: This type of RNN algorithm may be applied to intrusion detection and time series analysis.
4. Neural networks: One example of ML technology that draws inspiration from the structure and operation of the human brain is neural networks. They can be applied to many tasks, including intrusion detection, anomaly detection, and classification.
5. Autoencoder: This technique uses a specific kind of neural network that may be applied to anomaly and intrusion detection.

6. GAN: Data creation and anomaly detection are possible applications for this technique, a kind of neural network.
7. CNNs: These are a family of DL methods frequently employed for image and video analysis jobs, and they have been applied to jobs like intrusion detection, anomaly detection, and classification in cloud computing settings.
8. Adversarial training: By creating and using adversarial instances throughout the training process, this method teaches a model to be resistant to them.
9. Adversarial examples: These instances were created deliberately to trick ML models, and they were used to gauge how reliable ML models were for cloud computing security.
10. DBN: A type of deep neural network called a DBN is capable of feature extraction and categorization. This model can be employed to identify spam and phishing emails.
11. Capsule networks: A relatively new type of DL model called capsule networks is designed to better deal with spatial interactions between features. They have been utilized in security systems for image identification and categorization tasks.
12. Attention Mechanisms: Attention techniques are utilized to direct a DL model's attention to particular areas of the input data. They have been applied to intrusion detection and network traffic analysis.
13. Deep Residual Networks (ResNets): Deep neural networks called ResNets are made to deal with the vanishing gradient problem, which can happen while training very deep networks. They have been applied to classification and image recognition tasks in security systems.
14. VAE: An autoencoder class called VAE is employed to create fresh data samples. They have been used to create fake network traffic data for testing and refining security models and solutions.
15. Siamese networks: One class of deep neural network called a Siamese network compares two input data samples and produces a similarity score. They have been used to spot network traffic irregularities and malware samples that are similar or identical.

## 6 Discussion

The primary objective of this study was to evaluate and examine the patterns observed in published literature relating to ML and DL for cloud computing security. The study initially performed a bibliographic analysis by utilizing the Scopus database, which was crucial in identifying the areas, countries, institutions, and authors that demonstrated the most interest in this subject. Subsequently, the study utilized a systematic sampling approach to identify the trends and focal arguments of research papers published in this field to gain a more thorough understanding of ML and DL's role in cloud computing security. The study discovered a total of 4051 relevant documents as of mid-December 2023, with the literature on ML dating back to 2006 and the literature on DL beginning in 2016. This substantial volume of literature in a relatively short period underscores the significance of this subject.

As shown in Fig. 1, there has been a significant growth in ML and DL articles related to cloud computing security in recent years. This may relate to the combination of technological advancements, increased data availability, industry adoption, evolving threats, academic interest, and interdisciplinary collaboration. Exploring these factors further can provide valuable insights into the underlying drivers of this trend. Regarding contributions

by regions and countries, Asian nations such as India and China were found to be the most productive based on the number of documents published (1033 in India and 742 in China). The top 15 active countries account for the majority of ML and DL publications, highlighting the concentration of research efforts in these nations. Among the top 15 active countries, there is a diverse regional representation, with countries from Asia, Europe, the Middle East, and North America featuring prominently. This suggests widespread interest and involvement in ML and DL research for cloud security across different regions. In both ML and DL research, Asia emerges as the most active continent, with a significant number of countries contributing to the publications. The Middle East also demonstrates notable participation, particularly in DL research.

The top 15 active institutions, notably from China, Saudi Arabia, India, and other countries, demonstrate global collaboration and contributions to ML research. Similarly, in DL research, these regions, along with Singapore and Qatar, show strong representation among the top 15 institutions, highlighting collaborative efforts. These statistics underscore the global presence of research institutions in ML and DL for cloud security, reflecting collective efforts to address challenges through collaborative research. Taking into account the influence and output of scientific publications, the *Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, which includes subseries on artificial intelligence and bioinformatics, had the highest number of published articles (158), followed by *IEEE Access* (156), which is a multidisciplinary, open access journal that publishes in engineering, computer science, and materials science. *IEEE Access* ranked as a Q1 according to the evaluation by quartiles in SJR (0.93) with an H-index of 204.

The research trends in ML and DL for cloud computing security reveal eight primary focus areas, including anomaly detection, security automation, native cloud security, insider threats, security brokers, network traffic analysis, image-based analysis, and the impact of emerging technologies. A closer examination of the literature on ML and DL demonstrates that DL serves as an extension of ML, either by expanding upon the work of ML or by employing more advanced capabilities provided by DL. Additionally, it is worth noting that ML and DL are relatively new technologies in the realm of cloud computing security, and ongoing research in this field indicates that there is still significant potential for the development of new techniques and algorithms.

### 6.1 Machine learning and deep learning—finding cross analysis

To conduct a cross-analysis between the ML and DL findings, we identify common themes, compare methodologies, and highlight differences or complementary aspects. We break down the analysis based on the main themes identified in both the ML and DL discussed in Sect. 5.

- Anomaly detection: DL and ML algorithms are broadly used for anomaly detection in cloud security. ML techniques include monitoring and classification, detection and response, cloud data management, resource management, and the use of multiple ML algorithms and models. However, DL techniques, such as unsupervised autoencoders and generative models like VAE and GAN, are employed for anomaly detection. DL focus areas include traffic analysis, handling imbalanced datasets, multi-model solutions, real-time analysis, and cloud anomaly monitoring.

- Security automation: ML-based security automation involves automatic incident response, threat intelligence visualization, policy enforcement, security orchestration, threat intelligence feedback, and predictive modeling. ML techniques include automated analysis of threat intelligence data, policy enforcement, and security orchestration using ML algorithms. However, DL is utilized for automating anomaly detection, incident response, self-learning, and root cause analysis. DL focuses on the accuracy and low false-positive rates of automated detection systems and automating incident response and recovery processes.
- Cloud-native security: ML techniques are integrated into cloud-native security solutions for network security, management platforms, security policy management, and compliance management. However, DL is applied to native security analysis, incident management, cloud-agnostic solutions, and compliance management. DL focus areas include using DL to automate incident response, manage security incidents, and ensure compliance with regulations.
- Role of emerging technologies: ML and DL explore the integration of emerging technologies like blockchain, quantum-resistant cryptography, homomorphic encryption, secure multi-party computation, and zero-trust network access. They emphasize the potential benefits of these technologies in enhancing data security, privacy, and threat detection in cloud environments.

### 6.2 Practical implications

Both ML and DL offer practical implications for the industry, including improved security measures, automation of security processes, enhanced anomaly detection, and compliance management. While ML and DL share common objectives for enhancing cloud security, they employ different methodologies and techniques. ML tends to focus on traditional algorithms and supervised/unsupervised learning, while DL leverages neural network architectures and DL models for more complex pattern recognition tasks. Integrating both approaches can offer comprehensive solutions for handling various cloud computing security issues. ML and DL enable organizations to better protect their cloud environments, respond to security incidents more efficiently, and adapt to evolving threat landscapes. Additionally, ML is applied to identify insider threats, enhance CASB, and manage cloud-native security. However, DL techniques are used for image-based malware detection and network traffic analysis, focusing on malware behavior analysis, classification, variant detection, and visualization.

### 6.3 Challenges and research directions of machine learning and deep learning in the cloud security

While ML and DL are considered vital tools for cloud computing security, it is essential to understand the underlying data and the specific use case to select the correct algorithm, employ the proper feature engineering, and establish suitable thresholds to get the required performance (Pandey et al. 2023). Although ML and DL are not a one-size-fits-all solution, they should be used in conjunction with other security measures, including access control, network security, and incident management processes. The application of ML and DL for cloud security faces several difficulties, including the following:

1. Data privacy: Large volumes of data are needed to train ML and DL models, and this data must be protected to maintain user privacy (Narayanan and Muthukumar 2022). This might not be easy in cloud computing, where data is often processed and stored on shared infrastructure. Insider threats may be intricate and multidimensional. Therefore, it's crucial to remember that ML-based solutions alone might not be enough. A thorough security plan that may include technological, administrative, and physical measures should be in place to guard against insider threats (Mishra et al. 2023, 2022b). The implementation of robust data encryption and access control mechanisms can also help safeguard sensitive data. Employment techniques such as differential privacy can help to anonymize datasets while preserving their utility for training ML models (Nair et al. 2023). Additionally, utilizing federated learning approaches can help train models directly on decentralized data sources without exposing raw data to third parties (Abusitta et al. 2019). It is crucial to make sure that user privacy is maintained, given the growing usage of ML and DL in cloud security (Gupta et al. 2017). Creating methods for protecting privacy may be possible while enabling helpful research in this field (Alasmary et al. 2019).
2. Adversarial attacks: Adversarial attacks, in which a perpetrator tampers with the model's input to make it make the wrong judgments, are possible against ML and DL models (Saran et al. 2022). Models may be installed in untrusted contexts in cloud computing environments, making this particularly difficult. Adversarial ML research is crucial because cloud security increasingly uses ML models (Rathore et al. 2019). This entails creating techniques to strengthen the resistance of ML models to hostile cases and methods to recognize and fend off adversarial attacks. Moreover, incorporating techniques like input preprocessing and model regularization to enhance robustness against adversarial perturbations can be helpful (Ullah et al. 2022). Additionally, the employment of anomaly detection algorithms can help to identify and mitigate adversarial instances in real-time, coupled with dynamic model retraining to adapt to evolving attack strategies (Bhuvaneshwari and Kaythry 2023).
3. Scalability and performance: It might be challenging to apply ML and DL models in cloud systems where several users share resources since it is computationally costly to integrate ML and DL models with current infrastructure and security systems since these systems may have distinct needs and limitations (Heidari et al. 2022; Mishra et al. 2020). In addition, cloud computing security ML and DL models must abide by several laws and standards, including GDPR, HIPAA, and PCI-DSS (Mishra et al. 2022c). In order to effectively utilize cloud resources, the recommendations made here are centered on optimizing ML and DL algorithms for distributed computing systems (Heidari and Jabraeil Jamali 2023). Another recommendation is to modularize ML components and enable smooth integration with the current cloud infrastructure by leveraging containerization and microservice architecture (Gupta et al. 2017). Furthermore, to ensure optimal performance and resource utilization, the implementation of auto-scaling techniques can help to dynamically assign computing resources based on workload demands (Dasgupta et al. 2022).
4. Explainability: Many ML and DL models are difficult to read, making it challenging to comprehend the models' decision-making processes (AlAhmad et al. 2021). This may be particularly difficult regarding cloud computing security since it's crucial to understand a decision's justification (Hossain et al. 2020). The recommendations made here are centered on using explainable AI methods to clarify the variables affecting model predictions, such as feature importance analysis and model visualization (Gupta et al. 2017). Moreover, using decision trees or rule-based models in conjunction with intricate

ML and DL models can produce clear decision-making logic. The documentation of the model training and evaluation procedures may improve accountability and transparency in cloud security operations (Bhuvaneshwari and Kaythry 2023).

5. Generalization: It might be difficult for ML and DL models to generalize successfully to previously unknown data due to the possibility of novel attack types and abnormalities (Mahdavifar and Ghorbani 2019). In order to improve model generalization across various circumstances, the recommendations made here center on augmenting training datasets with a variety of representative samples (Dasgupta et al. 2022). Additionally, transfer learning techniques are used to deploy pre-trained models and modify them for use in particular cloud security tasks, which lessens the requirement for large amounts of labeled data (Bhuvaneshwari and Kaythry 2023). Overall, maintaining the model's performance and updating it frequently are recommended to take into account fresh threat intelligence and adjust to changing security issues (Lei et al. 2022).
6. Label bias: The performance of the models depends on the data quality, and bias in the data might lead to bias in the models. Although the data cannot always be indicative of real-world events, this might offer difficulty for cloud computing security (Yao et al. 2019; Krishnan et al. 2019). The recommendations made here center on the use of implementation data preparation techniques such as undersampling or oversampling to reduce label bias and guarantee that all classes are equally represented in training datasets (Belal and Sundaram 2022). To further promote equality and inclusion in cloud security applications, employment fairness-aware learning algorithms may be used to identify and reduce biased practices in model predictions (Krishnan et al. 2019). Furthermore, bias propagation in ML and DL models can be minimized by carrying out comprehensive evaluations of the quality of the data and establishing precise criteria for data collection and annotation (Gupta et al. 2017).

It is worth noting that ML, DL, and cloud computing security are all active research areas. New techniques and algorithms are evolving every day. The promising research areas found in this review are listed below.

1. More research is needed in the developing domains of automated incident response, image-based malware detection, and cloud-native security.
2. ML algorithms applied to cloud-native security, which entails utilizing ML algorithms to assess and safeguard cloud-native apps and services, is the subject of growing research.
3. In big and complicated cloud settings, there is a need to create more effective and efficient ways for identifying abnormalities and intrusions. Further developments in this area may bring up new elements and variables in ML and DL for security event correlation.
4. Federated learning presents a viable approach to augmenting security in cloud computing since it enables cooperation across several organizations.
5. As it enables the model to identify and stop various security risks, multi-task learning may benefit cloud computing security.
6. Cloud security is increasingly dependent on edge/fog computing as more and more devices and sensors are connected to the internet (Alzoubi and Aljaafreh 2023). To increase security and privacy, research in this field may focus on creating ML and DL approaches that can be used on edge and fog devices.
7. IoT-based data processing in cloud computing environments needs more sophisticated, secure operations using ML and DL.

### 6.4 Study limitations

There are certain limitations to the current investigation. They used only one database (Scopus) without considering other databases often utilized in the academic community, such as Dimensions and Web of Science. While this study exclusively utilizes the Scopus database, it's important to acknowledge its status as a comprehensive academic repository, encompassing many journals and conferences also indexed in the Web of Science, for example. While we recognize the value of other databases such as Web of Science, our decision to focus solely on Scopus was based on several factors, including accessibility, coverage of relevant literature, and familiarity with the platform. Nonetheless, it's possible that some sources are not included. Despite this limitation, the findings offer valuable insights into the emerging research theme, benefiting researchers in the field.

While retrieving literature from the Scopus database, there was a slight bias in favor of nations having a lot of articles indexed there. Scopus has a bias in favor of academic journals with English-language publications. As a result, publications published in languages other than English were not retrieved, potentially leaving out essential contributions made in those other languages, notwithstanding their preponderance. In the identified articles, non-English publications were primarily in Chinese, Turkish, and Portuguese, comprising 1.1% of ML and 1.8% of DL papers. It's important to note that we lack proficiency in these languages, limiting our ability to analyze the findings from these papers. This limitation may have impacted the comprehensiveness of our study, particularly in capturing insights from non-English sources.

As with all prior bibliometric research, the current study excluded gray literature. This study's search method may have produced some false-positive or false-negative findings. Thus, these constraints should be taken into account while interpreting the results. Nonetheless, the study outlines a strict approach for the chosen papers and uses a database widely regarded as trustworthy by experts. A study effort that may be utilized as a reference for future researchers can be presented thanks to the word choice, period, and number of documents used. Future studies might usefully apply our analytical methodology to another language (such as Chinese or Turkish) and provide a thorough worldwide understanding of the literature.

## 7 Conclusions

Cloud security is a critical concern as organizations increasingly rely on cloud computing for data storage, processing, and application hosting. Traditional security measures face challenges in detecting and mitigating sophisticated cyber threats targeting cloud environments. In response, ML and DL techniques have emerged as powerful tools to bolster cloud security. Massive data sets can be analyzed by ML algorithms to find trends and abnormalities that point to cyberattacks, while DL models excel at image-based threat detection and network traffic analysis. Although numerous papers have been published on utilizing ML and DL to bolster cloud security, there remains a dearth of comprehensive bibliographic reviews that synthesize the techniques, trends, and challenges in this domain. Therefore, this paper seeks to provide current insights into the research landscape of cloud computing security, focusing on the utilization of DL and ML techniques. This offers a novel perspective on the research area, expanding on previous studies. Additionally, the bibliometric

analysis conducted in this work allows for a systematic evaluation of scientific output and trends, contributing methodologically to the field.

The integration of ML and DL presents promising avenues to bolster the security of cloud computing systems, offering efficient means to detect and counteract security risks across various tasks. Our study highlights key trends in ML's role in cloud security, including anomaly detection, security automation, native security, insider threats, security brokers, and emerging technology roles. The study also highlights key trends in the DL's role in cloud security, including anomaly detection, security automation, native security, image-based detection, network traffic analysis, and emerging technology roles. However, challenges such as integration issues, performance concerns, and data privacy remain. Addressing these challenges requires proper system architecture, appropriate method selection, and the incorporation of explainable AI technologies. By navigating these challenges, researchers, policymakers, and practitioners can harness the full potential of ML and DL to fortify cloud computing security and safeguard data effectively. Researchers can explore different ML and DL techniques for securing cloud computing environments. Moreover, policymakers can incorporate these techniques into regulatory frameworks for enhanced security measures, and practitioners can deploy these solutions to proactively detect and respond to security threats. Future research may focus on automated incident response, image-based malware detection, and ML algorithms for cloud-native security. Future research may also focus on techniques to enhance anomaly detection in complex cloud environments. Additionally, future research in edge/fog computing and IoT-based data processing can enhance security and privacy in cloud environments.

**Author contributions** Author Contributions: Y.A. conceptualized and presented the idea. A.T. developed the theoretical analysis. A.M. supervised the findings of this research. Y.A. took the lead in writing the manuscript with participation of A.T., and A. M. All authors provided critical feedback and help in finalizing the research, results discussion and contributed to the final manuscript. All authors reviewed the manuscript.

**Funding** Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital)

## Declarations

**Competing interests** The authors declare no competing interests.



## References

Abbas Z, Myeong S (2023) Enhancing industrial cyber security, focusing on formulating a practical strategy for making predictions through machine learning tools in cloud computing environment. Electronics 12:2650

Abdel-Basset M, Mohamed M, Chang V (2018) NMCDA: A framework for evaluating cloud computing services. Futur Gener Comput Syst 86:12–29

Abdullahi M, Baashar Y, Alhussian H, Alwadain A, Aziz N, Capretz LF, Abdulkadir SJ (2022) Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review. Electronics 11:198

Abidi I, Kumar V, Sen R (2021) Practical attestation for edge devices running compute heavy machine learning applications. In: Proceedings of the Annual Computer Security Applications Conference (ACSAC), ACM. Austin, Texas, USA, 323–336

Abosata N, Al-Rubaye S, Inalhan G (2023) Customised intrusion detection for an industrial IoT heterogeneous network based on machine learning algorithms called FTL-CID. Sensors 23:321

Abou El Houda Z, Brik B, Ksentini A, Khoukhi L, Guizani M (2022) When federated learning meets game theory: A cooperative framework to secure iiot applications on edge computing. IEEE Trans Indust Inform 18:7988–7997

Abusitta A, Bellaiche M, Dagenais M, Halabi T (2019) A deep learning approach for proactive multi-cloud cooperative intrusion detection system. Futur Gener Comput Syst 98:308–318

Ahmad W, Rasool A, Javed AR, Baker T, Jalil Z (2022) Cyber security in iot-based cloud computing: A comprehensive survey. Electronics 11:16

Ahmed I, Anisetti M, Ahmad A, Jeon G (2023) A multilayer deep learning approach for malware classification in 5G-enabled IIoT. IEEE Trans Industr Inf 19:1495–1503

AlAhmad AS, Kahtan H, Alzoubi YI, Ali O, Jaradat A (2021) Mobile cloud computing models security issues: A systematic review. J Netw Comput Appl 190:103152

Alasmary H, Khormali A, Anwar A, Park J, Choi J, Abusnaina A, Awad A, Nyang D, Mohaisen A (2019) Analyzing and detecting emerging Internet of things malware: A graph-based approach. IEEE Int Things J 6:8977–8988

Aldallal A (1916) Toward efficient intrusion detection system using hybrid deep learning approach. Symmetry 2022:14

Aldhaheri A, Alwahedi F, Ferrag MA, Battah A (2023) Deep learning for cyber threat detection in IoT networks: A review. Int Things Cyber-Phys Syst 4:110–128

Alkadi O, Moustafa N, Turnbull B, Choo K-KR (2020) A deep blockchain framework-enabled collaborative intrusion detection for protecting IoT and cloud networks. IEEE Internet Things J 8:9463–9472

Aloqaily M, Otoum S, Al Ridhawi I, Jararweh Y (2019) An intrusion detection system for connected vehicles in smart cities. Ad Hoc Netw 90:101842

Alrashdi I, Alqazzaz A, Aloufi E, Alharthi R, Zohdy M, Ming H (2019) Ad-iot: Anomaly detection of iot cyberattacks in smart city using machine learning. In: Proceedings of the 9th Annual Computing and Communication Workshop and Conference (CCWC), IEEE. Las Vegas, NV, USA, pp 0305–0310

Alzoubi YI, Aljaafreh A (2023) Blockchain-fog computing integration applications: A systematic review. Cyber Inform Technol 23:3–37

Alzoubi Y, Gill A (2021) The critical communication challenges between geographically distributed agile development teams: Empirical findings. IEEE Trans Prof Commun 64:322–337

Alzoubi YI, Osmanaj VH, Jaradat A, Al-Ahmad A (2021) Fog computing security and privacy for the internet of thing applications: State-of-the-art. Secur Privacy 4:e145

Alzoubi YI, Gill A, Mishra A (2022a) A systematic review of the purposes of blockchain and fog computing integration: Classification and open issues. J Cloud Comput 11:1–36

Alzoubi YI, Al-Ahmad A, Kahtan H (2022b) Blockchain technology as a Fog computing security and privacy solution: An overview. Comput Commun 182:129–152

Alzoubi YI, Al-Ahmad A, Kahtan H, Jaradat A (2022c) Internet of things and blockchain integration: Security, privacy, technical, and design challenges. Future Int 14:216

Alzoubi YI, Topcu AE, Erkaya AE (2023) Machine learning-based text classification comparison: Turkish language context. Appl Sci 13:9428

Amin J, Anjum MA, Ibrar K, Sharif M, Kadry S, Crespo RG (2023) Detection of anomaly in surveillance videos using quantum convolutional neural networks. Image vis Comput 135:104710

Amiri Z, Heidari A, Navimipour NJ, Unal M, Mousavi A (2024) Adventures in data analysis: A systematic review of Deep Learning techniques for pattern recognition in cyber-physical-social systems. Multimedia Tools Applic 83:22909–22973

Angelopoulos A, Michailidis ET, Nomikos N, Trakadas P, Hatziefremidis A, Voliotis S, Zahariadis T (2020) Tackling faults in the industry 4.0 era—a survey of machine-learning solutions and key aspects. Sensors 20:109

Balamurugan V, Saravanan R (2019) Enhanced intrusion detection and prevention system on cloud environment using hybrid classification and OTS generation. Clust Comput 22:13027–13039
Balobaid AS, Shaik S, Komandur S (2024) A review on cyber security issues in IoT-based cloud computing. Intl J Intell Syst Appl Eng 12:278–285
Belal MM, Sundaram DM (2022) Comprehensive review on intelligent security defences in cloud: Taxonomy, security issues, ML/DL techniques, challenges and future trends. J King Saud Univ-Comput Inform Sci 34:9102–9131
Bhattacharya P, Tanwar S, Bodkhe U, Tyagi S, Kumar N (2019) Bindaas: Blockchain-based deep-learning as-a-service in healthcare 4.0 applications. IEEE Trans Net Sci Eng 8:1242–1255
Bhattacherjee SS, Kumar NS, Rajalakshmi P (2019) Emotion detection IoT enabled edge-node for citizen security. In: Proceedings of the 5th World Forum on Internet of Things (WF-IoT), IEEE. Limerick, Ireland, pp 925–930
Bhuvaneshwari A, Kaythry P (2023) A review of deep learning strategies for enhancing cybersecurity in networks: Deep learning strategies for enhancing cybersecurity. J Sci Ind Res 82:1316–1330
Brown P, Brown A, Gupta M, Abdelsalam M (2022) Online malware classification with system-wide system calls in cloud iaas. In: Proceedings of the 23rd International Conference on Information Reuse and Integration for Data Science (IRI), IEEE. San Diego, CA, USA, pp 146–151
Butt UA, Mehmood M, Shah SBH, Amin R, Shaukat MW, Raza SM, Suh DY, Piran MJ (2020) A review of machine learning algorithms for cloud computing security. Electronics 9:1379
Chen C, Liu B, Wan S, Qiao P, Pei Q (2020) An edge traffic flow detection scheme based on deep learning in an intelligent transportation system. IEEE Trans Intell Transp Syst 22:1840–1852
Chkirbene Z, Erbad A, Hamila R, Gouissem A, Mohamed A, Hamdi M (2020) Machine learning based cloud computing anomalies detection. IEEE Network 34:178–183
CybercrimeMagazine (2024) Cybercrime Magazine. https://cybersecurityventures.com/, accessed 2 April 2024.
Dai Y, Xu D, Zhang K, Maharjan S, Zhang Y (2020) Deep reinforcement learning and permissioned blockchain for content caching in vehicular edge computing and networks. IEEE Trans Veh Technol 69:4312–4324
Dalal S, Manoharan P, Lilhore UK, Seth B, Simaiya S, Hamdi M, Raahemifar K (2023) Extremely boosted neural network for more accurate multi-stage Cyber attack prediction in cloud computing environment. J Cloud Comput 12:1–22
Dasgupta D, Akhtar Z, Sen S (2022) Machine learning in cybersecurity: A comprehensive survey. J Defense Model Simul 19:57–106
Dey S, Ye Q, Sampalli S (2019) A machine learning based intrusion detection scheme for data fusion in mobile clouds involving heterogeneous client networks. Inform Fusion 49:205–215
Doriguzzi-Corin R, Millar S, Scott-Hayward S, Martinez-del-Rincon J, Siracusa D (2020) LUCID: A practical, lightweight deep learning solution for DDoS attack detection. IEEE Trans Netw Serv Manage 17:876–889
Dovom EM, Azmoodeh A, Dehghantanha A, Newton DE, Parizi RM, Karimipour H (2019) Fuzzy pattern tree for edge malware detection and categorization in IoT. J Syst Architect 97:1–7
Ferrag MA, Babaghayou M, Yazici MA (2020) Cyber security for fog-based smart grid SCADA systems: Solutions and challenges. J Inform Secur Appl 52:102500
Ferrag MA, Friha O, Hamouda D, Maglaras L, Janicke H (2022) Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning. IEEE Access 10:40281–40306
Gartner (2023) Gartner research. gartner.com. https://www.gartner.com/en/information-technology/research/research-index. Viewed 8 December 2023.
Ge M, Syed NF, Fu X, Baig Z, Robles-Kelly A (2021) Towards a deep learning-driven intrusion detection approach for Internet of Things. Comput Netw 186:107784
Ghazal TM, Hasan MK, Alshurideh MT, Alzoubi HM, Ahmad M, Akbar SS, Al Kurdi B, Akour IA (2021) IoT for smart cities: Machine learning approaches in smart healthcare—A review. Future Int 13:218
Goh GD, Sing SL, Yeong WY (2021) A review on machine learning in 3D printing: Applications, potential, and challenges. Artif Intell Rev 54:63–94
Gu T, Liu K, Dolan-Gavitt B, Garg S (2019) Badnets: Evaluating backdooring attacks on deep neural networks. IEEE Access 7:47230–47244
Gupta C, Johri I, Srinivasan K, Hu Y-C, Qaisar SM, Huang K-Y (2017) A systematic review on machine learning and deep learning models for electronic information security in mobile networks. Sensors 2022:22
Gupta M, Abdelsalam M, Khorsandroo S, Mittal S (2020) Security and privacy in smart farming: Challenges and opportunities. IEEE Access 8:34564–34584

Gupta I, Singh AK, Lee C-N, Buyya R (2022) Secure data storage and sharing techniques for data protection in cloud environments: A systematic review, analysis, and future directions. IEEE Access 10:71247–71277
Hameed SS, Selamat A, Abdul Latiff L, Razak SA, Krejcar O, Fujita H, Ahmad Sharif MN, Omatu S (2021) A hybrid lightweight system for early attack detection in the IoMT fog. Sensors 21:8289
Hassija V, Chamola V, Saxena V, Jain D, Goyal P, Sikdar B (2019) A survey on IoT security: Application areas, security threats, and solution architectures. IEEE Access 7:82721–82743
Heidari A, Jabraeil Jamali MA (2023) Internet of Things intrusion detection systems: a comprehensive review and future directions. Clust Comput 26:3753–3780
Heidari A, Navimipour NJ, Unal M (2022) Applications of ML/DL in the management of smart cities and societies based on new trends in information technologies: A systematic literature review. Sustain Cities Soc 85:104089
Heidari A, Navimipour NJ, Unal M (2023) A secure intrusion detection platform using blockchain and radial basis function neural networks for internet of drones. IEEE Internet Things J 10:8445–8454
Hernandez-Jaimes ML, Martinez-Cruz A, Ramírez-Gutiérrez KA, Feregrino-Uribe C (2023) Artificial intelligence for IoMT security: a review of intrusion detection systems, attacks, datasets and Cloud-Fog-Edge architectures. Internet of Things 23:100887
Herrera-Franco G, Montalván-Burbano N, Carrión-Mero P, Apolo-Masache B, Jaya-Montalvo M (2020) Research trends in geotourism: A bibliometric analysis using the scopus database. Geosciences 10:379
Hossain MS, Muhammad G, Guizani N (2020) Explainable AI and mass surveillance system-based healthcare framework to combat COVID-I9 like pandemics. IEEE Network 34:126–132
Huang H, Wang Y, Zong H (2022) Support vector machine classification over encrypted data. Appl Intell 52:5938–5948
IACSIT (2024) The international association of computer science and information technology. https://www.iacsit.org/, accessed 2 April 2024.
IDC. Idc marketscape worldwide managed security services 2020 vendor assessment. International Data Coorporation. https://www.idc.com/getdoc.jsp?containerId=US46235320. Viewed 15 December 2023. 2020.
ISSA (2024) The information systems security association. https://www.issa.org/, accessed 2 April 2024.
Jaber AN, Rehman SU (2020) FCM–SVM based intrusion detection system for cloud computing environment. Clust Comput 23:3221–3231
Jia Y, Zhong F, Alrawais A, Gong B, Cheng X (2020) Flowguard: An intelligent edge defense mechanism against IoT DDoS attacks. IEEE Int Things J 7:9552–9562
Kasongo SM (2023) A deep learning technique for intrusion detection system using a recurrent neural networks based framework. Comput Commun 199:113–125
Khudzari JM, Kurian J, Tartakovsky B, Raghavan GV (2018) Bibliometric analysis of global research trends on microbial fuel cells using Scopus database. Biochem Eng J 136:51–60
Kornaros G (2022) Hardware-assisted machine learning in resource-constrained IoT environments for security: Review and future prospective. IEEE Access 10:58603–58622
Krishnan P, Duttagupta S, Achuthan K (2019) VARMAN: Multi-plane security framework for software defined networks. Comput Commun 148:215–239
Kumar P, Kumar R, Srivastava G, Gupta GP, Tripathi R, Gadekallu TR, Xiong NN (2021) PPSF: A privacy-preserving and secure framework using blockchain-based machine-learning for IoT-driven smart cities. IEEE Trans Net Sci Eng 8:2326–2341
Kumar A, Umurzoqovich RS, Duong ND, Kanani P, Kuppusamy A, Praneesh M, Hieu MN (2022) An intrusion identification and prevention for cloud computing: from the perspective of deep learning. Optik 270:170044
Kumar S, Prethi KA, Singh S, Lourens M, Patil N (2022) Role of machine learning in managing cloud computing security. In Proceedings of the 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), IEEE. Greater Noida, India, 2366–2369
Lahande PV, Kaveri PR (2022) Reinforcement learning applications for performance improvement in cloud computing—A systematic review. In: Aurelia S, Hiremath SS, Subramanian K, Biswas SK (eds) Sustainable advanced computing: lecture notes in electrical engineering, vol 840. Springer, Singapore, pp 91–112
Landman T, Nissim N (2021) Deep-Hook: A trusted deep learning-based framework for unknown malware detection and classification in Linux cloud environments. Neural Netw 144:648–685
Lei W, Pang Z, Wen H, Hou W, Han W (2022) FDI attack detection at the edge of smart grids based on classification of predicted residuals. IEEE Trans Industr Inf 18:9302–9311

Lin H, Xue Q, Feng J, Bai D (2022) Internet of things intrusion detection model and algorithm based on cloud computing and multi-feature extraction extreme learning machine. Digital Communications and Networks, In Press, https://doi.org/10.1016/j.dcan.2022.09.021
Liu D, Shen J, Wang A, Wang C (2020) Secure real-time image protection scheme with near-duplicate detection in cloud computing. J Real-Time Image Proc 17:175–184
Ma X, Wu J, Xue S, Yang J, Zhou C, Sheng QZ, Xiong H, Akoglu L (2023) A comprehensive survey on graph anomaly detection with deep learning. IEEE Trans Knowl Data Eng 35:12012–12038
Mahdavifar S, Ghorbani AA (2019) Application of deep learning to cybersecurity: A survey. Neurocomputing 347:149–176
Makkar A, Ghosh U, Sharma PK (2021) Artificial intelligence and edge computing-enabled web spam detection for next generation IoT applications. IEEE Sens J 21:25352–25361
MarketsandMarkets (2023) Artificial intelligence (AI) market. MarketsandMarkets.com. https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-market-74851580.html. Accessed 5 Jan 2024
Miles MB, Huberman AM (1994) Qualitative data analysis: An expanded sourcebook. Sage, Beverly Hills, USA
Mishra S, Sharma SK, Alowaidi MA (2020) Multilayer self-defense system to protect enterprise cloud. Comput, Mater Contin 66:71–85
Mishra A, Alzoubi YI, Gill AQ, Anwar MJ (2022b) Cybersecurity enterprises policies: A comparative study. Sensors 22:538
Mishra A, Alzoubi YI, Anwar MJ, Gill AQ (2022c) Attributes impacting cybersecurity policy development: An evidence from seven nations. Comput Secur 120:102820
Mishra A, Jabar TS, Alzoubi YI, Mishra KN (2023) Enhancing privacy-preserving mechanisms in Cloud storage: A novel conceptual framework. Concurr Comput: Pract Exp 35:e7831
Mishra DP (2024) A review of cloud security solutions: Leveraging machine learning and deep learning techniques. Available at SSRN 4704331, SSRN: https://ssrn.com/abstract=4704331. Accessed 5 Jan 2024
Mishra S, Tyagi AK (2022) Emerging trends and techniques in machine learning and Internet of things-based cloud applications. In: Tyagi AK, Sreenath N (eds) Handbook of research of internet of things and cyber-physical systems, 1st edn. Apple Academic Press: CRC Press. Taylor Francis Group, pp 149–167
Mishra N, Singh R, Yadav S (2022) Detection of DDoS vulnerability in cloud computing using the perplexed bayes classifier. Computational Intelligence and Neuroscience 2022
Morgan, S (2022) Boardroom cybersecurity 2022 report. Cybercrime Magazine. https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021/. Viewed 31 November 2023.
Mostafa SA, Ahmad IA (2018) Recent developments in systematic sampling: A review. J Stat Theor Pract 12:290–310
Mothukuri V, Parizi RM, Pouriyeh S, Huang Y, Dehghantanha A, Srivastava G (2021a) A survey on security and privacy of federated learning. Futur Gener Comput Syst 115:619–640
Mothukuri V, Khare P, Parizi RM, Pouriyeh S, Dehghantanha A, Srivastava G (2021b) Federated-learning-based anomaly detection for IoT security attacks. IEEE Internet Things J 9:2545–2554
Moustafa N (2021) A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets. Sustain Cities Soc 72:102994
Moustafa N, Hu J, Slay J (2019) A holistic review of network anomaly detection systems: A comprehensive survey. J Netw Comput Appl 128:33–55
Nahmias D, Cohen A, Nissim N, Elovici Y (2020) Deep feature transfer learning for trusted and automated malware signature generation in private cloud environments. Neural Netw 124:243–257
Nair AK, Sahoo J, Raj ED (2023) Privacy preserving federated learning framework for IoMT based big data analysis using edge computing. Computer Standards & Interfaces 86:103720
Narayanan E, Muthukumar B (2022) A machine learning framework for providing data integrity and confidentiality for sensitive data cloud applications. Int J Syst Assur Eng Manag. https://doi.org/10.1007/s13198-022-01741-y,1-12
Nissim N, Lahav O, Cohen A, Elovici Y, Rokach L (2019) Volatile memory analysis using the MinHash method for efficient and secured detection of malware in private cloud. Comput Secur 87:101590
Oltsik J (2022) Technology perspectives from cybersecurity professionals. https://www.issa.org/. Viewed 31 November 2023.
Pandey BK, Veeramanickam M, Ahmad S, Rodriguez C, Esenarro D (2023) ExpSSOA-deep maxout: Exponential shuffled shepherd optimization based deep maxout network for intrusion detection using big data in cloud computing framework. Comput Secur 124:102975

Parra GDLT, Rad P, Choo K-KR, Beebe N (2020) Detecting Internet of things attacks using distributed deep learning. J Netw Comput Appl 163:102662
PonemonInstitute (2024) Advancing responsible information management. https://www.ponemon.org/, accessed 2 April 2024.
Prasad VM, Bharathi B (2022) A novel trust negotiation protocol for analysing and approving IoT edge computing devices using machine learning algorithm. Int J Comput Net Appl 9:712–723
Priya S, Maddikunta PKR, Parimala M, Koppu S, Gadekallu TR, Chowdhary CL, Alazab M (2020) An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture. Comput Commun 160:139–149
Quraishi SJ (2022) Machine learning approach for cloud computing security. In: Proceedings of the 3rd International Conference on Intelligent Engineering and Management (ICIEM), IEEE. London, United Kingdom, pp 158–163
Rahila S, Khonde S (2019) SAAS: Attack detection and prevention with forensic in cloud environment. Intl J Innov Technol Exploring Eng 9:1199–1203
Ramchandran A, Sangaiah AK (2020) Unsupervised deep learning system for local anomaly event detection in crowded scenes. Multimedia Tools Appl 79:35275–35295
Rao SM, Jain A (2024) Advances in malware analysis and detection in cloud computing environments: A review. Int J Safety Secur Eng 14:225
Rathore S, Kwon BW, Park JH (2019) BlockSecIoTNet: Blockchain-based decentralized security architecture for IoT network. J Netw Comput Appl 143:167–177
Ravi N, Shalinie SM (2020) Learning-driven detection and mitigation of DDoS attack in IoT via SDN-cloud architecture. IEEE Int Things J 7:3559–3570
Ravinder M, Kulkarni V (2023) A review on cyber security and anomaly detection perspectives of smart grid. In: Proceedings of the 5th international conference on smart systems and inventive technology (ICSSIT), IEEE. Tirunelveli, India, pp 692–697
Rjoub G, Bentahar J, Abdel Wahab O, Saleh Bataineh A (2021) Deep and reinforcement learning for automated task scheduling in large-scale cloud computing systems. Concurr Comput: Pract Experience 33:e5919
Rosero D, Díaz N, Trujillo C (2021) Cloud and machine learning experiments applied to the energy management in a microgrid cluster. Appl Energy 304:117770
Roy I, Modak R, Ghosh E, Rahaman SN, Chatterjee S, Majumder K, Shaw RN, Ghosh A (2023) A review on machine learning based security in edge computing environment. In Advanced Communication and Intelligent Systems. ICACIS 2023. Communications in Computer and Information Science, Shaw, R.N., Paprzycki, M., Ghosh, A., Eds.; Springer, Cham, 1921, 120–137.
Salem IE, Al-Saedi KH (2023) Enhancing cloud security through the integration of deep learning and data mining techniques: A comprehensive review. Period Eng Nat Sci 11:176–192
SANS (2024) The most trusted source for cyber security training, certification and research. https://www.sans.org/mlp/middle-east-turkey-africa/, accessed 2 April 2024.
Saran M, Yadav RK, Tripathi UN (2022) Machine learning based security for cloud computing: A survey. Int J Appl Eng Res 17:332–337
Saranya T, Sridevi S, Deisy C, Chung TD, Khan MA (2020) Performance analysis of machine learning algorithms in intrusion detection system: A review. Proc Comput Sci 171:1251–1260
Sarker IH, Khan AI, Abushark YB, Alsolami F (2023) Internet of things (iot) security intelligence: A comprehensive overview, machine learning solutions and research directions. Mob Net Appl 28:296–312
Spitzner L (2022) SANS 2022 security awareness report. SANS. https://www.sans.org/blog/sans-2022-security-awareness-report/. Viewed 14 December 2023.
Sugi SSS, Ratna SR (2020) A novel distributed training on fog node in IoT backbone networks for security. Soft Comput 24:18399–18410
Sweileh WM (2018) Research trends on human trafficking: A bibliometric analysis using Scopus database. Glob Health 14:106
Thakkar A, Lohiya R (2021) A review on machine learning and deep learning perspectives of IDS for IoT: recent updates, security issues, and challenges. Arch Comput Methods Eng 28:3211–3243
Thilagam T, Aruna R (2021) Intrusion detection for network based cloud computing by custom RC-NN and optimization. ICT Express 7:512–520
Tian Z, Luo C, Qiu J, Du X, Guizani M (2020) A distributed deep learning system for web attack detection on edge devices. IEEE Trans Industr Inf 16:1963–1971
Tian P, Chen Z, Yu W, Liao W (2021) Towards asynchronous federated learning based threat detection: A DC-Adam approach. Comput Secur 108:102344

Topcu AE, Alzoubi YI, Elbasi E, Camalan E (2023) Social media zero-day attack detection using TensorFlow. Electronics 12:3554

Tuli S, Basumatary N, Gill SS, Kahani M, Arya RC, Wander GS, Buyya R (2020) HealthFog: An ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated IoT and fog computing environments. Futur Gener Comput Syst 104:187–200

Ullah F, Srivastava G, Ullah S (2022) A malware detection system using a hybrid approach of multi-heads attention-based control flow traces and image visualization. J Cloud Comput 11:1–21

Vinolia A, Kanya N, Rajavarman V (2023) Machine learning and deep learning based intrusion detection in cloud environment: a review. In: Proceedings of the 5th International Conference on Smart Systems and Inventive Technology, IEEE. Tirunelveli, India, pp 952–960.

Vu L, Nguyen QU, Nguyen DN, Hoang DT, Dutkiewicz E (2022) Deep generative learning models for cloud intrusion detection systems. IEEE Trans Cybern 53:565–577

Wu M, Song Z, Moon YB (2019) Detecting cyber-physical attacks in CyberManufacturing systems with machine learning methods. J Intell Manuf 30:1111–1123

Wu H, Li X, Deng Y (2020) Deep learning-driven wireless communication for edge-cloud computing: opportunities and challenges. J Cloud Comput 9:1–14

Wu Z, Wang L, Xu Z, Li H, Yang J (2022) GPU virtualization technology and security issues: A survey. J Cyber Secur 7:30–58

Yao H, Gao P, Zhang P, Wang J, Jiang C, Lu L (2019) Hybrid intrusion detection system for edge-based IIoT relying on machine-learning-aided detection. IEEE Network 33:75–81

Yi T, Chen X, Zhu Y, Ge W, Han Z (2023) Review on the application of deep learning in network attack detection. J Netw Comput Appl 212:103580

Yıldız M (2023) History of machine learning. https://clarusway.com/history-of-machine-learning/, accessed 6 January 2024.

Yu S, Chen X, Zhou Z, Gong X, Wu D (2021) When deep reinforcement learning meets federated learning: Intelligent multitimescale resource management for multiaccess edge computing in 5G ultradense network. IEEE Int Things J 8:2238–2251

Zhang L, Lai S, Xia J, Gao C, Fan D, Ou J (2022) Deep reinforcement learning based IRS-assisted mobile edge computing under physical-layer security. Physical Commun 55:101896

Zhou X, Xu X, Liang W, Zeng Z, Yan Z (2021) Deep-learning-enhanced multitarget detection for end–edge–cloud surveillance in smart IoT. IEEE Int Things J 8:12588–12596

Zhou Y, Wang R, Mo X, Li Z, Tang T (2023) Robust hierarchical federated learning with anomaly detection in cloud-edge-end cooperation networks. Electronics 12:112