Towards Effective Anomaly Detection: Machine Learning Solutions in Cloud Computing

Hussain Almajed, Abdulrahman Alsaqer, Abdullah Albuali
Department of Computer Networks and Communications
College of Computer Sciences and Information Technology
King Faisal University
Al-Ahsa, 31982 Saudi Arabia

***Abstract*—Cloud computing has transformed modern Information Technology (IT) infrastructures with its scalability and cost-effectiveness but introduces significant security risks. Moreover, existing anomaly detection techniques are not well equipped to deal with the complexities of dynamic cloud environments. This systematic literature review shows the advancements in Machine Learning (ML) solutions for anomaly detection in cloud computing. The study categorizes ML approaches, examines the datasets and evaluation metrics utilized, and discusses their effectiveness and limitations. We analyze supervised, unsupervised, and hybrid ML models showing their advantages in dealing with a certain threat vector. It also discusses how advanced feature engineering, ensemble learning and real-time adaptability can improve detection accuracy and reduce false positives. Some key challenges, such as dataset diversity and computational efficiency, are highlighted, along with future research directions to improve ML based anomaly detection for robust and adaptive cloud security. Hybrid approaches are found to increase the accuracy reaching up to 99.85% and reduces the number of false positives. This review provides a comprehensive guide to researchers aiming to enhance anomaly detection in cloud environments.**

***Keywords*—Anomaly; cloud; machine learning; detection**

I. Introduction

An important part of modern IT infrastructure today is cloud computing, which offers flexible, scalable and cost effective solutions for businesses and individuals over the internet [1],[2]. Offers an on demand access to computing resources such as servers, storage, databases, and applications, which can be rapidly provisioned and released with minimal management effort. Cloud computing has become widely adopted in different industries from healthcare to finance, entertainment to education due to its benefits. Despite these advantages, the use of cloud services has introduced security challenges that must be addressed to guarantee the reliability and trustworthiness of cloud systems [3].

Back in 2022, a leading health insurer in Australia called Medibank stored sensitive information about 10 million customer accounts in its cloud based systems, and the unlucky company suffered a massive data breach which revealed all their customer's data [4], [5]. While the company refused to pay a ransom, hackers then started to leak the stolen data on dark web forums. It highlights the critical requirement for better anomaly detection in cloud environments. In addition, cyberattacks on cloud computing environments are becoming more prevalent and cause significant data breaches. The number of breaches within the cloud environment also increased, from 35% of businesses in 2022 to 39% at 2023 [6].

Cloud environments are dynamic and elastic, which makes them vulnerable to attacks from malicious actors, therefore effective anomaly detection is crucial to keep cloud environment secure. The cloud infrastructure is by nature shared where many tenants may use the same physical resources, making the risk of potential security breaches even higher [7]. Furthermore, managing various virtualized environments, and the constant scaling of resources, makes it difficult to establish a stable security baseline. In these environments security threats can vary from external threat, like Distributed Denial of Service (DDoS) attacks to internal threat, losing control over an insider, unauthorized access, data breaching and configuration errors.

Anomaly detection is identifying unusual patterns or behaviour which may indicate security breach, failure of system or performance issues [8] [9]. An effective anomaly detection method is necessary to reduce the impact of these threats by providing timely and proactive responses. Traditional rule based detection methods based on predefined signatures or rules struggle to keep pace with the complexity and evolution of cloud environments. These methods are unable to detect previously unseen or novel threats, particularly with the diversity and scale of cloud services. Attack patterns evolve rapidly and the cloud environment is always changing, for that we need more adaptive and more intelligent approaches [10].

ML steps in when it provides advanced algorithmic tools to process huge amounts of data and react to new threats by identifying anomalies [11] [12]. Anomaly detection with ML has the capability to offer more secure cloud systems by means of automated, intelligent monitoring of cloud systems. Because ML techniques learn from data, find complex relationships and get better over time, they are particularly well suited to the cloud [13]. With these characteristics, ML is a promising approach to identifying security anomalies that would likely be missed by traditional approaches.

Therefore, cloud computing has revolutionized how the organizations manage and access to the IT resources by providing many advantages and at the same time introducing various security issues. Anomaly detection helps to identify abnormal activities in the cloud, which can indicate abnormal threats. Anomaly detection is a powerful capability of ML which improves the security posture of cloud environments with

adaptive, data driven techniques. In this systematic literature review, we will explore in detail the current state of the art of ML based anomaly detection in cloud computing, the challenges faced and future research directions to address the evolving threat landscape.

Following this introduction, Section II provides background information on cloud computing, anomaly detection, and a definition of ML. Section V describes the research methodology used in this study. Section VI reviews related work in the field. After that, Section VII will illustrate a case study regarding the research study. In Section VIII, the results of the review will be presented and discussed. Section IX then highlights key challenges and outlines open directions for future research. Finally, Section X synthesizes the key findings and concludes the study.

II. Background

A. Cloud Computing Overview

Cloud computing is an on demand network access to a shared pool of configurable computing resources such as a network, applications, servers, storage, or services, in which the providers deliver the resources on demand because they are-scalable, elastic and vary as per your need [14], [15]. The next section represents the cloud service models of the cloud and the cloud deployment models.

B. Cloud Service Model

There are three main service models for cloud computing which offer a varying level of control, flexibility and management. Fig. 1 shows the different cloud service models with examples.

- Infrastructure as a Service (IaaS): The model virtualizes these resources on demand. Users can deploy applications and the configuration settings, without managing or controlling the underlying cloud infrastructure.
- Platform as a Service (PaaS): This model provides a development setting where developers create and deploy applications without having to know how many processors.
- Software as a Service (SaaS): This model is used to deliver the applications over the web where they are consumed by the consumers through web portals.

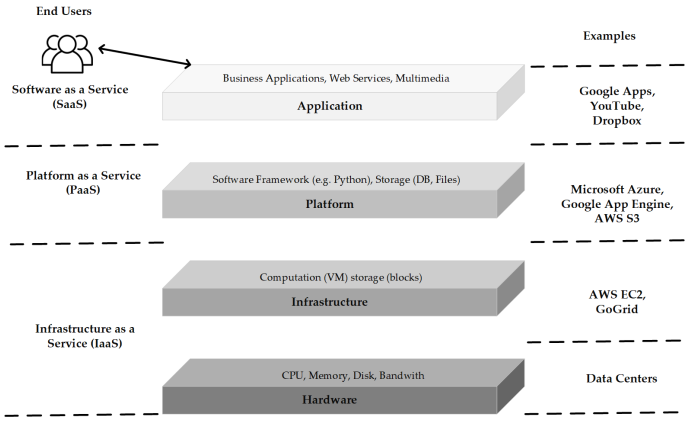

Fig. 1. Cloud service models with examples.

C. Cloud Deployment

The cloud deployment models as shown in Fig. 2 define the way by which cloud resources are managed and offered to the users. There are four main types:

- Private Cloud: The cloud infrastructure is delivered solely for use by one organization (a business unit) that consists of multiple consumers [15].
- Public Cloud: The cloud infrastructure is delivered for general public open use. It may be managed, owned, and operated by a corporation, academic or government organization, or a mix.
- Community Cloud: It is a cloud computing environment where multiple organizations with similar goals and security requirements share a cloud.
- Hybrid Cloud: Two or more distinct cloud infrastructures (private, community, or public) that have been independently operated and connected using standard or proprietary technology.

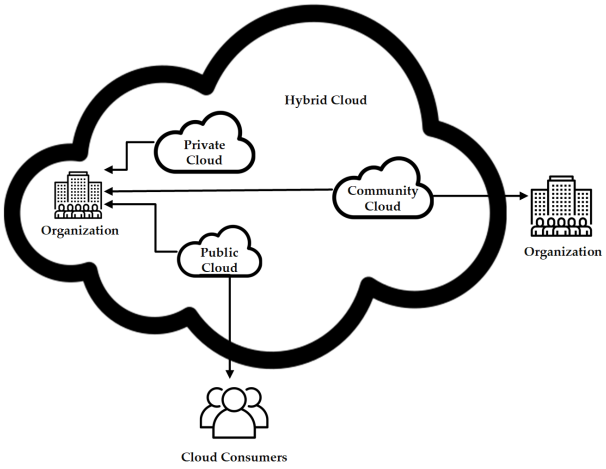

Fig. 2. Cloud deployment models.

D. Cloud Threat

It any event, situation or action which could lead to compromise of the confidentiality, integrity, or availability of cloud computing resources, data or services, including access, data breaches, service disruptions or other security violations, whether deliberate or accidental, that affect the security and trustworthiness of cloud computing environments [16]. The cloud threat can be categorized based Confidentiality,Integrity, and Availability (CIA) tried in the next section.

1) Classification of threat-based CIA:

- Confidentiality: There are many confidentiality threats in cloud computing environments. When unauthorized access of sensitive data stored in the cloud leads to data breaches and privacy violations, it is called Data Breaches [17], [18]. Moreover, Shared Technology Vulnerabilities e.g. hypervisor vulnerabilities and cross Virtual Machine (VM) side channel attacks are at risk as a result of shared infrastructure and multi tenancy [19].
- Integrity: Data tampering is one of the integrity threats in cloud computing, which refers to the unauthorized modifications of data stored in cloud which adversely affects the accuracy and reliability of data. Data integrity can also be compromised by Malicious Insiders, system administrators, and former employee. Application Programming Interface (APIs) with such vulnerabilities are insecure, which can easily be exploited in data manipulation.
- Availability: Denial of Service (DoS) Attacks threaten the availability of cloud computing by overwhelming cloud services with traffic making the cloud services unavailable to the legitimate users.

2) Attacks in cloud: The cloud can be attacked by diverse attacks and for better understanding security threats and vulnerabilities in cloud computing can be broadly classified into five main categories application-based,storage-based, VM-based network-based, Identity and Access Management. Fig. 3 shows list of some attacks based in the five categories.

1) Network-based attacks: Attacks related to network communications and configurations in cloud computing environments [20], [21]. Examples such as flooding attacks, Structured Query Language (SQL) injection attack, spoofing.
2) VM-Based Attacks: Attacks related to virtualization technology and hypervisor vulnerabilities. Examples such as Hypervisor Vulnerabilities, VM Escape, and VM Image Sprawl.
3) Storage-based Attacks: Attacks related to data confidentiality, integrity, and availability in cloud storage. Examples include Data Breaches, Data Isolation, and Data Backup and Redundancy.
4) Application-Based Attacks: The attacks are aimed at cloud infrastructure running applications. Examples include web services, malware infusion, and shared design vulnerabilities.
5) Identity and Access Management based attacks: Threats related to managing identities and providing secure and efficient access to data [18]. Examples include Identity Management, authorization, authentication, access control, and federation management.

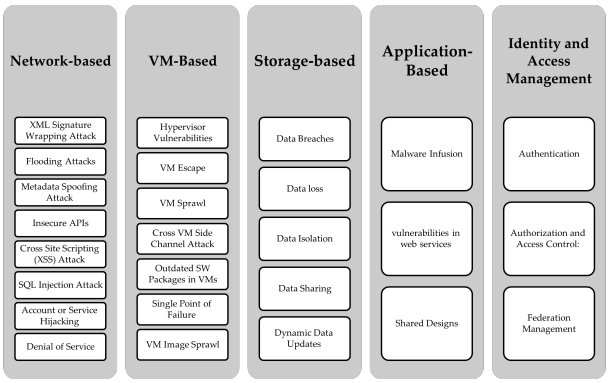

Fig. 3. List of cloud attacks based in category.

E. Anomaly Detection Definition

Finding data points, patterns or behaviors that are very different from the norm is known as anomaly detection [22]. When considering computing and cybersecurity, anomalies are indicative of possible issues, including security breaches, system failures and fraudulent activities or unusual behavior. Anomalies are outliers basically, data that doesn't conform to expected patterns or historical trends. Anomaly detection is used to detect these deviations early, so that organizations can take timely corrective action. Anomaly detection is important for reliability, security, and minimizing operational risk in modern systems and in the face of the variety of data and system complexity. Anomaly detection can be applied to a wide range of domains such as finance, healthcare, industrial monitoring and cloud computing in which real time and accurate anomaly detection can prevent a major loss or damage [13].

1) Anomaly Detection Techniques: The techniques can be categorized into three main types: statistical methods, ML techniques, and hybrid techniques.

- Statistical Methods: These techniques use statistical models to specify what normal behavior of a system is. Statistical methods build a baseline distribution from historical data, and flag any data point outside this distribution as an anomaly. Three common statistical techniques are z-score analysis, hypothesis testing and time series modeling [23]. These are easy to implement statistical methods, but can have difficulty with high dimensional data and in capturing complex patterns in dynamic environments.
- ML Approaches: Anomaly detection has gained popularity with ML due to its ability to learn directly from data while adjusting to evolving patterns. A variety of models are used in ML approaches to understand normal behavior and identify deviations which could indicate anomalies. These models are very useful in dynamic environments such as cloud environments where traditional models can fail to capture evolving patterns.
- Hybrid Techniques: Statistical, ML, or other domain specific anomaly detection methods could be combined to improve accuracy, robustness. Hybrid methods combine the strengths of different methods, filling the shortcomings of single techniques like increasing detection accuracy or reducing false positives. A hybrid approach may be a statistical model to identify potential anomalies, and apply a ML algorithm to validate it further. In complex environments like the cloud, these methods are effective where adaptive

learning and baseline modeling are both necessary to cope with dynamic changes.

Choice of anomaly detection technique is conditioned by data nature, availability of labeled data set, system complexity and desired detection accuracy and computational efficiency trade off. Each technique has pros and cons, and in most practical cases, a set of techniques are combined to obtain the best performance in real world settings.

F. Machine Learning

ML is a field of artificial intelligence concerned with enabling systems to learn and act based on data. The basic idea falls under training models to come up with patterns and predict without being explicitly programmed for the task that is required [24]. Due to its capacity to learn and get better with time, ML is now a must in many domains. ML is one of the important roles in cloud computing to improve security by automated detection of unusual or potentially malicious behaviors. ML models operating off large datasets can offer advanced and intelligent solutions to complicated issues such as anomaly detection.

G. Type of ML

Depending on their way of learning and the types of tasks, ML can be divided into various types.

1) Supervised Learning: Supervised learning is a type of ML where a model is trained using a labeled dataset, where input data is given along with corresponding outputs or labels [25]. By identifying patterns in the data, the model learns to map inputs to outputs. Supervised learning is very powerful in the case of anomaly detection in cloud computing, if there are a lot of labeled normal and anomalous behaviors [26]. For example, ML classification techniques commonly applied include classifications like Decision Trees (DT), Support Vector Machines (SVM), and Random Forests (RF), etc to classify activities as normal or anomalies. The major drawback of supervised learning, is that it is very hard to collect a large number of labeled data that are representative of rare events such as security breaches or insider threats [27]. Table I shows some of the supervised models.

2) Unsupervised Learning: It is a ML approach that does not demand labeled datasets. It does not look for normal behavior, nor does it look for anomalous behavior, instead it looks for patterns or groupings in the data without prior knowledge of what is normal and what is anomalous [25]. These techniques work by identifying deviation from known patterns, and are therefore particularly well suited to detecting new and previously unseen threats. One of the problems with unsupervised learning is that it is hard to tell the difference between benign deviations and actual security events without labeled data, and as a result, can have very high false positive rates. Table II shows some of the unsupervised models.

3) Reinforcement Learning (RL): It is learning what to do and how to map cases to actions to maximize a numerical reward signal. Unlike supervised and unsupervised learning, rather it is trial and error based sequential decision making with the agent's goal being to maximize cumulative rewards by taking the best possible actions [25]. RL can be used for developing adaptive security systems in cloud computing for anomaly detection, which models optimal reaction to threats as time elapsed. One of the reasons that RL is particularly appealing for cloud security is that it allows for real time adaptation to new and evolving threats [28]. RL has the potential to help improve the robustness and adaptability of cloud based anomaly detection systems.

TABLE I. LIST OF SUPERVISED ML MODELS SHOWING THEIR ADVANTAGES AND DISADVANTAGES

Model	Definition	Advantages	Disadvantages
RF [17]	It is algorithm used for classification and regression, where the data gets split into subsets and we train several DTs.	Automatic processing of missing values and no need to transform variables, good performance with many variables and large data, and high accuracy	Slow learning, large memory footprint, and difficult interpretation.
SVM [17]	A binary classification method that creates a hyperplane to divide two target values.	High accuracy, works well with high-dimensional data, and memory efficient	Poor performance with noisy data, and long training time.
Naïve Bayes (NB) [17]	Collection of classification algorithms based on Bayes theorem.	Easy to understand and configure, Fast and small memory footprint, and can learn from small data.	Failure to predict rare events, and possible overfitting.
Logistic Regression (LR) [17]	Statistical method for analyzing data with dichotomous outcomes.	Good outcomes with a small number of variables and easy to implement	Requires many samples for training, and not easy result interpretation.
DTs [17]	It used algorithm for classification and regression by dividing the data into areas of similar characteristics.	Easy to maintain and understand, good results with small data, and intuitive	Large memory footprint, tendency for overfitting, and, high variation in generated models.

TABLE II. LIST OF UNSUPERVISED ML MODELS SHOWING THEIR ADVANTAGES AND DISADVANTAGES

Model	Definition	Advantages	Disadvantages
K-Means Clustering [17]	Builds a hierarchy of clusters by connecting adjacent clusters.	Not sensitive to distance selection, accepts noisy data, does not require predetermination of the number of clusters.	Complex algorithm (O(n power 3)), cannot process large data volumes.
Hierarchical Clustering [17]	A binary classification method that creates a hyperplane to divide two target values.	High accuracy, works well with high-dimensional data, and memory efficient	Long training time, and poor performance with noisy data.

H. Ensemble Learning

It is a technique that uses multiple models to obtain better performance than any one model alone. Ensemble learning is an idea where predictions from several models of different knowledge and strategies are averaged, letting the resulting system handle overfitting and perform better [24]. Anomaly detection systems in cloud environments are improved by ensemble methods like bagging, boosting and stacking. Nevertheless, ensemble models are computationally expensive, and

their complexity makes them difficult to deploy in real time cloud environments[29].

III. ML's Role in Real-time Anomaly Detection in Cloud Computing

Fig. 4 shows several advantages of using traditional ML techniques for real time anomaly detection in cloud computing environments:

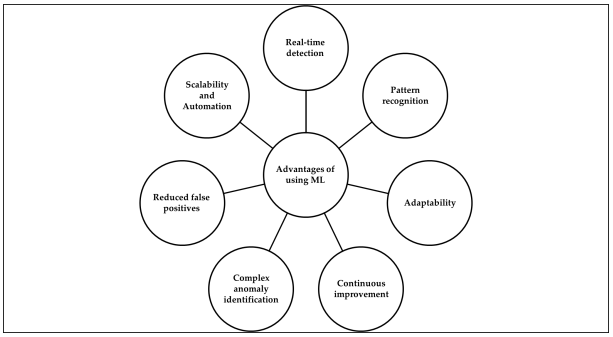

Fig. 4. Advantages of using traditional ML.

1) Real-time detection: ML is real time and immediately alerts when there are suspicious activities which is critical for responding to risks quickly [30].
2) Pattern recognition: On historical data, ML formulates complex patterns and trends and highlights normal activities against anomalies.
3) Adaptability: The ML models learn, update and change, almost continuously, to accommodate the evolving data patterns in a dynamic cloud environment.
4) Continuous improvement: Anomaly detection capabilities are made better and better by ML models through retraining with new data. This iterative improvement keeps the models effective at detecting new threats and adapting to ever changing cloud environment patterns.
5) Complex anomaly identification: Compared to other anomaly detection approaches, ML is unique because it can deal with multivariate anomalies from sources like unusual access patterns of users, different behavior in network traffic.
6) Reduced false positives: ML algorithms are trained to understand the unique behaviour of the organization's cloud environment so normal activities are less likely to trigger anomalies. With this precise tuning, security teams don't have to be overwhelmed with false alerts, but can concentrate on real threats.
7) Scalability and Automation: In cloud environment, ML consumes large volume of generated data and conducts an automated anomaly detection, reducing the manual intervention.

IV. Objective

The objectives of this research are as follows:

- To conduct a comprehensive literature review of existing research on anomaly-based ML detection in cloud computing.
- To investigate the ML techniques used, datasets used, and their accuracy.
- To examine the cloud computing models used.
- To develop a taxonomy for the systematic literature review to categorize and analyze the research findings.
- To identify the challenges and advantages of anomaly-based ML detection in cloud computing.
- To outline potential future directions for research in anomaly-based ML detection in cloud computing.

By addressing these objectives, this review aims to offer a clear understanding of the current landscape of anomaly detection using ML in cloud computing, highlight the barriers that need to be overcome, and propose directions for future innovations that can help secure cloud environments more effectively. The insights gained from this review can be valuable for both researchers and practitioners in the field of cloud security, guiding future research efforts and helping organizations implement effective anomaly detection solutions.

V. Research Methodology

We follow a systematic approach to review the existing literature on anomaly based ML detection in cloud computing, following the Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) guidelines. It includes defining the research questions, selecting databases, developing search strings, establishing of inclusion exclusion criteria, and applying a quality assessment framework. The methodology is organized as follows:

A. Research Questions (RQ)

The following research questions (RQs) guide this study to provide a structured analysis of anomaly detection models in cloud computing environments:

- RQ1: What anomaly-based ML techniques are applied in cloud computing environments, and how are these models classified?
- RQ2: What datasets and evaluation metrics are used in the assessment of these models?
- RQ3: What are the primary challenges and benefits of using anomaly-based ML detection in cloud environments?
- RQ4: What gaps exist in the current literature, and what future research directions are suggested for advancing anomaly-based detection in cloud computing?

B. Data Sources and Search Strategy

To ensure comprehensive coverage of relevant studies, the search was conducted across the following academic databases:

- IEEE Xplore
- MDPI
- SpringerLink
- ScienceDirect

- ACM Digital Library

The keywords used for the selection based on the related research objectives:

("Anomaly Detection" OR "Anomaly") AND ("Machine Learning" OR "ML") AND ("Cloud Computing" OR "CC" OR "Cloud")

Only peer reviewed journal articles, conference papers published between 2020 and 2024 were considered to capture recent developments.

C. Inclusion and Exclusion Criteria

To filter search results for relevant studies, we established the following inclusion and exclusion criteria:

1) Inclusion Criteria:

- Studies that focus on anomaly detection using ML within cloud computing environments.
- Peer-reviewed journal articles, conference papers.
- Studies that provide empirical results or evaluations using datasets relevant to cloud settings.
- Publications written in English.

2) Exclusion Criteria:

- Studies not related to anomaly detection in ML applications for cloud computing.
- Publications that only provide theoretical models without empirical validation.
- Non-peer-reviewed sources such as theses, white papers, and editorials.

D. Study Selection Process

The study selection process adhered to the PRISMA framework, proceeding in three stages:

- Initial Screening: All retrieved articles were screened by titles and abstracts to exclude irrelevant studies and choose those meeting the inclusion criteria for full text review.
- Full-Text Review: Full texts of selected articles were reviewed to determine their relevance and quality. Excluded articles that did not provide detailed information on ML techniques, datasets or empirical evaluations.
- Data Extraction and Coding: A standardized form was used to extract data from the final set of articles, including anomaly detection techniques, datasets, evaluation metrics, as well as identified challenges and benefits.

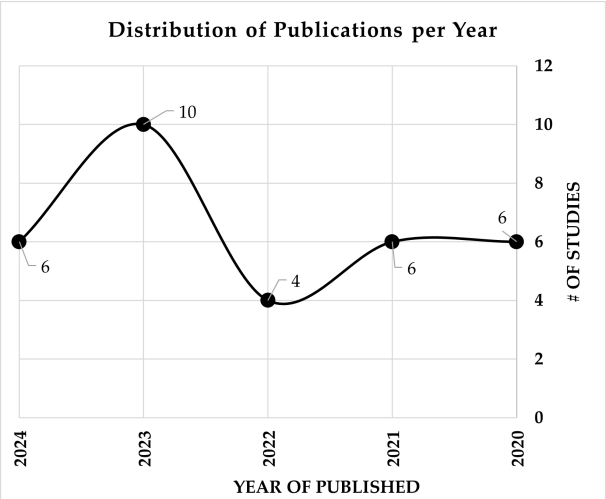

Fig. 5. Distribution of the selected studies per publication year.

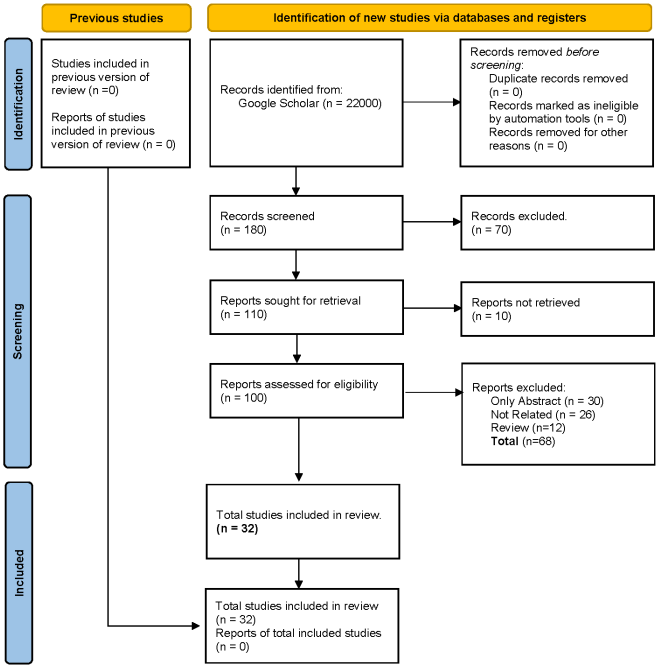

Fig. 6. PRISMA flow chart for the selection process.

E. Selection Results

With the application of selection criteria, 70 papers were excluded and 110 papers were selected for further review. Of these, 10 papers were not retrieved, and 100 articles were assessed for eligibility. At this stage, 68 articles were excluded leaving 32 articles in the final SLR. Fig. 5 shows distribution of the selected studies per publication year.

Fig. 6 shows the PRISMA flow diagram showing each stage of the process is presented in PRISAM.

VI. Literature Review

A. Supervised Models

Talpur et al. [31] presents a robust framework for DDoS attacks using evolutionary algorithms with ML models. They propose an innovative hybrid methodology that combines Extreme Gradient Boosting(XGBoost)-Genetic Algorithm(GA) Optimization, RF-GA Optimization, and SVM-GA Optimization with the Tree-based Pipelines Optimization Tool (TPOT). It automates the optimization of ML pipelines to enhance accuracy. Datasets such as KDD Cup 99 and CIC-IDS 2017 were used for the study, which achieved high accuracy scores of 99.99% for XGBoost-GA and SVM-GA, and 99.50% for RF-GA using 10-fold cross validation. Although effective, the methodology is limited by an increase in computational complexity resulting from multiple detection models and optimization phases.

Alduailij et al. [32] introduces an effective approach to DDoS attack detection in cloud environments. The main contribution is the use of Mutual Information (MI) and Random Forest Feature Importance (RFFI) for feature selection to reduce misclassification errors. The authors evaluate five ML models RF, Gradient Boosting (GB), Weighted Voting Ensemble (WVE), K-Nearest Neighbor (KNN), and LR. The method demonstrated to achieve high accuracy rates using the CICIDS 2017 and CICDDoS 2019 datasets, with RF reaching 99.997% accuracy and lowest misclassification errors when trained using 19 features. The study notes that KNN has a higher computational cost and models such as LR and GB need better parameter tuning. Although these limitations exist, the research proves that MI and RFFI feature selection can increase DDoS attack detection accuracy over different ML models.

DASARI and KALURI [33] suggests a hierarchical ML approach optimized for hyperparameter tuning to boost intrusion detection in networks due to DDoS attacks. The research uses the CICIDS 2017 dataset and preprocesses the data with normalization and balancing techniques such as Min-Max scaling and SMOTE. Feature selection is executed via the LASSO method, and the selected features are fed into five ML classifiers XGBoost, Light Gradient-Boosting Machine (LightGBM), CatBoost, RF, and DT. Model accuracy was improved through hyperparameter optimization. Of all these, LightGBM had the highest classification accuracy of 99.77%, better than all other models. It also mentions future areas of improvements in handling real time data and adaptive attacks. The work contributes to improving Intrusion Detection System (IDS) capabilities with hierarchical ML techniques for high precision and recall.

Mishra et al. [34] introduce a perplexed Bayes classifier model for identifying and mitigating DDoS attacks in cloud computing environments. This classifier uses the NSL-KDD dataset which contains DDoS attack scenarios and features. The innovation is in using correlation based feature selection to improve classification accuracy to a 99%. This method is benchmarked against the established algorithms of NB and RF and found to be more accurate, sensitive, and specific. Furthermore, they compare perplexed Bayes classifier against nature inspired feature selection techniques such as GA and Particle Swarm Optimization (PSO) and show that perplexed Bayes classifier achieves 2% to 8% higher accuracy.

Parameswarappa et al. [35] propose a new intrusion detection system for cloud computing based on ML and deep learning techniques to boost security. UNSW-NB15 dataset was used by the authors for developing and testing their model which consists of multiple classifiers, LR, KNN, DT, RF, Extra Trees, GB, and Multilayer Perceptron (MLP). It focuses on preprocessing by using K best feature selection for optimizing classification tasks. Models detected cloud anomalies and attacks with a detection rate of 97.68% by RF. The model improves precision and reduce false positives, but is limited by its dependence on labeled datasets and its use in broader cloud environments. This suggests further work of integrating advanced data mining, deep learning techniques into the existing anomaly detection process to increase its accuracy on various anomalies.

Advanced ML techniques are developed by M et al. [36] to enhance data security in cloud computing environment. They evaluate three ML models such as RF, Deep Neural Network (DNN), and Q-Learning across different experiments. RF model was reliable in categorizing security threats, with 95% accuracy and balanced precision of 0.92, recall of 0.96, and F1 score of 0.94. Also, DNN model demonstrated good performance with an accuracy of 97%, a recall of 0.98 and an F1 score of 0.96 and it could recognize complex patterns in cloud data. The research used cloud system data sets of logs, network traffic, and access patterns for anomaly detection and adaptive security response. The resource intensive nature of the deep learning models, difficulties to reduce false positives in Q-Learning and ethical issues, such as privacy preservation are also limitations. In particular, this work calls for the continuous optimization of ML models for cloud security to cope with the ever changing threat landscape.

ABUBAKAR et al. [37] propose a hybrid DDoS detection and mitigation mechanism using an optimized SVM is combined with SNORT Intrusion Prevention System (IPS). This integrated approach identifies malicious traffic early and mitigates attacks by rerouting or dropping suspicious packets. The methodology uses the KDDCup99 and DARPA datasets, while the abnormalities in real time network traffic are analyzed. The results show that the system achieves superior average detection accuracy of 97.9% compared to traditional SNORT IPS, Probabilistic Neural Networks (PNN) and Back Propagation methods. While the method has high accuracy and low false positives, it suffers from two limitations: multithreading and zero-day attack detection. The model focuses on supervised learning configurations for traffic behavior analysis and protocol validation, which is effective but limited by dataset quality and scope.

A sophisticated cloud IDS is introduced by BAKRO et al. [38] which utilizes a hybrid feature selection method combined with a RF classifier. The proposed methodology integrates Information Gain (IG), Chi-Square and PSO for selecting relevant features which increases the model accuracy and reduces the data dimensionality. The Synthetic Minority Over-sampling Technique (SMOTE) is used to address data imbalance, and robust performance in multi-class attack detection is ensured. The proposed system shows its effectiveness while detecting different attack types on the UNSW NB15 and Kyoto datasets with 98% and 99% accuracy rates, respectively. However, it

suffer from reliance on resource-intensive feature selection methods. The system is shown to significantly improve detection rates but may not necessarily generalize to real world cloud environments without further dataset diversity.

BAKRO et al. [39] presents a novel cloud IDS. The hybrid feature selection technique Grasshopper Optimization Algorithm (GOA) and GA aims to optimize feature selection and enhance IDS performance by improving classification accuracy and reducing computational complexity. This hybrid approach helps optimize feature selection while increasing accuracy of the classfied data and reducing amounts of computation. The model uses an RF classifier trained on the selected features with an ADAptive SYNthetic (ADASYN) algorithm for minority oversampling and RUS for majority class balancing. The proposed system is evaluated on three datasets including UNSW-NB15, CIC-DDoS2019, and CIC Bell DNS EXF 2021 and achieves accuracy of 98%, 99% and 92% respectively. However, because it is dependent on the specific datasets to evaluate on, and may not be generalizable. It achieved improvements in True Positive Rate (TPR) and False Positive Rate (FPR) along with better performance than state of the art classifiers like SVM, AlexNet and XGBoost.

A novel framework for cloud anomaly detection using a Secure Packet Classifier (SPC) is proposed by Chkirbene et al. [40] The SPC combines two ML algorithms selected based on accuracy and computational efficiency and leverages collaborative filtering. The main focus of the model is anomaly detection and classifying different types of attack which is crucial for targeted counter measures. Using the UNSW-NB15 dataset, the model delivered an impressive improvement on accuracy, detecting 81% of anomalies with a lower FPRthan the traditional methods. The work is constrained by the fact that the model relies on specific dataset properties and does not generalize to other datasets without significant retraining.

Aldallal and Alisa [41] propose to develop a hybrid IDS for cloud computing environments, which integrates GA for feature selection and SVM for classification. A novel fitness function is used by the system to measure the performance of the intrusion detection system, combining F1-score, accuracy and TPR to ensure balance and accuracy of the detection system. They used CICIDS2017 dataset for evaluation, results show that the proposed model provided up to 5.74% improvement on detection accuracy over benchmarks, while demonstrating its effectiveness. Although the system performed well, it required data preprocessing, including cleaning missing or corrupted entries, and relied on predefined datasets for evaluation rather than real-time data.

Jaber and Rehman [42] propose an IDS for cloud computing environments by combining Fuzzy C-Means (FCM) clustering with SVM. The hybrid FCM–SVM model proposed can overcome the limitations of conventional IDS including high false alarm rates and poor accuracy. The proposed system is implemented using the NSL-KDD dataset, with clustering used to group data points for improving SVM performance in anomaly detection. The system shows a capability to classify different types of network attacks with an accuracy of 97.37% for User to Root (U2R) attacks, 98.46% for Remote to Local (R2L) attacks and 98.85% for Probe attacks.

AlSaleh et al. [43] proposes a novel ML approach for detecting DDoS attacks in cloud computing settings using a Bayesian Convolutional Neural Network (BaysCNN). To achieve significant improvements in DDoS detection accuracy, BaysCNN uses a 19 layer architecture with an average accuracy of 99.66% across 13 multi class attacks. The study also improves model performance using the Bayesian-based Convolutional Neural Network with Data Fusion (BaysFusCNN) approach, which combines features from different sources, yielding a better accuracy of 99.79%. This research demonstrates that these models can tackle challenges including distinguishing application-layer attacks and real time detection. Bayesian methods are also used in the models to estimate uncertainties to improve reliability. The limitation is the dependence on the CICDDoS2019 dataset for which the results are not generalizable to other datasets and environments.

Sherubha et al. [44] propose a novel anomaly detection mechanism through an auto-encoder for feature selection and a NB classifier for classification. The approach improves the ability of existing IDS to deal with unlabeled data and reduces redundancy and noise in datasets. On NSL-KDD dataset, the model shows a detection accuracy of 93% which is better than traditional methods like J48 and RF. The main contribution of this study is the combination of unsupervised learning for dimensionality reduction and NB for robust classification, which achieves superior performance in detecting network anomalies. However, the approach is tested only on a static dataset, which limits its real time applicability and ability to address zero day attacks. The results of this research highlight the possibility of application of hybrid methodologies in order to improve intrusion detection in cloud computing environments.

Moreira et al. [45] propose ISAD; an intelligent system for anomaly detection in smart environments based on the integration between Fog and cloud computing. The system uses ML techniques to process network traffic and detect unusual behavior, offloading the data processing overhead to the Fog and cloud environments. A fog layer is used to perform raw network traffic data processing, feature extraction, and transmit filtered data to the cloud for dynamic anomaly detection using ML models. The system achieves high accuracy especially with RF, achieving 98.7% accuracy with Microsoft Azure. The CICIDS dataset is used, which represents real network traffic scenarios. The system shows robust performance, but it is reliant on fog and cloud environment computational infrastructure and has reduced recall in some ML configurations, which poses challenges in generalizing anomalies.

The authors, Alshammari and Aldribi, presents a lightweight ML based framework to boost IDS for detecting network anomalies [46]. In order to assess its performance, the framework utilizes the ISOT-CID dataset by incorporating novel features, more specifically the 'rambling feature', in classification. Also, In this study, six ML models such as DT, RF, and KNN are evaluated by using cross-validation and split validation techniques. Therefore, results show that DT and RF models are the most accurate, with 100% accuracy in both validation strategies. The novel feature addition and data preprocessing make the dataset better in quality, making the training of the models effective. While the model worked well, it requires large datasets and is not ready for real time deployment because of latency issues. This work points to future work, where deep learning approaches will be integrated to

overcome these limitations and better refine anomaly detection in real-time cloud environments.

Al-jumaili and Bazzi investigate the use of ML models to improve IDS in cloud environments, that suffer from dynamic threats and false positives [47]. The research compares and analyzes algorithms like DT, RF, XGBoost and SVM and finds XGBoost as the best effective model with 99.63% accuracy as it has a great GB capability. They use NSL KDD dataset, which is well known for its rich network intrusion patterns and perform robust preprocessing such as label encoding and data scaling to improve the performance of the model. However, since the research is based on synthetic datasets, it is not applicable in real world scenarios and more realistic cloud traffic should be validated. The results help explain how ML can be used to develop robust and scalable IDS solutions for the ever changing cloud landscape.

Naiem et al. [48] proposes a new framework to optimize the Gaussian Naïve Bayes (GNB) classifier for DDoS detection in cloud environments. The research acknowledges the GNB's limitations, dependency on feature independence, and sensitivity to the zero frequency problem by addressing them with an iterative feature selection process and advanced preprocessing. Also, feature selection techniques such as the Pearson Correlation Coefficient (PCC), MI and Chi-squared tests are used, and the SMOTE algorithm is used to address data imbalance. They demonstrate a 2% improvement in accuracy and substantial gains in precision, recall and F1-score. In addtion , the approach improves GNB's performance on the CICD2018 dataset to the level of other classifiers such as RF and SVM with the simplicity and computational efficiency.

Aslan et al. [49] propose a new cloud based malware detection system focusing on intelligent behavior analysis. The proposed Cloud-Based Behavior-Centric Model (CBCM) collects execution traces of suspicious files in VMs, identifies relevant behaviors, and extracts discriminative features. Both learning based and rule based detection agents process these features. They used ML classifiers such as RF and logistic model trees with 99.83% accuracy and a 0.6% FPRon a dataset of 10,000 samples for RF. Also, real time detection is further augmented by rule based agent. In addtion, the work uses cloud scalability to efficiently analyze malware, showing the high accuracy and speed of detection compared to traditional methods. Limitations, however, exist in the form of difficulty in detecting advanced obfuscated malware, and the requirement for broader dataset diversity. This research greatly enhances the malware detection efficiency in cloud computing environments.

Mehmood et al. [50] offer an advanced framework of privilege escalation attacks detection and mitigation within cloud computing environments, which is based on ensemble learning. The research works on a customized dataset from multiple CERT dataset files, using ML algorithms RF, Adaptive Boosting (AdaBoost), XGBoost and LightGBM to classify and mitigate insider threats. Moreover, the highest accuracy (97%) was achieved by LightGBM, which was better than RF (86%), AdaBoost (88%) and, XGBoost (88.27%). This was achieved by pre processing the dataset, training the models, and tuning the hyperparameters to solve the specific attack scenarios thereby leading to a robust detection mechanism. Insider threat research concludes that insider threats, in particular those resulting from privilege abuse, are especially critical and suggests ensemble learning for increased classification accuracy. Although the study yields promising accuracy, it is limited to a single dataset and is challenged in recognizing subtle attack patterns, which suggests future exploration of the diversity of datasets.

Bamasag et al. [51] introduce the Real-Time DDoS flood Attack Monitoring and Detection (RT– AMD) model which is aimed at mitigating the effects of DDoS attacks on cloud computing environments. The RT-AMD model works to utilize ML algorithms such as RF, KNN, NB, and DT, with high accuracy, to detect abnormal network traffic. The model is evaluated on the DDoS-2020 dataset, which contains balanced attack and normal traffic records for Transmission Control Protocol (TCP), Domain Name System (DNS), and Internet Control Message Protocol (ICMP) protocols, and achieves an accuracy of 99.38% in real time detection. The incremental learning capability further improves real time adaptation and detection without a costly retraining process. The study reveals the scope of expansion in its impressive accuracy and performance, including the incorporation of various DDoS types and evaluating on other cloud environments. This is key to advancing secure, real time cloud operations.

Chkirbene et al. [52] introduce a novel ML based weighted class classification scheme to tackle the challenges of anomaly detection in cloud computing, in the context of class imbalance problem. The system improves the classification accuracy of rare attack classes by integrating supervised learning with past node behavior and a weight optimization algorithm. The approach involves training a DT classifier on the UNSW-NB15 data set and using historical data to determine decision weights and obtains 95% accuracy. This framework significantly enhances multi class detection capabilities and is resilient to underrepresentation of minority attack classes. However, the reliance on historical data and the computational cost of weight optimization restrict the model's adaptability in dynamic real time environments.

In their work, Sambangi and Gondi investigate the use of MLR to detect DDoS attacks in cloud environments [53]. The study uses the CICIDS2017 dataset, focusing on the Friday afternoon traffic logs and applies an IG based feature selection technique to select critical attributes, reducing the dimensionality from 79 to 16 and then to 6. It is shown that the MLR model achieves 73.79% accuracy with 16 attributes. In addition, residual plots and fit charts are used by the authors to visualize the model's ability to differentiate between benign and attack traffic. The research is restricted to single day log data and does not explore ensemble or deep learning methods. This work contributes to enhancing the DDoS detection efficiency by simplifying the feature selection process and focusing on the regression analysis to handle the attack classification issues.

In response to DDoS attacks in cloud environments, Wani et al. [54] propose a robust IDS. They created a unique dataset with 21 attributes by using the CloudStack platform for experimentation and using Tor Hammer for generating malicious traffic. Also, the researchers evaluated six ML models: DT, NB, RF, C4.5, and SVM, K-Means. The SVM algorithm showed the best accuracy of 99.7% among the algorithms such as C4.5 (98.7%) and RF (97.6%). Furthermore, to evaluate the performance of the system, the dataset was analyzed by using Weka tool and evaluating the performance using precision,

recall metric. However, the study shows that data imbalance is a challenge, but does not consider more general attack scenarios or feature selection optimization. They contribute by showing that SVM is effective in detecting anomalies but their work is limited to specific tools and attack types.

P et al. [55] propose a ML framework for detection of phishing attacks in distributed cloud systems. In particular, the authors use supervised learning algorithms such as NB, SVM, and DT to detect phishing attacks. The study evaluates and compares the performances of these algorithms in terms of accuracy by using IDS generated dataset and shows that DT is the best method with slightly slower than other algorithms. The work aims at solving the critical challenge of resource management in cloud systems, which is often exploited for phishing attacks. Although the proposed model has a good detection accuracy, its response time is slower than the current standards and it depends on pre processed datasets which does not allow it to adapt to real time. The results highlight the importance of feature reduction and classification methodology for improving detection rates and reducing false alarms in cloud computing environments.

Kushwah and Ranga propose a novel system based on a Voting based Extreme Learning Machine (V-ELM) to detect DDoS attacks in cloud computing environments [56]. Unlike conventional single layer neural network approaches, the system uses multiple Extreme Learning Machines (ELM) with a majority vote mechanism to improve detection accuracy and reduce false alarms. The proposed model is evaluated using NSL-KDD and ISCX datasets and the accuracies of 99.18% and 92.11% respectively are shown. Accuracy and false positive rates for the system are shown to be superior to traditional models such as RF and Adaboost. But, this requires great amounts of labeled training data which can constrain it. Finally, the research mainly focuses on the challenges of high detection accuracy, false positive rates, and fast training and offers an efficient scalable DDoS attack detection solution.

Guezzaz et al. [57] propose a cloud based intrusion detection model by using the RF algorithm and feature engineering for improved anomaly detection. The research discusses the ever growing security challenges in cloud environments including unauthorized intrusions and real time attack detection. For the proposed framework, the feature set is reduced to only two important attributes of the NSL-KDD and Bot-IoT datasets while using data visualization to ease the feature engineering process. It achieved accuracy rates of 98.3% and 99.99% on these datasets. Although the model was able to achieve high precision and accuracy, it has low recall, meaning that it cannot detect all attacks of some types. It shows that RF can be a better classifier than SVM and DNN with a minimal feature set, and thus the work demonstrates the potential for using a minimal feature set for effective classification. Yet, there is room for improvement in recall and evaluation has yet to be done across various datasets. Table III presents a summary of supervised ML models, datasets and their accuracy rates as demonstrated by the related works.

The attacks, advantages, and disadvantages of supervised models for anomaly detection in cloud computing are highlighted in Table IV.

TABLE III. SUMMARY OF SUPERVISED MODELS AND THEIR ACCURACY RATES BASED ON RELATED WORK

Author	Year	ML Models Used	Dataset	Accuracy Rate
Talpur et al. [31]	2024	XGBoost-GA, SVM-GA	KDD Cup 99, CIC-IDS 2017	99.99%
Alduailij et al. [32]	2022	RF	CICIDS 2017, CI-CDDoS 2019	99.997%
DASARI and KALURI [33]	2024	LightGBM	CICIDS 2017	99.77%
Mishra et al. [34]	2022	Perplexed Bayes Classifier	NSL-KDD+	99%
Parameswarappa et al. [35]	2023	RF	UNSW-NB15	97.68%
M et al. [36]	2024	DNN	Cloud-based	97%
ABUBAKAR et al. [37]	2020	SVM	KDDCup99 and DARPA	97.9%
BAKRO et al. [38]	2023	RF	Kyoto	99%
BAKRO et al. [39]	2024	GOA-GA with RF	CIC-DDoS2019	99%
Chkirbene et al. [40]	2021	SPC	UNSW-NB15	81%
Aldallal and Alisa [41]	2021	GA with SVM	CICIDS2017	99.65%
Jaber and Rehman [42]	2020	FCM with SVM	NSL-KDD	98.85%
AlSaleh et al. [43]	2024	BaysFusCNN	CICDDoS2019	99.79%
Sherubha et al. [44]	2023	NB	NSL-KDD	93%
Moreira et al. [45]	2021	RF	CICIDS	98.7%
Alshammari and Aldribi [46]	2021	RF	ISOT-CID	100%
Al-jumaili and Bazzi [47]	2023	XGBoost	NSL-KDD	99.63%
Naiem et al. [48]	2023	GNB	CICD2018	97.57% with PCC-IM
Aslan et al. [49]	2021	RF	Various sources and forming (7000 malware, 3000 benign)	99.83% with cross-validation
Mehmood et al. [50]	2023	LightGBM	customized dataset derived from multiple CERT dataset files	97%
Bamasag et al. [51]	2022	RF	DDoS-2020	99.38%
Chkirbene et al. [52]	2020	DT	UNSW-NB15	95%
Sambangi and Gondi [53]	2020	MLR	CICIDS2017(Friday afternoon traffic logs)	73.79% using 16 features
Wani et al. [54]	2020	SVM	Custom dataset with 21 attributes	99.7%
P et al. [55]	2023	DT	IDS-generated dataset	87%
Kushwah and Ranga [56]	2020	V-ELM	NSL-KDD	99.18%
Guezzaz et al. [57]	2023	RF	Bot-IoT	99.99%

B. Unsupervised Models

Shanthi and Maruthi [58] introduced a new method to build anomaly-based IDS in cloud computing environment with the combination of Isolation Forest and SVM models. The proposed system aims to improve the efficiency and accuracy of detecting anomalous activities in large and complex network datasets. The study uses the NSL-KDD dataset to evaluate the performance of both models with Isolation Forest attaining an accuracy of 99% and SVM of 95%. Isolation Forest isolates anomalies via recursive random splits, and a supervised binary classifier SVM learns to identify anomalies by learning the normal vs anomaly distinction. Although Isolation Forest

TABLE IV. SUMMARY OF THE ATTACKS FOCUS, ADVANTAGES, AND DISADVANTAGES OF SUPERVISED MODELS IN ANOMALY DETECTION IN CLOUD COMPUTING

Ref	Attacks	Advantages	Disadvantages
[31]	DDoS	Superior pipeline optimization with TPOT.	Increased computational complexity.
[32]	DDoS	Integration of MI and RFFI for feature selection.	LR and GB require tuning to reduce errors.
[33]	DDoS	Hyperparameter tuning significantly improved overall performance metrics.	Limited to CICIDS dataset.
[34]	DDoS	Efficient feature selection method.	Dependent on pre-processed and structured datasets.
[35]	DDoS	Integration of historical and real-time decisions.	Dependency on labeled datasets.
[36]	Data breaches, unauthorized access	Ability to detect complex patterns.	Resource-intensive computation, privacy concerns.
[37]	DDoS	Early detection and mitigation of attacks.	Limited handling of zero-day attacks.
[38]	DoS, worms, and exploits.	High accuracy, robust detection rates.	Resource-intensive feature selection.
[39]	DoS, DDoS, and DNS attacks	Enhanced classification accuracy.	Limited evaluation datasets.
[40]	Analysis, Backdoor, DoS, Exploits, and others	Effective classification of attack types.	Requires significant retraining for new environments.
[41]	Brute-force attacks, SQL Injection, and others	Effective feature selection	Requires significant preprocessing for corrupted data.
[42]	U2R, R2L, Probe and DoS	High accuracy, low false alarm rates.	Dependency on pre-defined datasets.
[43]	Multi-class DDoS attacks	High accuracy.	Potential computational overhead for real-time scenarios.
[44]	DoS, probe, R2L and U2R	Effective feature selection	Ineffective against zero-day attacks
[45]	Network anomalies, including DDoS	Reduces data processing overhead by fog preprocessing.	Recall performance varies across models.
[46]	Malicious network traffic	Novel features like "rambling" improved detection performance.	Limited real-time deployment capability.
[47]	Network intrusions, anomalies, Unknown and zero-day attacks	Balanced precision and recall.	High computational cost for XGBoost.
[48]	DDoS	Iterative feature selection improves accuracy and reduces overfitting.	GNB still underperforms compared to more advanced classifiers.
[49]	Malwares	Combines ML and rule-based approaches for robustness.	Imbalanced dataset distribution of malware types.
[50]	Privilege escalation attacks (horizontal and vertical) and insider threats.	Efficient in identifying privilege escalation.	Limited dataset diversity restricts generalizability.
[51]	DDoS flood attacks targeting the network/transport layer (TCP, DNS, ICMP)	Real-time detection, high accuracy, uses incremental learning, and adaptive model.	Limited to specific attack types.
[52]	Analysis, Backdoor, DoS, Exploits, Fuzzers, Reconnaissance, Shellcode, Worms.	Enhances detection accuracy for rare attack classes.	Relies heavily on historical data; computationally intensive weight optimization.
[53]	DDoS	Demonstrates effectiveness of MLR	Limited to single-day traffic logs.
[54]	DDoS	Custom dataset tailored to cloud-specific scenarios.	Dataset imbalance issues.
[55]	Phishing attacks	Highlighted the importance of feature reduction to enhance detection speed and accuracy.	Slower response time.
[56]	DDoS	High detection accuracy.	Requires large labeled datasets.
[57]	General anomaly detection.	Low computational overhead due to reduced feature set.	Poor recall, and limiting generalizability.

has high accuracy and adaptability, it is also dependent on strong parameter tuning like contamination factor, and SVM is dependent on data portioning. The work suggests that careful feature selection and dimensionality reduction are necessary to enhance detection abilities.

Ntambo and Adeshina present a proactive anomaly detection model for detecting anomalies in VM resource usage in cloud environments with emphasis on the multitenancy vulnerabilities in public clouds [59]. The model uses Isolation Forest and One-Class Support Vector Machine (OCSVM) algorithms to detect anomalies in deviations of VM metrics such as CPU, memory usage and disk throughput. The dataset used is from the Grid Workload Archive, consisting of time series VM resource data. The results show that OCSVM gains the best accuracy with F1 = 0.97 for hourly and F1 = 0.89 for daily series compared to Isolation Forest. However, these results rely on specific datasets and only consider a limited set of real world scenarios. In this work, they contribute a scalable approach to augmenting cloud security with real time anomaly detection capabilities for VM resources.

Table V presents a summary of unsupervised ML models, datasets and their accuracy rates as demonstrated by the related works.

The attacks, advantages, and disadvantages of unsupervised models for anomaly detection in cloud computing are highlighted in Table VI.

TABLE V. SUMMARY OF UNSUPERVISED MODELS AND THEIR ACCURACY RATES BASED ON RELATED WORK

Author	**Year**	**ML Models Used**	**Dataset**	**Accuracy Rate**
Shanthi and Maruthi [58]	2023	Isolation Forest	NSL-KDD	99%
Ntambo and Adeshina [59]	2021	OCSVM	Grid Workload Archive	not specified.

C. Hybrid Models

Wang et al. [60] proposes a hybrid anomaly detection system for cloud computing environments using a Stacked Contractive Autoencoder (SCAE) and a SVM. A deep learning based SCAE for unsupervised feature extraction is proposed which transforms raw network traffic data into low dimensional robust representations. An SVM is used to classify these features for malicious activities. The methodology is evaluated on two benchmark datasets, KDD Cup 99 and NSL-KDD. Experimental results show that the model achieves good detection rates and has an accuracy of 87.33% in multi class classification tasks. The research also points out the limitations

TABLE VI. SUMMARY OF THE ATTACKS FOCUS, ADVANTAGES, AND DISADVANTAGES OF UNSUPERVISED MODELS IN ANOMALY DETECTION IN CLOUD COMPUTING

Ref	Attacks	Advantages	Disadvantages
[58]	Network-based such as intrusions	High accuracy, robust feature extraction, efficient handling of large datasets.	Dependence on hyperparameter tuning, limited response to encrypted traffic.
[59]	Anomalies in VM resource usage, including stealthy attacks exploiting multitenancy in public clouds	Combines VM metrics for improved anomaly detection.	Limited generalizability due to specific datasets.

of the classifier in identifying less common attack types and identifies directions for further optimizing the classifier. The results of this work suggest that hybrid models are a viable solution to the scalability and precision problems inherent in cloud IDS.

A hybrid clustering and classification-based approach to intrusion detection in distributed cloud computing environments is proposed by Samunnisa et al. [61]. They introduces a ML-based anomaly detection system that combines K-means clustering with RF classifiers to classify malicious activities across five types DoS, Probe, U2R, R2L, and normal. The model is tested using NSL-KDD and KDDCup99 datasets and shows high accuracy and low false alarm rates with 99.78% detection rate and 0.09% false alarm rate for NSL-KDD dataset. The methodology uses threshold-based functions and measures accuracy, detection rate and the area under the curve. The study points out that the datasets used are out dated and do not necessarily reflect the current network threats. It demonstrates that hybrid models can improve IDS but the use of the outdated datasets may limit their applicability to modern, evolving attack patterns..

Megouache et al. [62] present a new framework for intrusion detection that combines clustering and classification in cloud environments. It uses the K-means clustering to label previously unlabeled datasets and is able to use the ELM classifier to quickly identify and prevent malicious activities. Also, the proposed system is tested on the KDD99 dataset, where an accuracy of 99.2% in detecting non legitimate users is achieved. The main innovation is in the integration of clustering and classification to perform data segmentation and intrusion detection in an optimized manner, dealing with issues like scalability and false positives. In addition, the approach uses probabilistic methods to minimize data loss risks and improve real time cloud security. Still, the method is limited by the time needed to perform matrix operations to train large scale datasets. In summary, the work provides a high speed, accurate, and scalable solution for intrusion detection in the cloud, which is of significant importance to cloud security, but requires further work to improve scalability and processing efficiency.

Table VII presents a summary of hybrid models, datasets and their accuracy rates as demonstrated by the related works.

The attacks, advantages, and disadvantages of hybrid models for anomaly detection in cloud computing are highlighted in Table VIII.

D. Taxonomy of the Research

The research taxonomy Fig. 7 has been systematically organized in order to categorize ML methods used in the cloud computing and hence give a clear picture and structure to the field. The taxonomy categorized ML as supervised, unsupervised and hybrid techniques. Furthermore, this taxonomy benefits researchers and practitioners to select the specific dataset and to select the most appropriate technique for their use cases. Also, the purpose of this taxonomy ultimately is to provide a framework for driving innovation in and making better decisions about using ML to solve cloud computing issues.

TABLE VII. SUMMARY OF SUPERVISED MODELS AND THEIR ACCURACY RATES BASED ON RELATED WORK

Author	**Year**	**ML Models Used**	**Dataset**	**Accuracy Rate**
Wang et al. [60]	2022	SCAE and SVM	NSL-KDD	87.33%
Samunnisa et al. [61]	2023	K-means with RF	NSL-KDD	99.85%
Megouache et al. [62]	2024	K-means Clustering, ELM	KDD99	99.2% for detecting non-legitimate users

VII. CASE STUDY: PRIVILEGE ESCALATION ATTACK DETECTION IN CLOUD COMPUTING USING ML

A. Analyzing Real-World Case Studies

The study by Mehmood et al. [50] is a good example of providing ML solutions to a real world inspired scenario that addresses the issue of privilege escalation attacks in cloud computing. An example of privilege escalation attacks is cases where attackers abuse the faulty system vulnerability, misconfiguration or inadequate access control to receive increased access to resources or information. These attacks can be classified across horizontal privilege escalation or where an attacker gains access to another user's privileges, as well as vertical privilege escalation or where the attacker gains the higher level of access like administrative or root privileges. Based in that and without a doubt, privilege escalation attacks can grant an attacker unauthorized access to sensitive information or cause disruption of the critical system operations leading to through severe data breaches.

For this study, a customized CERT dataset was used to aggregate user behavior logs from various sources to emulate real world insider activities. Furthermore, malicious activities like unauthorized file copy to deleting files and abnormal system access patterns were also part of these activities.

The proposed methodology is to design a ML enabled insider threat detection system to classify and address privilege escalation attacks. The dataset was then preprocessed using carefully designed strategies to remove outliers, manage missing values and select appropriate features. The dataset was able

TABLE VIII. SUMMARY OF THE ATTACKS FOCUS, ADVANTAGES, AND DISADVANTAGES OF HYBRID MODELS IN ANOMALY DETECTION IN CLOUD COMPUTING

Ref	Attacks	Advantages	Disadvantages
[60]	R2L, U2R, Probe, and DoS.	Efficient feature extraction; high accuracy in detecting major attack types.	Limited detection of less-represented attack types.
[61]	R2L, U2R, Probe, and DoS.	High accuracy, low false alarm rates.	Relies on outdated datasets.
[62]	Anomalies and malicious user identification.	High accuracy in intrusion detection.	Computational inefficiency with large datasets due to matrix operations.

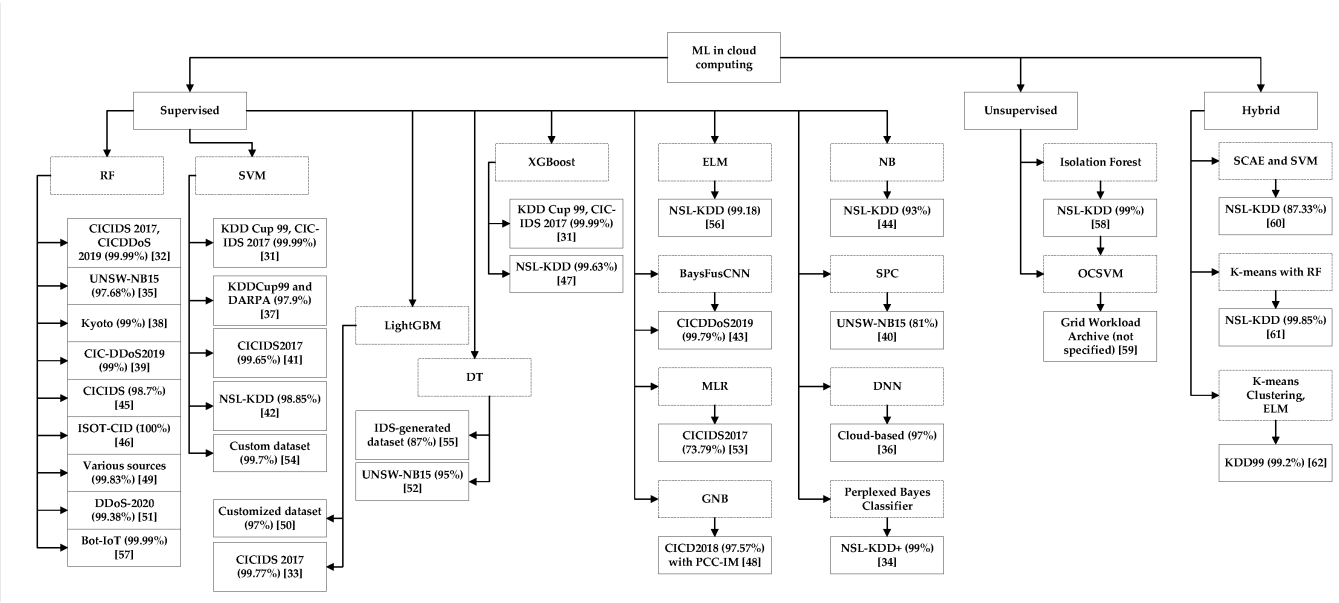

Fig. 7. Taxonomy of the Research

to simulate real world activities to provide an effective foundation in evaluating ML algorithms in real-world scenarios. In addition to ensuring the practical applicability of the models, this approach brought to light common challenges, including imbalanced data and effective feature engineering.

B. Insights from Actual Implementations

For detecting and classifying insider threats, the study employed four ML algorithms: AdaBoost, RF, XGBoost and LightGBM. The results of these algorithms were evaluated based on accuracy, precision, recall and F1 score for each algorithm. It is to note that, for the largest dataset, the LightGBM algorithm achieved the highest accuracy of 97%, due to its good leaf wise growth technique and also its performance on large dataset. Key insights from the implementation include:

- Algorithm Selection and Optimization: The results illustrated that using an ensemble approach is important to improve prediction accuracy by combining multiple models. The boosting techniques, especially LightGBM, were very good at catching complex patterns in high dimensional data, and are very effective for real world insider threat detection.
- Feature Engineering: Feature selection and preprocessing were emphasized as being critical. To improve model performance, features, that were irrelevant like ”employee” or ”file tree” were removed. The algorithms further gained stability on the learning of data by having data normalized and aggregated.
- Challenges in Data Quality: A large part of the first preprocessing phase involved handling missing values and outliers in the dataset. For example, while imputing the missing values in ‘File Copy’ feature, missing values were filled with the value which is found in the dataset patterns, hence not affecting the training process.
- Performance Metrics and Comparative Analysis: The performance metrics and confusion matrices helped them locate weaknesses and strengths of each algorithm. The high accuracy and low false alarm rate of LightGBM indicate that this solution can be used as a robust solution in real world deployments. Table IX shows the comparative analysis of performance metrics for the ML models in case study.
- Mitigation Strategies: The study not only discusses how detection of privilege escalation attacks is possible, but also suggests how such attacks can be suppressed to make them infeasible. As such, these include developing multifactor authentication and models of behavioral biometrics and secure access controls.

TABLE IX. LIST OF SUPERVISED ML MODELS SHOWING THEIR ADVANTAGES AND DISADVANTAGES

Algorithm	Accuracy	Precision	Recall	F1-Score	False Alarm Rate
RF	86%	86%	85%	85%	0.19
AdaBoost	88.27%	88%	86%	86%	0.16
XGBoost	89%	88.27%	87%	87%	0.13
LightGBM	**97%**	**97%**	**95%**	**95%**	**0.11**

C. Lessons Learned and Best Practices

The successful application of ML models in this case study highlights several best practices for implementing anomaly detection in cloud computing:

- Datasets are adapted to reflect real world scenarios to keep things practical and makes for better algorithm training.
- Using ensemble methods like LightGBM and XGBoost, combining the power of multiple methods, we have much higher accuracy and better reliability in threat detection.
- Preprocessing impact, They address the data quality issue of missing values and irrelevant features since these issues decrease model performance.
- The detection capabilities are complimented with preventive measures to include multifactor authentication and access control to provide additional security.
- Employees should be educated about cybersecurity best practice to reduce insider threats.

As indicated in this case study, ML algorithms have the potential to adequately deal with these complex cloud security challenges. These implementations provide valuable insights to guide anomaly detection projects in cloud environments in the future.

VIII. RESULTS AND DISCUSSION

The reviewed papers highlight several ML models that excel in anomaly detection for cloud computing environments. RF is shown to be the dominant supervised model as shown in Fig. 8, capable of reaching accuracy rates of 99.997% when complemented with powerful feature selection techniques such as MI and RFFI [32]. Other top performers include SVM, which may achieve accuracy of 99.99%, using techniques such as GAs or clustering [31], [41]. LightGBM and XGBoost, which are tree-based ensemble methods, have competitive accuracy 99.77% and higher when they are hyperparameter tuned [33], [47].

Anomaly detection associated with unsupervised and hybrid models is also critical. In unsupervised scenarios, Isolation Forest is very effective with 99% accuracy using recursive random splits [58]. Other hybrid models such as K-means clustering with a RF classifier deliver remarkable performance, and achieved an accuracy up to 99.85% [61]. Other innovative approaches, for example using autoencoders with a classifier such as SVM, were explored that can extract robust features and still achieve detection accuracy as high as 87.33% [60].

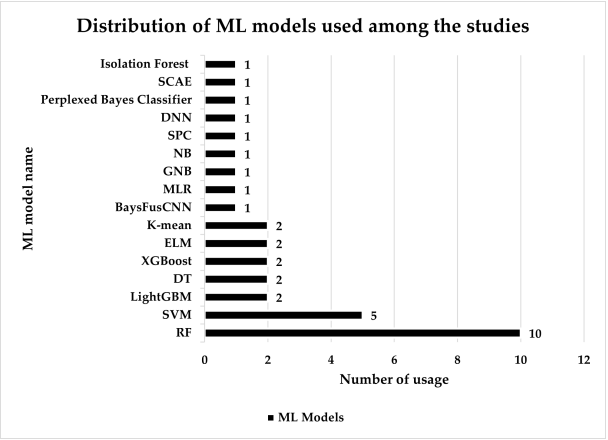

Fig. 8. Distribution of ML models used among the studies.

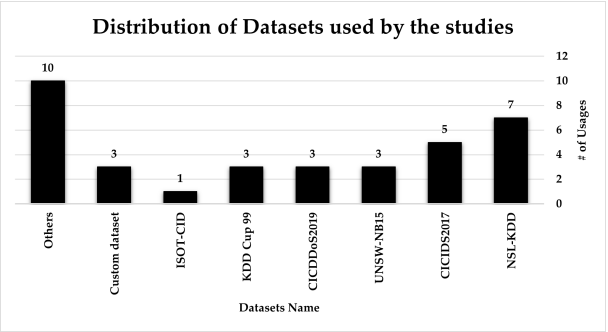

Fig. 9. Distribution of datasets used by the studies.

The challenges in imbalanced datasets and multi class anomaly detection are addressed by these hybrid and unsupervised models [52], [38].

Several advantages of using ML for anomaly detection in cloud environments are observed from the reviewed studies. Data dimensionality is reduced effectively, providing higher precision to a model using advanced feature selection techniques like LASSO, Chi-Square tests, and GAs [33], [38]. Often, hybrid approaches, combining clustering and supervised classification or unsupervised feature extraction with supervised classification are found to increase the accuracy and reduce the number of false positives [61], [62]. A second strength is scalability, and some studies have exploited cloud and fog computing to offload heavy computations and realize real time anomaly detection [45], [51].

However, the studies also identify notable limitations. Techniques such as genetic optimization and Bayesian networks improve accuracy add computational complexity, restricting their real time applicability [31], [43]. In many models, the datasets are highly dependent and based on CICIDS2017 and NSL KDD as shown in Fig. 9, which may limit their applicability in various and evolving cloud environments [41], [42]. Even high performing models such as DNNs continue to struggle with false positives [36]. Additionally, the experimental results are promising, but most systems are only tested on static datasets, so they are unable to handle real time detection or adapt to zero-day attacks [44], [53].

Feature handling is a critical factor for improving ML performance. Class imbalance and datasets are normalized using preprocessing techniques like Min Max scaling, SMOTE and ADASYN, that help improve model accuracy [33], [38]. LASSO and clustering (e.g., FCM–SVM hybrids) are adopted to perform feature sets optimization [42], [62]. In addition, several studies introduce new features, including rambling features and behavioral analytics, which can be used to enhance anomaly detection in a particular cloud scenario [46], [49].

RF and SVM models generally achieve high precision and recall, but perform poorly in detecting complex attack patterns [36], [43]. Techniques such as probabilistic methods and ELMs highlight speed and computational efficiency but the scalability is a concern with large scale datasets [62], [34]. The integration of fog and cloud improves system scalability at the cost of additional infrastructure [45], [51].

From these studies, valuable insights are drawn that emphasize the need for real time detection capability, including incremental learning in RT-AMD models that enable the models to adapt to dynamic threats [51]. However, the usage of outdated or synthetic datasets like KDD Cup 99 shows a need to have various and realistic benchmarks [31], [42]. Adaptive models, such as Bayesian methods and RL, are also studied by some to respond to changing threats, but these solutions are generally not optimized [43], [36]. Moreover, [36] emphasize ethical considerations especially on privacy preservation, in deep learning based systems.

IX. Challenges and Open Directions

- Development of Realistic Datasets: Future research should focus on creating and utilizing realistic, diverse, and up to date datasets that follow the real-world traffic patterns, including zero-day attacks and multi vector threats. This gap can be addressed by collaborative datasets derived from actual cloud systems.
- Scalable and Low-Latency Models: By optimizing ML models for computational efficiency, especially in the context of deep learning frameworks, it can help to deploy them in real time, high traffic environments. Incremental learning techniques, like those of RT-AMD, are promising areas for expansion [51].
- Advanced Feature Engineering: Additionally, incorporating novel features, behavioral analytic or domain specific characteristics rambling features proposed by [46] improve anomaly precision and reduce false positives.
- Integration of Hybrid Approaches: Addressing challenges such as data imbalance, generalization across datasets are possible by combining the strengths of unsupervised and supervised models such as the use of autoencoders for feature extraction and classes like SVM for detection.
- Enhanced Validation: Several, very different, datasets need to be validated across multiple models to promote generalizability. By benchmarking against real-time traffic from cloud providers [47], it can better evaluate their effectiveness in practical applications.
- Reduction of False Positives: Future studies include transferring ensemble methods or an advanced voting method (V-ELM) with the objective of reducing false positives and improving detection reliability [56].

X. Conclusion

Anomaly detection on cloud computing has emerged as an important application of ML and has provided us with robust tools to address evolving security challenges. RF and LightGBM models are very accurate on structured data and Isolation Forest capable of handling novel threats. Hybrid models that combine clustering with classification can deal with imbalanced datasets and with more complex attack patterns. Despite these advancements, limitations persist include the dependency on out-of-date dataset, computational inefficiencies, and high false positive rates. Future work must focus on developing realistic datasets that capture the dynamic nature of cloud environments, design of scalable models for real-time detection, and improvement of hybrid approaches to tradeoff between accuracy and adaptability. Techniques like incremental learning, advanced feature engineering, and ensemble methods should be prioritized to minimize false positives and improve model generalizability. Addressing these challenges will solidify the role of ML in protecting cloud infrastructures, towards making computing secure and resilient for a wide range of applications.

Funding

This work was funded by King Faisal University, Saudi Arabia. [Project No. GRANT KFU250561].

Acknowledgment

This work was supported through the Annual Funding track by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Project No. GRANT KFU250561].

Conflicts of Interest

All authors declare no conflict of interest.

References

[1] A. Bento, F. Araujo, and R. Barbosa, "Cost-availability aware scaling: Towards optimal scaling of cloud services," *Journal of Grid Computing*, vol. 21, no. 4, Dec. 2023. [Online]. Available: http://dx.doi.org/10.1007/s10723-023-09718-2

[2] H. Tabrizchi and M. Kuchaki Rafsanjani, "A survey on security challenges in cloud computing: issues, threats, and solutions," *The Journal of Supercomputing*, vol. 76, no. 12, p. 9493–9532, Feb. 2020. [Online]. Available: http://dx.doi.org/10.1007/s11227-020-03213-1

[3] M. K. Sasubilli and V. R, "Cloud computing security challenges, threats and vulnerabilities," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*. IEEE, Jan. 2021, p. 476–480. [Online]. Available: http://dx.doi.org/10.1109/ICICT50816.2021.9358709

[4] B. Morris-Grant, *Medibank data breach: Hackers release more sensitive customer information on dark web*, November 10 2022, accessed: 2024-12-05. [Online]. Available: https://www.abc.net.au/news/2022-11-10/medibank-data-breach-latest-dark-web-leak/101632746

[5] ImmuniWeb, *Top 10 Cloud Security Incidents in 2022*, 2022, accessed: 2024-12-05. [Online]. Available: https://www.immuniweb.com/blog/top-10-cloud-security-incidents-in-2022.html

[6] T. Group, *2023 Cloud Security: Cyberattacks and Data Breaches*, 2023, accessed: 2024-12-05. [Online]. Available: https://cpl.thalesgroup.com/about-us/newsroom/2023-cloud-security-cyberattacks-data-breaches-press-release

[7] F. Khoda Parast, C. Sindhav, S. Nikam, H. Izadi Yekta, K. B. Kent, and S. Hakak, "Cloud computing security: A survey of service-based models," *Computers & Security*, vol. 114, p. 102580, Mar. 2022. [Online]. Available: http://dx.doi.org/10.1016/j.cose.2021.102580

[8] R. Foorthuis, "On the nature and types of anomalies: a review of deviations in data," *International Journal of Data Science and Analytics*, vol. 12, no. 4, p. 297–331, Aug. 2021. [Online]. Available: http://dx.doi.org/10.1007/s41060-021-00265-1

[9] R. Singh, N. Srivastava, and A. Kumar, "Machine learning techniques for anomaly detection in network traffic," in *2021 Sixth International Conference on Image Information Processing (ICIIP)*. IEEE, Nov. 2021, p. 261–266. [Online]. Available: http://dx.doi.org/10.1109/ICIIP53038.2021.9702647

[10] R. Rajab Asaad and S. R. M. Zeebaree, "Enhancing security and privacy in distributed cloud environments: A review of protocols and mechanisms," *Academic Journal of Nawroz University*, vol. 13, no. 1, p. 476–488, Mar. 2024. [Online]. Available: http://dx.doi.org/10.25007/ajnu.v13n1a2010

[11] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine learning for anomaly detection: A systematic review," *IEEE Access*, vol. 9, p. 78658–78700, 2021. [Online]. Available: http://dx.doi.org/10.1109/ACCESS.2021.3083060

[12] S. A. Ali, D. Sujatha, R. Michael, G. Ramesh, and M. Agoramoorthy, "Leveraging machine learning for real-time anomaly detection and self-repair in iot devices," in *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)*. IEEE, Nov. 2023, p. 982–986. [Online]. Available: http://dx.doi.org/10.1109/ICCSAI59793.2023.10421539

[13] H. Hojjati, T. K. K. Ho, and N. Armanfard, "Self-supervised anomaly detection in computer vision and beyond: A survey and outlook," *Neural Networks*, vol. 172, p. 106106, Apr. 2024. [Online]. Available: http://dx.doi.org/10.1016/j.neunet.2024.106106

[14] W. Voorsluys, J. Broberg, and R. Buyya, "Introduction to cloud computing," *Cloud computing: Principles and paradigms*, pp. 1–41, 2011.

[15] P. Mell, "The nist definition of cloud computing," *NIST Special Publication*, pp. 800–145, 2011. [Online]. Available: https://faculty.winthrop.edu/domanm/csci411/Handouts/NIST.pdf

[16] V. Ashktorab, S. R. Taghizadeh *et al.*, "Security threats and countermeasures in cloud computing," *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, vol. 1, no. 2, pp. 234–245, 2012.

[17] K. Dineva and T. Atanasova, "Systematic look at machine learning algorithms–advantages, disadvantages and practical applications," *International Multidisciplinary Scientific GeoConference: SGEM*, vol. 20, no. 2.1, pp. 317–324, 2020.

[18] T. Islam, D. Manivannan, and S. Zeadally, "A classification and characterization of security threats in cloud computing," *Int. J. Next-Gener. Comput*, vol. 7, no. 1, pp. 268–285, 2016.

[19] J. B. Hong, A. Nhlabatsi, D. S. Kim, A. Hussein, N. Fetais, and K. M. Khan, "Systematic identification of threats in the cloud: A survey," *Computer Networks*, vol. 150, pp. 46–69, 2019.

[20] A. Babaei, P. M. Kebria, M. M. Dalvand, and S. Nahavandi, "A review of machine learning-based security in cloud computing," *arXiv preprint arXiv:2309.04911*, 2023.

[21] S. Iqbal, M. L. M. Kiah, B. Dhaghighi, M. Hussain, S. Khan, M. K. Khan, and K.-K. R. Choo, "On cloud security attacks: A taxonomy and intrusion detection and prevention as a service," *Journal of Network and Computer Applications*, vol. 74, pp. 98–120, 2016.

[22] X. Xia, X. Pan, N. Li, X. He, L. Ma, X. Zhang, and N. Ding, "Gan-based anomaly detection: A review," *Neurocomputing*, vol. 493, p. 497–535, Jul. 2022. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2021.12.093

[23] J. Jot and P. L. S. Sharma, "Study of anomaly detection in iot sensors," *International Journal for Research in Applied Science and Engineering Technology*, vol. 11, no. 8, p. 767–774, Aug. 2023. [Online]. Available: http://dx.doi.org/10.22214/ijraset.2023.55226

[24] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Systems with Applications*, vol. 166, p. 114060, Mar. 2021. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2020.114060

[25] M. Ravinder and V. Kulkarni, "A review on cyber security and anomaly detection perspectives of smart grid," in *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE, Jan. 2023, p. 692–697. [Online]. Available: http://dx.doi.org/10.1109/ICSSIT55814.2023.10060871

[26] M. Adiban, S. M. Siniscalchi, and G. Salvi, "A step-by-step training method for multi generator gans with application to anomaly detection and cybersecurity," *Neurocomputing*, vol. 537, p. 296–308, Jun. 2023. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2023.03.056

[27] R. Qi, C. Rasband, J. Zheng, and R. Longoria, "Detecting cyber attacks in smart grids using semi-supervised anomaly detection and deep representation learning," *Information*, vol. 12, no. 8, p. 328, Aug. 2021. [Online]. Available: http://dx.doi.org/10.3390/info12080328

[28] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, Mar. 2021. [Online]. Available: http://dx.doi.org/10.1007/s42979-021-00592-x

[29] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, p. 757–774, Feb. 2023. [Online]. Available: http://dx.doi.org/10.1016/j.jksuci.2023.01.014

[30] M. Namdev, S. Jayasundar, M. Babur, D. A. Vidhate, and S. Yerasuri, "Enhancing security in cloud computing with anomaly detection using machine learning," *Tuijin Jishu/Journal of Propulsion Technology*, vol. 44, no. 3, pp. 1923–1931, 2023. [Online]. Available: https://www.propulsiontechjournal.com/index.php/journal/article/view/622

[31] F. Talpur, I. A. Korejo, A. A. Chandio, A. Ghulam, and M. S. H. Talpur, "Ml-based detection of ddos attacks using evolutionary algorithms optimization," *Sensors*, vol. 24, no. 5, p. 1672, Mar. 2024. [Online]. Available: http://dx.doi.org/10.3390/s24051672

[32] M. Alduailij, Q. W. Khan, M. Tahir, M. Sardaraz, M. Alduailij, and F. Malik, "Machine-learning-based ddos attack detection using mutual information and random forest feature importance method," *Symmetry*, vol. 14, no. 6, p. 1095, May 2022. [Online]. Available: http://dx.doi.org/10.3390/sym14061095

[33] S. Dasari and R. Kaluri, "An effective classification of ddos attacks in a distributed network by adopting hierarchical machine learning and hyperparameters optimization techniques," *IEEE Access*, vol. 12, p. 10834–10845, 2024. [Online]. Available: http://dx.doi.org/10.1109/ACCESS.2024.3352281

[34] N. Mishra, R. K. Singh, and S. K. Yadav, "Detection of ddos vulnerability in cloud computing using the perplexed bayes classifier," *Computational Intelligence and Neuroscience*, vol. 2022, p. 1–13, Jul. 2022. [Online]. Available: http://dx.doi.org/10.1155/2022/9151847

[35] P. Parameswarappa, T. Shah, and G. R. Lanke, "A machine learning-based approach for anomaly detection for secure cloud computing environments," in *2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, Jan. 2023, p. 931–940. [Online]. Available: http://dx.doi.org/10.1109/IDCIoT56793.2023.10053518

[36] M. Dhinakaran, M. Sundhari, S. Ambika, V. Balaji, and R. Rajasekaran, "Advanced machine learning techniques for enhancing data security in cloud computing systems," in *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*. IEEE, Feb. 2024, p. 1598–1602. [Online]. Available: http://dx.doi.org/10.1109/IC2PCT60090.2024.10486559

[37] R. Abubakar, A. Aldegheishem, M. Faran Majeed, A. Mehmood, H. Maryam, N. Ali Alrajeh, C. Maple, and M. Jawad, "An effective mechanism to mitigate real-time ddos attack," *IEEE Access*, vol. 8, p. 126215–126227, 2020. [Online]. Available: http://dx.doi.org/10.1109/ACCESS.2020.2995820

[38] M. Bakro, R. R. Kumar, A. Alabrah, Z. Ashraf, M. N. Ahmed, M. Shameem, and A. Abdelsalam, "An improved design for a cloud intrusion detection system using hybrid features selection approach with ml classifier," *IEEE Access*, vol. 11, p. 64228–64247, 2023. [Online]. Available: http://dx.doi.org/10.1109/ACCESS.2023.3289405

[39] M. Bakro, R. R. Kumar, M. Husain, Z. Ashraf, A. Ali, S. I. Yaqoob, M. N. Ahmed, and N. Parveen, "Building a cloud-ids by hybrid bio-inspired feature selection algorithms along with random forest model," *IEEE Access*, vol. 12, p. 8846–8874, 2024. [Online]. Available: http://dx.doi.org/10.1109/ACCESS.2024.3353055

[40] Z. Chkirbene, R. Hamila, A. Erbad, S. Kiranyaz, N. Al-Emadi, and M. Hamdi, "Cooperative machine learning techniques for cloud intrusion detection," in *2021 International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, Jun. 2021, p. 837–842. [Online]. Available: http://dx.doi.org/10.1109/IWCMC51323.2021.9498809

[41] A. Aldallal and F. Alisa, "Effective intrusion detection system to secure data in cloud using machine learning," *Symmetry*, vol. 13, no. 12, p. 2306, Dec. 2021. [Online]. Available: http://dx.doi.org/10.3390/sym13122306

[42] A. N. Jaber and S. U. Rehman, "Fcm–svm based intrusion detection system for cloud computing environment," *Cluster Computing*, vol. 23, no. 4, p. 3221–3231, Mar. 2020. [Online]. Available: http://dx.doi.org/10.1007/s10586-020-03082-6

[43] I. AlSaleh, A. Al-Samawi, and L. Nissirat, "Novel machine learning approach for ddos cloud detection: Bayesian-based cnn and data fusion enhancements," *Sensors*, vol. 24, no. 5, p. 1418, Feb. 2024. [Online]. Available: http://dx.doi.org/10.3390/s24051418

[44] P. Sherubha, S. Sasirekha, A. D. K. Anguraj, J. V. Rani, R. Anitha, S. P. Praveen, and R. H. Krishnan, "An efficient unsupervised learning approach for detecting anomaly in cloud." *Comput. Syst. Sci. Eng.*, vol. 45, no. 1, pp. 149–166, 2023.

[45] D. A. Moreira, H. P. Marques, W. L. Costa, J. Celestino, R. L. Gomes, and M. Nogueira, "Anomaly detection in smart environments using ai over fog and cloud computing," in *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2021, pp. 1–2.

[46] A. Alshammari and A. Aldribi, "Apply machine learning techniques to detect malicious network traffic in cloud computing," *Journal of Big Data*, vol. 8, no. 1, p. 90, 2021.

[47] M. I. S. Al-jumaili and J. Bazzi, "Cyber-attack detection for cloud-based intrusion detection systems," *Mesopotamian Journal of CyberSecurity*, vol. 2023, pp. 170–182, 2023.

[48] S. Naiem, A. E. Khedr, A. M. Idrees, and M. I. Marie, "Enhancing the efficiency of gaussian naïve bayes machine learning classifier in the detection of ddos in cloud computing," *IEEE Access*, vol. 11, pp. 124 597–124 608, 2023.

[49] Ö. Aslan, M. Ozkan-Okay, and D. Gupta, "Intelligent behavior-based malware detection system on cloud computing environment," *IEEE Access*, vol. 9, pp. 83 252–83 271, 2021.

[50] M. Mehmood, R. Amin, M. M. A. Muslam, J. Xie, and H. Aldabbas, "Privilege escalation attack detection and mitigation in cloud using machine learning," *IEEE Access*, vol. 11, pp. 46 561–46 576, 2023.

[51] O. Bamasag, A. Alsaeedi, A. Munshi, D. Alghazzawi, S. Alshehri, and A. Jamjoom, "Real-time ddos flood attack monitoring and detection (rt-amd) model for cloud computing," *PeerJ Computer Science*, vol. 7, p. e814, 2022.

[52] Z. Chkirbene, A. Erbad, R. Hamila, A. Gouissem, A. Mohamed, and M. Hamdi, "Machine learning based cloud computing anomalies detection," *IEEE Network*, vol. 34, no. 6, pp. 178–183, 2020.

[53] S. Sambangi and L. Gondi, "A machine learning approach for ddos (distributed denial of service) attack detection using multiple linear regression," in *Proceedings*, vol. 63, no. 1. MDPI, 2020, p. 51.

[54] A. R. Wani, Q. Rana, and N. Pandey, "Machine learning solutions for analysis and detection of ddos attacks in cloud computing environment," *Int. J. Eng. Adv. Technol*, vol. 9, no. 3, pp. 2205–2209, 2020.

[55] P. Preethi, P. Ramadevi, K. Akshaya, S. Sangamitra, and A. Pritikha, "Analysis of phishing attack in distributed cloud systems using machine learning," in *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*. IEEE, 2023, pp. 1–5.

[56] G. S. Kushwah and V. Ranga, "Voting extreme learning machine based distributed denial of service attack detection in cloud computing," *Journal of Information Security and Applications*, vol. 53, p. 102532, 2020.

[57] H. Attou, A. Guezzaz, S. Benkirane, M. Azrour, and Y. Farhaoui, "Cloud-based intrusion detection approach using machine learning techniques," *Big Data Mining and Analytics*, vol. 6, no. 3, pp. 311–320, 2023.

[58] K. Shanthi and R. Maruthi, "Machine learning approach for anomaly-based intrusion detection systems using isolation forest model and support vector machine," in *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, Aug. 2023, p. 136–139. [Online]. Available: http://dx.doi.org/10.1109/ICIRCA57980.2023.10220620

[59] P. Ntambu and S. A. Adeshina, "Machine learning-based anomalies detection in cloud virtual machine resource usage," in *2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS)*. IEEE, 2021, pp. 1–6.

[60] W. Wang, X. Du, D. Shan, R. Qin, and N. Wang, "Cloud intrusion detection method based on stacked contractive auto-encoder and support vector machine," *IEEE Transactions on Cloud Computing*, vol. 10, no. 3, p. 1634–1646, Jul. 2022. [Online]. Available: http://dx.doi.org/10.1109/TCC.2020.3001017

[61] K. Samunnisa, G. S. V. Kumar, and K. Madhavi, "Intrusion detection system in distributed cloud computing: Hybrid clustering and classification methods," *Measurement: Sensors*, vol. 25, p. 100612, Feb. 2023. [Online]. Available: http://dx.doi.org/10.1016/j.measen.2022.100612

[62] L. Megouache, A. Zitouni, S. Sadouni, and M. Djoudi, "Machine learning for cloud data classification and anomaly intrusion detection," *Revue des Sciences et Technologies de l'Information-Série ISI: Ingénierie des Systèmes d'Information*, vol. 29, no. 05, pp. 1809–1819, 2024.