International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Proactive Detection of Attacks on Cloud-based Applications using Machine Learning

S. Rekha Garikamukkala*[1], V. Ravi Sankar [2]

Submitted:14/03/2024 **Revised**: 29/04/2024 **Accepted**: 06/05/2024

Abstract: This study thoroughly examines how machine learning techniques may be used to proactively detect assaults on cloud-based services. The security of cloud-based systems has become crucial due to the growing dependence on cloud computing for a wide range of applications in many industries. Conventional security methods frequently fail to identify advanced threats that take use of weaknesses in cloud infrastructures and apps. Therefore, there is a critical requirement for security measures that are proactive and adaptable, able to detect and address new threats as they develop instantly. This study focuses on incorporating machine learning methods to enhance the security of cloud-based systems. Our platform utilizes past data on system behaviors, network traffic patterns, and application interactions to use machine learning in distinguishing regular operations from abnormal activity that may signal possible assaults. Our method creates strong detection systems that can adjust to changing threat environments by utilizing feature extraction, dimensionality reduction, and model training. Our technique focuses on creating a comprehensive detection system that includes anomaly detection, intrusion detection, and behavior analysis. Our system is versatile in identifying various assaults such as DDoS attacks, SQL injection, cross-site scripting, and data exfiltration attempts by combining supervised, unsupervised, and semi-supervised learning approaches. Our technique aims to reduce false positives and negatives by detecting discriminative features and reducing noise, thereby enhancing detection accuracy and reliability. The scalability and efficiency of the proposed framework are crucial due to the dynamic and resource-limited nature of cloud infrastructures. We investigate lightweight machine learning techniques and distributed computing architectures that can easily integrate with cloud settings while reducing computational overhead. This study showcases the effectiveness and durability of our proactive detection architecture in protecting cloud-based apps from various cyber threats, through thorough testing and assessment using real-world datasets and simulated attack scenarios. Organizations may reduce the risks of cyber-attacks and protect vital assets, data integrity, and user trust in cloud computing ecosystems by adopting a proactive security approach based on machine learning insights.

Keywords: *Machine Learning, Cloud-Based Applications, Proactive Detection, Cybersecurity, Anomaly Detection, Intrusion Detection*

1. Introduction

Cloud computing is a fundamental technology in the current digital transformation era, changing the ways computing, storage, and service provision are approached. Cloud-based apps have brought about increased scalability, agility, and cost-effectiveness for enterprises in many industries. This allows them to move beyond the limitations of traditional IT infrastructure and adopt a more flexible and responsive operational approach. Despite the numerous advantages of cloud computing, the threat of cybersecurity poses a significant risk to the security and privacy of sensitive data stored in cloud infrastructures.

The fast increase in cloud-based services has brought about a new era of cyber dangers and vulnerabilities, with malevolent individuals aiming to take advantage of holes in cloud structures and applications. The security concerns faced by enterprises globally in cloud-based deployments are diverse and constantly changing, including DDoS assaults, data breaches, and ransomware campaigns. Traditional security methods focused on perimeter defenses and signature-based detection are frequently inadequate in dealing with the complexity and magnitude of contemporary cyber threats, exposing businesses to potential exploitation and penetration. The combination of machine learning and cyber security has emerged as a powerful force to enhance the effectiveness and resilience of cloud-based defenses in light of the increasing threat landscape. Machine learning is a fundamental change in cyber security that enables businesses to transition from static rule sets and manual intervention to a more proactive and adaptable security approach. Machine learning algorithms can detect abnormal activity that may indicate future attacks in real-time by using data-driven analytics, pattern recognition, and predictive modeling. Proactive detection involves anticipating and preventing new risks before they develop into security breaches.

Proactive detection uses historical data and contextual insights to identify new and changing threats accurately and quickly, instead of relying on predetermined signatures and established attack patterns like reactive techniques. Machine learning algorithms can analyze large amounts of

[1]*Research Scholar, Department of Computer Science and Engineering, GITAM (Deemed to be University), Hyderabad, India.*
Email:sunitha.garikamukkala@gmail.com
ORCID: 0009-0006-3816-984X
**(Corresponding Author)*
[2]*Associate Professor, Department of Computer Science and Engineering, GITAM (Deemed to be University), Hyderabad, India.*
Email:rvadali@gitam.edu
ORCID: 0000-0002-9104-9509

information to identify patterns and correlations, allowing for autonomous decision-making and adaptive responses in dynamic cloud settings through repeated model training, validation, and refinement.

Furthermore, the inherent scalability and adaptability of machine learning frameworks make them suitable for the intricate and dispersed characteristics of cloud infrastructures. As enterprises increasingly use cloud computing for important workloads and data-heavy applications, the demand for scalable and resource-efficient security measures becomes more crucial. Machine learning techniques, including supervised, unsupervised, deep learning, and reinforcement learning, provide a flexible set of tools for identifying and reducing various cyber risks in different cloud contexts, such as public, private, and hybrid installations.

This study aims to clarify the ideas, methodology, and ramifications of proactively detecting assaults on cloud-based services using machine learning. We want to analyze the effectiveness, obstacles, and potential advancements of machine learning-based security systems in protecting cloud infrastructures from cyber-attacks by examining current literature, case studies, and experimental results.

2. Foundational Method for Attack Detection

The fundamental approach for identifying and preventing attacks serves as the basis of strong cybersecurity strategies, equipping companies with the required tools and knowledge to protect their digital assets from a wide range of threats. This section explores the core ideas and approaches that support efficient attack detection systems. It establishes the foundation for proactive threat mitigation in dynamic and diverse computing environments. Our objective is to provide practitioners and researchers with a thorough grasp of the techniques and algorithms used to identify and address harmful activity by explaining the fundamental principles of anomaly detection, intrusion detection, and behavior analysis. Using machine learning and data-driven analytics, we examine how feature engineering, model selection, and assessment metrics play a crucial role in improving the effectiveness and scalability of attack detection systems. In addition, we analyze the practical elements and compromises involved in creating and executing attack detection systems, including important aspects such as computing burden, rate of false positives, and ability to respond in real-time. This section aims to provide stakeholders with the required information and insights to strengthen their defenses and reduce the risks associated with new cyber threats by combining theoretical insights with practical implementations.

Further, the mathematical model for the foundational method is furnished here.

Assuming that the dataset, DS[] is the collection network characteristics with n number of parameters denoted as P[i] and the class variable is denoted as C[i]. Thus, this can be formulated as,

$$DS[] =< P[], C[] > \tag{1}$$

The traditional method directs that based on the class variable, C[], the dataset must be clusters. Assuming that, X[] is the collection of clusters, then the following formulation can be furnished,

$$X[] = \sum_{j=1}^{m} \sum_{i=1}^{n} \left| C[i] - Mean\{C[]\} \right| \tag{2}$$

Once the clusters are formed, the relevant parameters can be classified based on the formed clusters. The following method helps in achieving the same,

$$DS1[].X[i] = \prod_{X[i].C[i]} DS[] \tag{3}$$

Where, DS1[] is the newly formed classified dataset.

Further, for any given test set, TS[], the validation can be formed as,

$$A \leftarrow Iff\ TS[] \subset DS1[].X[m] \tag{4}$$

Where A is event, which is detected as Attack.

Or,

$$NA \leftarrow Iff\ TS[] \not\subset DS1[].X[m] \tag{5}$$

Where NA is event, which is detected as Normal or Not Attack.

However, this method does not completely enable the attack detection process and the drawbacks are furnished in the further sections in this paper.

3. Understanding of Attack Types

In the age of cloud computing, when data and applications are progressively moving to distant servers and virtualized environments, the security situation has grown more intricate and fluid than ever. Cloud-based networks, while their exceptional scalability and flexibility, are also appealing to unscrupulous individuals who aim to exploit weaknesses and compromise valuable data. Organizations must possess a comprehensive understanding of the many forms of assaults that might specifically target networks based on cloud technology. This knowledge is crucial for strengthening their defensive measures and efficiently reducing possible risks. This note provides a thorough examination of several attack types frequently observed in cloud systems, including detailed insights into their methodology, implications, and solutions for mitigating them.

A. Denial of Service (DoS) and Distributed Denial of Service (DDoS) Attacks:

Denial of Service (DoS) and Distributed Denial of Service (DDoS) assaults continue to be widespread risks in the field of cloud computing. These attacks are designed to interrupt the availability of services and overwhelm network resources. A Denial of Service (DoS) attack occurs when a solitary entity inundates a certain system or network with an overwhelming number of requests, resulting in the system becoming unavailable to authorized users. DDoS assaults, in contrast, entail organized endeavors from several hacked devices or botnets to coordinate a substantial surge of traffic, intensifying the damage and rendering response more difficult. In order to protect against Denial of Service (DoS) and Distributed Denial of Service (DDoS) assaults, cloud providers and organizations utilize traffic filtering, rate limiting, and distributed mitigation mechanisms to detect and counteract unwanted network traffic, all while ensuring that their services remain accessible.

B. Incidents of data breaches and unauthorized access:

Data breaches pose a substantial risk to the security and integrity of data held in cloud-based networks. These breaches can happen through several methods, such as taking advantage of improperly set access restrictions, vulnerable authentication protocols, or compromised credentials. Malicious individuals may unlawfully remove sensitive material, such as client data, intellectual property, or financial records, which can have significant repercussions for impacted firms in terms of regulatory adherence, harm to reputation, and financial detriment. Deploying resilient encryption, stringent access restrictions, and vigilant monitoring methods are essential measures for reducing the likelihood of data breaches and unlawful entry in cloud settings.

C. Incidents of Malware and Ransomware Attacks:

Malicious software and ransomware assaults provide significant risks to networks that rely on cloud technology, since they infiltrate systems and encrypt vital data in order to collect ransom payments from those affected. The assaults can spread through several channels, such as harmful email attachments, hacked websites, or weak software components. After gaining access to the cloud environment, malware has the ability to spread horizontally, infecting several systems and putting the security of data and applications at risk. Organizations utilize a mix of endpoint protection, network segmentation, and threat intelligence to efficiently identify, control, and resolve infections caused by malware and ransomware assaults, therefore reducing the associated risks.

D. Risks posed by individuals with insider access and misuse of privileges:

Insider attacks, whether intentional or accidental, pose a major problem for cloud-based networks, as trusted individuals with special access may use their privileges to undermine systems or steal important data. Insider dangers can appear in several ways, such as dissatisfied workers, negligent contractors, or hacked accounts with excessive privileges. To effectively address insider threats, it is necessary to employ a comprehensive strategy that includes user behavior analytics, access restrictions, and privilege monitoring. This approach enables the proactive identification of abnormal activity and the mitigation of potential hazards.

E. SQL Injection and Cross-Site Scripting (XSS) Attacks:

SQL injection and cross-site scripting (XSS) attacks continue to be significant risks for online applications hosted in cloud settings. These attacks enable malicious actors to alter SQL queries or insert harmful scripts into web pages, resulting in the extraction of sensitive data or the hijacking of user sessions. These vulnerabilities frequently arise from poor input validation, inadequate parameterization, or unsafe coding techniques. In order to reduce the likelihood of SQL injection and XSS attacks, businesses regularly do code reviews, deploy web application firewalls (WAFs), and sanitize user inputs to avoid the malicious exploitation of vulnerabilities.

To effectively protect cloud-based networks from various attack types, it is essential to adopt a proactive and multi-layered strategy. This strategy should include threat intelligence, strong authentication systems, encryption, and constant monitoring. Organizations may enhance their defenses and protect their assets from developing cyber threats in the dynamic world of cloud computing by comprehending the methodology, implications, and mitigation measures related to different attack types.

4. Recent Works

Cloud computing has significantly transformed the field of information technology by providing unparalleled scalability, flexibility, and resource accessibility. The paradigm change has introduced new security concerns, such as denial-of-service (DoS) and distributed denial-of-service (DDoS) attacks, which are serious risks to cloud infrastructures. Researchers and practitioners have put significant effort into creating strong ways to identify and prevent assaults, using a variety of strategies from classic to advanced approaches.

Janitza Nicole Punto Gutierrez and Kilhung Lee (2020) suggested a specialized filtering system designed to identify slow-rate denial-of-service (DoS) assaults in cloud systems. Their plan focuses on distinguishing between genuine and malicious traffic by analyzing traffic patterns and detecting abnormalities that suggest active assaults. Nagarathna Ravi and S. Mercy Shalinie (2020) explored the field of Internet of Things (IoT) and introduced a learning-based method

combined with Software-Defined Networking (SDN) and cloud architecture to identify and counteract DDoS attacks, strengthening the IoT environment against malicious actions [2].

Gopal Singh Kushwah and Virender Ranga (2020) researched extreme learning machines for detecting DDoS attacks in cloud computing. They proposed a distributed detection framework that utilizes interconnected nodes' collective intelligence to improve detection accuracy and efficiency [3]. Shweta Gumaste and colleagues (2020) developed a new method using Apache Spark to identify DDoS attacks in OpenStack-based private clouds, demonstrating the effectiveness of big data analytics in enhancing cloud security against cyber threats [4].

Omar Abdel Wahab, Jamal Bentahar, Hadi Otrok, and Azzam Mourad (2020) discussed the difficulties of identifying VM-based DDoS assaults in cloud environments. They recommended implementing efficient load distribution techniques to enhance attack detection and enhance system durability. Aanshi Bhardwaj, Veenu Mangat, and Renu Vig (2020) suggested using a Hyperband-tuned deep neural network with a stacked sparse autoencoder to detect DDoS attacks. This approach leverages artificial intelligence and deep learning to strengthen cloud security against malicious intrusions.

Abhishek Agarwal, Ayush Prasad, Rishabh Rustogi, and Sweta Mishra (2021) developed a deep learning method to detect and prevent fraudulent resource consumption attacks in cloud environments, emphasizing the importance of adaptive and intelligent security measures in protecting cloud infrastructures. Prasanna Balaji Narasingapuram and M. Ponnavaikko (2021) introduced a new system for detecting attacks and encrypting data in distributed cloud computing. They highlighted the importance of encryption in reducing security risks and safeguarding data privacy.

Meghana G. Raj and Santosh Kumar Pani (2021) performed a meta-analytic assessment of advanced intrusion detection methods in cloud computing settings, offering a comprehensive overview of current techniques and highlighting new developments and obstacles in cloud security. Parul Singh and Virender Ranga (2021) investigated ensemble learning methods for detecting attacks and intrusions in cloud computing. They utilized numerous detection models to improve detection accuracy and resilience.

S. Velliangiri, P. Karthikeyan, and V. Vinoth Kumar (2021) researched optimization-based deep networks to detect distributed denial-of-service (DDoS) attacks in cloud computing. They recommended using optimization techniques to improve the learning process and boost the efficiency of attack detection mechanisms. Reddy SaiSindhuTheja and Gopal K. Shyam (2021) introduced a metaheuristic algorithm-driven feature selection technique combined with recurrent neural networks to identify DoS attacks. This study demonstrates the effectiveness of using hybrid approaches to enhance the security of cloud infrastructures against cyber threats.

Isaac Odun-Ayo, Williams Toro-Abasi, Marion Adebiyi, and Oladapo Alagbe (2021) developed a real-time detection system for cross-site scripting attacks on cloud-based web applications using deep learning, emphasizing the significance of proactive security measures in reducing web-based vulnerabilities [13]. Fargana J. Abdullayeva (2021) presented a method for detecting advanced persistent threat assaults using an autoencoder and softmax regression algorithm. This study highlighted the effectiveness of anomaly detection approaches in strengthening cloud infrastructures against complex attacks.

Sandeep Kautish, A. Reyana, and Ankit Vidyarthi (2022) introduced SDMTA, a method for identifying and addressing DDoS weaknesses in hybrid cloud setups by combining the advantages of cloud and on-premises systems to enhance security. Ahmed Abdullah Alqarni (2022) proposed using a majority vote-based ensemble strategy for detecting DDoS attacks, leveraging collective intelligence to improve detection accuracy and resilience [16].

M. Arunkumar and K. Ashok Kumar (2022) studied machine learning methods to identify harmful attacks in cloud computing. They demonstrated the effectiveness of support vector machines and optimization algorithms in strengthening cloud infrastructures against cyber threats [17]. Omaimah Bamasag and colleagues (2022) created a methodology for monitoring and detecting real-time DDoS flood assaults, emphasizing the significance of proactive monitoring and prompt response to reduce DDoS attacks.

Fargana J. Abdullayeva (2022) broadened the focus to E-government clouds and introduced a method for detecting distributed denial-of-service attacks using data clustering, highlighting the importance of customized security measures in specific cloud settings [19]. Theyazn H.H. Aldhyani and Hasan Alkahtani (2022) investigated artificial intelligence algorithms for detecting economic denial of sustainability (EDoS) attacks in cloud computing settings, emphasizing the growing risks and the necessity for flexible security solutions [20].

Vasily Desnitsky, Andrey Chechulin, and Igor Kotenko (2022) proposed a multi-aspect-based method for detecting attacks in IoT clouds, emphasizing the intricate nature of IoT environments and the importance of comprehensive security tactics [21]. Ryuga Kaneko and Taiichi Saito (2023) discussed the need of using proactive threat intelligence and security monitoring to detect cookie bomb attacks in cloud computing settings and mitigate new risks.

Yousef Sanjalawe and Turke Althobaiti (2023) suggested combining ensemble feature selection with deep learning to

enhance DDoS attack detection in cloud computing, utilizing machine learning and feature engineering to improve attack detection skills. S. Balasubramaniam et al. (2023) supported the use of optimization-enabled deep learning for detecting DDoS attacks, highlighting the significance of optimization methods in improving the effectiveness and scalability of detection systems [24].

In 2023, B. Dhiyanesh and colleagues investigated the iterative dichotomiser posteriori approach for detecting service attacks in cloud computing. Their study demonstrated the adaptability of machine learning techniques in combating various cyber threats. Muhammad Mehmood and colleagues (2023) addressed the detection and prevention of privilege escalation attacks in cloud computing through the application of machine learning. Their work emphasizes the changing landscape of cyber threats and the necessity for flexible security protocols. Animesh Kumar, Sandip Dutta, and Prashant Pranav (2023) focused on supervised learning for detecting attacks in cloud systems, highlighting the significance of labeled datasets and supervised methods in developing strong detection models [27]. Sasha Mahdavi Hezavehi and Rouhollah Rahmani (2023) presented a new approach for detecting DDoS attacks using anomalies. They utilized third-party auditors to enhance security and improve the ability to identify attacks.

In 2023, M. Arunkumar and K. Ashok Kumar introduced GOSVM, a Gannet optimization-based support vector machine designed for detecting harmful attacks in the cloud. Their work highlights the effectiveness of nature-inspired optimization techniques in strengthening cloud infrastructures against cyber threats. Adel Binbusayyis (2024) proposed using a combination of VGG19 and 2D-CNN for intrusion detection in the FOG-cloud environment, emphasizing the need of customized detection methods in new computing models [30].

The literature study discusses a wide range of methods and strategies used to identify and address different types of cyber risks in cloud computing settings. Researchers and practitioners are constantly developing and upgrading classic machine learning methods and cutting-edge deep learning algorithms to protect cloud infrastructures from emerging threats and ensure their security. The summary of the work is provided here [Table – 1].

Table 1. Summary of Related Works

Author, Year	***Proposed Method***	***Research Limitations***
Janitza Nicole Punto Gutierrez, & Kilhung Lee (2020)[1]	Attack-based Filtering Scheme	May lack scalability for rapidly evolving attack patterns
Nagarathna Ravi, & S. Mercy Shalinie (2020)[2]	Learning-Driven Detection, SDN-Cloud Architecture	Dependency on accurate learning models and SDN infrastructure
Gopal Singh Kushwah, & Virender Ranga (2020)[3]	Voting Extreme Learning Machine	Sensitivity to noisy or unbalanced data
Shweta Gumaste, D. G. Narayan, Sumedha Shinde, & K. Amit (2020)[4]	Detection using Apache Spark	Resource-intensive, requires specialized expertise for deployment
Omar Abdel Wahab, Jamal Bentahar, Hadi Otrok, & Azzam Mourad (2020)[5]	Optimal Load Distribution for VM-Based DDoS Attack Detection in Cloud	Complexity in determining optimal thresholds and load balancing policies
Aanshi Bhardwaj, Veenu Mangat, & Renu Vig (2020)[6]	Hyperband Tuned Deep Neural Network with Stacked Sparse Autoencoder	High computational requirements during training and tuning
Abhishek Agarwal, Ayush Prasad, Rishabh Rustogi, & Sweta Mishra (2021)[7]	Deep Learning Approach for Fraudulent Resource Consumption Attack Detection in Cloud	Reliance on large datasets and computational resources for effective training
Prasanna Balaji Narasingapuram, & M. Ponnavaikko (2021)[8]	Novel Attack Detection and Encryption Framework for Distributed Cloud Computing	Overhead introduced by encryption processes and potential impact on system performance
Meghana G. Raj, & Santosh Kumar Pani (2021)[9]	Meta-analytic Review of Intelligent Intrusion Detection Techniques in Cloud Computing	Reliance on existing literature and potential biases in selection and interpretation of studies
Parul Singh, & Virender Ranga (2021)[10]	Ensemble Learning Approach for Attack and Intrusion Detection in	Sensitivity to ensemble model configurations and potential overfitting

Author, Year	*Proposed Method*	*Research Limitations*
	Cloud Computing	
S. Velliangiri, P. Karthikeyan, & V. Vinoth Kumar (2021)[11]	Optimization-Based Deep Networks for DDoS Attack Detection in Cloud Computing	Complexity in optimization algorithms and tuning parameters
Reddy SaiSindhuTheja, & Gopal K. Shyam (2021)[12]	Metaheuristic Algorithm with Recurrent Neural Network for DoS Attack Detection in Cloud	Limited generalizability to different attack scenarios and system configurations
Isaac Odun-Ayo, Williams Toro-Abasi, Marion Adebiyi, & Oladapo Alagbe (2021)[13]	Real-time Detection of Cross-Site Scripting Attacks using Deep Learning on Cloud-based Web Apps	Dependency on real-time data feeds and potential delays in response times
Fargana J. Abdullayeva (2021)[14]	Autoencoder and Softmax Regression for Advanced Persistent Threat Attack Detection in Cloud	Sensitivity to model hyperparameters and potential overfitting
Sandeep Kautish, A. Reyana, & Ankit Vidyarthi (2022)[15]	SDMTA: DDoS Attack Detection and Mitigation in Hybrid Cloud Environment	Dependency on accurate detection thresholds and potential overhead in mitigation processes
Ahmed Abdullah Alqarni (2022)[16]	Majority Vote-Based Ensemble Approach for DDoS Attack Detection in Cloud Computing	Sensitivity to ensemble model configurations and potential overfitting
M. Arunkumar, & K. Ashok Kumar (2022)[17]	Malicious Attack Detection using Machine Learning Techniques in	Limited generalizability across diverse attack scenarios and potential

Author, Year	*Proposed Method*	*Research Limitations*
	Cloud Computing	bias in training data
Omaimah Bamasag, Alaa Alsaeedi, Asmaa Munshi, Daniyal Alghazzawi, Suhair Alshehri, & Arwa Jamjoom (2022)[18]	Real-time DDoS Flood Attack Monitoring and Detection Model for Cloud Computing	Dependency on real-time data feeds and potential delays in response times
Fargana J. Abdullayeva (2022)[19]	Data Clustering for DDoS Attack Detection in E-Government Cloud	Sensitivity to clustering algorithm parameters and potential bias in cluster assignment
Theyazn H.H. Aldhyani, & Hasan Alkahtani (2022)[20]	AI Algorithm-Based Detection of Economic Denial of Sustainability Attacks in Cloud	Complexity in determining economic indicators and potential false positives
Vasily Desnitsky, Andrey Chechulin, & Igor Kotenko (2022)[21]	Multi-Aspect Based Approach to Attack Detection in IoT Clouds	Dependency on accurate feature extraction methods and potential scalability issues
Ryuga Kaneko, & Taiichi Saito (2023)[22]	SIEM Monitored Detection of Cookie Bomb Attacks in Cloud Computing Environment	Complexity in SIEM configuration and potential overhead in monitoring
Yousef Sanjalawe, & Turke Althobaiti (2023)[23]	Ensemble Feature Selection and Deep Learning for DDoS Attack Detection in Cloud Computing	Sensitivity to ensemble model configurations and potential overfitting
S. Balasubramaniam, C. Vijesh Joe, T. A. Sivakumar, A. Prasanth, K. Satheesh Kumar,	Optimization-Enabled Deep Learning for DDoS Attack Detection in	Complexity in optimization algorithms and tuning parameters

Author, Year	*Proposed Method*	*Research Limitations*
V. Kavitha, & Rajesh Kumar Dhanaraj (2023)[24]	Cloud Computing	
B. Dhiyanesh, K. Karthick, R. Radha, & Anita Venaik (2023)[25]	Iterative Dichotomiser Posteriori Method for Service Attack Detection in Cloud Computing	Sensitivity to decision boundary parameters and potential bias in training data
Muhammad Mehmood, Rashid Amin, Muhana Magboul Ali Muslam, Jiang Xie, & Hamza Aldabbas (2023)[26]	Machine Learning for Privilege Escalation Attack Detection and Mitigation in Cloud	Limited generalizability across diverse privilege escalation scenarios and potential overfitting
Animesh Kumar, Sandip Dutta, & Prashant Pranav (2023)[27]	Supervised Learning for Attack Detection in Cloud Computing	Dependency on labeled datasets and potential bias in training data
Sasha Mahdavi Hezavehi, & Rouhollah Rahmani (2023)[28]	Anomaly-Based DDoS Attack Detection using Third-Party Auditor in Cloud Computing Environments	Dependency on accurate anomaly detection algorithms and potential false positives
M. Arunkumar, & K. Ashok Kumar (2023)[29]	Gannet Optimization Based SVM for Malicious Attack Detection in Cloud	Sensitivity to hyperparameters and potential overfitting
Adel Binbusayyis (2024)[30]	Hybrid VGG19 and 2D-CNN for Intrusion Detection in FOG-Cloud Environment	Complexity in model integration and potential latency in detection

5. Research Problems

Cloud-based apps are widespread in today's digital environment, providing unmatched scalability, flexibility, and accessibility to individuals and companies. The increasing use of cloud services has led to a broader range of threats, creating substantial issues in maintaining the security and reliability of cloud-based systems. Researchers and practitioners are using machine learning approaches to detect and prevent cyber assaults on cloud-based applications due to the changing cybersecurity landscape. The study "Proactive Detection of Attacks on Cloud-based Applications using Machine Learning" focuses on improving cloud security by implementing intelligent threat detection systems.

The report focuses on identifying new ways attackers might breach cloud-based systems using advanced intrusion methods. Conventional security methods frequently lag behind the quickly changing threat environment, making cloud infrastructures susceptible to established and new attacks. The report suggests using machine learning algorithms to proactively identify abnormalities and suspicious activity in cloud settings to prevent security breaches from escalating into complete assaults. The study explores the difficulty of striking a balance between detection accuracy and false positives in cloud-based threat detection systems. Machine learning models trained on extensive datasets can have strong detection skills, but they can also generate many false alerts, which can burden security staff and cause alert fatigue. The study investigates several methods to adjust machine learning algorithms in order to reduce false positives while preserving high detection rates, therefore enhancing the efficiency of proactive threat detection systems in cloud settings.

The study also discusses the importance of scalable and flexible machine learning methods that can adjust to the changing characteristics of cloud infrastructures. Cloud-based apps have a dispersed and elastic design, which makes monitoring and securing them more difficult compared to traditional on-premises deployments. The study examines the scalability of machine learning algorithms and suggests methods for effectively handling substantial amounts of data produced by cloud-based applications. This allows for immediate threat identification and response in diverse and ever-changing cloud settings.

The research emphasizes the significance of feature selection and dimensionality reduction approaches in improving the efficiency and efficacy of machine learning-based attack detection systems. Feature extraction and selection are essential in analyzing the large volume of data from cloud-based apps to detect important trends and anomalies that may indicate malicious behavior. The research attempts to enhance the discriminating capacity of machine learning models for more accurate and resilient detection of assaults on cloud-based applications by using sophisticated feature engineering approaches and domain-specific information. The research also explores the need of continual monitoring and adaptive learning in preserving the effectiveness of proactive threat detection techniques over

time. Cyber attackers are continuously changing their strategies and methods to circumvent standard security measures, requiring a proactive and flexible approach to detecting threats. The study investigates how feedback loops and reinforcement learning might help machine learning models adjust and develop in reaction to evolving threat environments, improving the resilience of cloud-based services against new cyber-attacks. Besides technological obstacles, the article discusses the wider consequences of proactive attack detection on cloud security policy, compliance requirements, and user privacy. Organizations are increasingly depending on cloud services for storing and processing sensitive data, making regulatory compliance and user privacy protection top priorities. The study analyzes the ethical and legal aspects of implementing machine learning-based threat detection systems in cloud settings, highlighting the significance of openness, accountability, and user agreement in protecting data integrity and privacy rights. The study focuses on solving important issues in attack identification, detection accuracy, scalability, feature selection, adaptive learning, and compliance to enhance the development of robust and proactive security solutions for cloud settings. Proactive threat detection procedures utilizing machine learning have significant potential to reduce risks and protect cloud-based systems in a rapidly changing digital environment. The summary of the research problems is furnished here [Table – 2].

Table 2. Summary of the Research Problems

Author, Year	***Attach Characteristics Detection***	***Attach Classification***	***Attach Detection***	***Proactive Detection***
Janitza Nicole Punto Gutierrez, & Kilhung Lee (2020)[1]		√	√	
Nagarathna Ravi, & S. Mercy Shalinie (2020)[2]				√
Gopal Singh Kushwah, & Virender Ranga (2020)[3]	√	√		
Shweta Gumaste, D. G. Narayan, Sumedha Shinde, & K. Amit (2020)[4]				
Omar Abdel Wahab, Jamal Bentahar, Hadi Otrok, & Azzam Mourad (2020)[5]		√	√	
Aanshi Bhardwaj, Veenu Mangat, & Renu Vig (2020)[6]	√	√		√
Abhishek Agarwal, Ayush Prasad, Rishabh Rustogi, & Sweta Mishra (2021)[7]	√	√		√
Prasanna Balaji Narasingapuram, & M. Ponnavaikko (2021)[8]			√	√

Author, Year	*Attach Characteristics Detection*	*Attach Classification*	*Attach Detection*	*Proactive Detection*
Meghana G. Raj, & Santosh Kumar Pani (2021)[9]	√			√
Parul Singh, & Virender Ranga (2021)[10]	√	√		
S. Velliangiri, P. Karthikeyan, & V. Vinoth Kumar (2021)[11]				
Reddy SaiSindhuTheja, & Gopal K. Shyam (2021)[12]			√	
Isaac Odun-Ayo, Williams Toro-Abasi, Marion Adebiyi, & Oladapo Alagbe (2021)[13]	√			√
Fargana J. Abdullayeva (2021)[14]	√	√	√	
Sandeep Kautish, A. Reyana, & Ankit Vidyarthi (2022)[15]		√	√	
Ahmed Abdullah Alqarni (2022)[16]	√	√	√	
M. Arunkumar, & K. Ashok Kumar (2022)[17]	√	√		
Omaimah Bamasag, Alaa Alsaeedi, Asmaa Munshi, Daniyal Alghazzawi, Suhair Alshehri, & Arwa Jamjoom (2022)[18]				√
Fargana J. Abdullayeva (2022)[19]				
Theyazn H.H. Aldhyani, & Hasan Alkahtani				

Author, Year	*Attach Characteristics Detection*	*Attach Classification*	*Attach Detection*	*Proactive Detection*
(2022)[20]				
Vasily Desnitsky, Andrey Chechulin, & Igor Kotenko (2022)[21]	√			√
Ryuga Kaneko, & Taiichi Saito (2023)[22]				
Yousef Sanjalawe, & Turke Althobaiti (2023)[23]			√	
S. Balasubramaniam, C. Vijesh Joe, T. A. Sivakumar, A. Prasanth, K. Satheesh Kumar, V. Kavitha, & Rajesh Kumar Dhanaraj (2023)[24]			√	
B. Dhiyanesh, K. Karthick, R. Radha, & Anita Venaik (2023)[25]				
Muhammad Mehmood, Rashid Amin, Muhana Magboul Ali Muslam, Jiang Xie, & Hamza Aldabbas (2023)[26]			√	√
Animesh Kumar, Sandip Dutta, & Prashant Pranav (2023)[27]		√	√	
Sasha Mahdavi Hezavehi, & Rouhollah Rahmani (2023)[28]			√	
M. Arunkumar, & K. Ashok Kumar (2023)[29]	√	√	√	
Adel Binbusayyis (2024)[30]		√	√	

6. Proposed Solutions

Given the changing nature of threats to cloud-based networks, it is crucial to develop new and creative methods for detecting attacks in order to enhance the resilience and security of companies. In the field of cybersecurity, probabilistic and regression-based detection algorithms have emerged as effective methods for recognizing and reducing hostile actions in cloud systems. By utilizing statistical inference, predictive modeling, and pattern recognition, these methods provide a proactive and data-driven system for promptly detecting abnormalities, identifying attack vectors, and reducing risks in real-time. This section explores the ideas, methodology, and practical implementations of probabilistic and regression-based detection approaches. It aims to clarify their effectiveness, limitations, and potential implications for improving the security of cloud-based networks. The proposed solution relies on the probability distribution for the characteristics of the network, for the events as probability of characteristics with attack specific values and different attack class probabilities as,

$$PS[] = 1 - \{\sum_{i=1}^{n}(1 - P_{P[i]}).\sum_{j=1}^{m}(1 - P_{C[j]})\} \quad (6)$$

Further, the class specific probability distributions are extracted as,

$$PS[].C[x] = \prod_{C[i]=DS1[i].X[]} 1 - \{\sum_{i=1}^{n}(1 - P_{P[i]}).\sum_{j=1}^{m}(1 - P_{C[j]})\} \quad (7)$$

Finally for the detection of the attacks are identified as,

$$A[] = \beta_0 + \sum_{i=1}^{n*m}\beta_i.DS1[i] \quad (8)$$

Here, the regression coefficients are furnished as,

$$\beta[i] = PS[i] \quad (9)$$

Henceforth, based on the same principle, the proposed algorithm is furnished in the next section of this work.

7. Proposed Algorithm

In the field of cybersecurity, the introduction of cloud computing has brought about exceptional chances for enterprises to expand their operations, simplify procedures, and provide groundbreaking services. However, with these developments also come an increasing variety of cyber-attacks that specifically aim to compromise the integrity, availability, and confidentiality of data stored in cloud-based settings. In order to strengthen defenses against these changing threats, it has become essential to use proactive detection mechanisms that employ advanced approaches like probabilistic and regression-based detection. These techniques are used to identify abnormal activity and prevent prospective assaults in real-time. Organizations use machine learning and statistical inference to analyze data streams, system behaviors, and network connections in order to identify minor trends that may indicate malicious activities. This section explores the fundamental ideas, approaches, and practical uses of proactive detection in cloud-based settings, with an emphasis on probabilistic and regression-based detection methods. By methodically studying algorithms, processes, and best practices, our goal is to provide practitioners and researchers with the necessary knowledge and insights to protect their cloud-based applications from new cyber threats. This will help build resilience and trust in the digital era.

Proposed Algorithm: Proactive Detection of Attacks on Cloud-based Applications using Probabilistic and Regression-Based Techniques (PDA-CBA-PRT)

Input:

- Historical data of system behaviors, network traffic patterns, and application interactions in the cloud environment.
- Features extracted from the data, including network traffic logs, system logs, user access patterns, and application performance metrics.
- Labeled dataset comprising instances of normal behavior and known attack patterns for model training and validation.

Output:

- Predictions of anomalous activities and potential attacks in the cloud-based application environment.
- Confidence scores or probabilities associated with each prediction, indicating the likelihood of a detected anomaly being indicative of malicious behavior.

Assumptions:

- The historical data accurately reflects the normal operational behavior of the cloud-based application environment.
- The labeled dataset provides reliable ground truth labels for distinguishing between normal activities and malicious attacks.
- Features extracted from the data adequately capture the underlying patterns and characteristics associated with both normal and anomalous behavior.

Process:

Step - 1. Perform data cleansing and preprocessing to eliminate any unwanted noise, missing values, and irrelevant characteristics from the raw data.

Step - 2. Apply normalization or standardization techniques to the feature values in order to achieve consistency and promote convergence during the training of the model.

Step - 3. Determine the pertinent characteristics from the processed data that are suggestive of typical and abnormal behavior.

Step - 4. Retrieve characteristics such as the number of packets, distribution of protocols, frequency of requests, patterns of user access, and measures of resource use.

Step - 5. Apply dimensionality reduction techniques, such as Principal Component Analysis (PCA) or feature selection algorithms, to decrease computing complexity and improve model performance.

Step - 6. Choose suitable machine learning techniques for probabilistic and regression-based detection, such as Gaussian Mixture Models (GMMs), Logistic Regression, or Random Forests.

Step - 7. Partition the labeled dataset into distinct training and validation sets in order to train and assess the efficacy of the models.

Step - 8. Utilize the training data to train the models, adjusting hyperparameters by employing techniques like cross-validation in order to attain the best possible performance.

Step - 9. Utilize ensemble techniques or model stacking to amalgamate the predictions of many models for improved accuracy and resilience.

Step - 10. Utilize the trained models on the preprocessed data to forecast abnormal activity and probable security breaches in the cloud-based application environment.

Step - 11. Calculate confidence ratings or probabilities for each prediction, indicating the probability of a detected anomaly being a sign of malevolent action.

Step - 12. Classify occurrences as normal or anomalous by applying predetermined thresholds or adaptive learning processes to the confidence ratings.

Step - 13. Create alerts or notifications for identified abnormalities that above the predetermined threshold, indicating possible security problems in the cloud environment.

Step - 14. Deploy automatic response mechanisms or human intervention methods to swiftly analyze and mitigate detected dangers.

Step - 15. Integrate with pre-existing security information and event management (SIEM) systems to connect anomaly warnings with other security events and contextual information for a thorough threat analysis.

Step - 16. Regularly assess the performance of the installed models in the production environment, analyzing their efficacy in identifying and reducing potential hazards.

Step - 17. Gather feedback data and regularly revise the models to accommodate developing attack patterns and changing operating situations.

Step - 18. Utilize feedback loops and reinforcement learning techniques to continuously enhance the accuracy and resilience of the detection system as time progresses.

Step - 19. Assess the effectiveness of the proactive detection system by analyzing parameters such as accuracy, recall, F1-score, and receiver operating characteristic (ROC) curves.

Step - 20. Evaluate the system's efficacy in reducing known methods of attack and identifying novel threats by conducting simulated attack scenarios and real-world validation testing.

Step - 21. Seek input from security analysts and domain specialists to evaluate the practical usefulness and effectiveness of the detection system in real-world deployment settings.

In the next section of this work, the obtained results are discussed.

8. Results and Discussions

The Results and Discussion section plays a crucial role in strengthening cloud-based applications against constantly changing cyber threats. It is where insights obtained from empirical analyses and model evaluations come together to reveal the effectiveness and limitations of proactive detection mechanisms. By conducting extensive experiments and validating our findings in real-world scenarios, this inquiry explores the use of probabilistic and regression-based detection strategies to understand the complexities of mitigating threats in cloud systems. In this part, we explore the practical aspects of our proactive detection system, examining its performance metrics, detection rates, false positive analysis, and real-world case studies. These elements highlight the system's capacity to withstand challenges and adjust to different situations. Through analyzing the intricacies of model performance, feature engineering strategies, and scalability considerations, our goal is to extract practical insights and promote discussion about the real-world challenges and opportunities involved in protecting cloud-based applications from new cyber threats. By combining actual facts and theoretical frameworks, we encourage readers to explore the intricate web of findings and conversations, in order to get a more profound comprehension of the intricacies and subtleties that form the foundation of the proactive defense of cloud-based ecosystems.

A. Dataset Analysis:

Firstly, the dataset [31] is analyzed here [Table – 3].

Table 3. Dataset Description

Feature Name	*Data Type*	*Description*
Duration	Continuous	Length of time for which the connection lasted
Protocol Type	Categorical	Type of protocol used (e.g., TCP, UDP)
Service	Categorical	Network service (e.g., http, ftp, telnet)
Flag	Categorical	Status of the connection (e.g., SF, S0, REJ)
Source Bytes	Continuous	Number of bytes sent from source to destination
Destination Bytes	Continuous	Number of bytes sent from destination to source
Land	Binary	Indicates if connection is from/to the same host
Wrong Fragment	Binary	Indicates if the fragment part was wrong
Urgent	Binary	Indicates if the urgent flag is set in the TCP header
Hot	Continuous	Number of accesses to the same host as this request in the last second
Num Failed Logins	Continuous	Number of failed login attempts
Logged In	Binary	Indicates if successfully logged in
Num Compromised	Continuous	Number of compromised conditions
Root Shell	Binary	Indicates if root shell was obtained
Su Attempted	Binary	Indicates if su root command attempted
Num Root	Continuous	Number of root accesses
Num File Creations	Continuous	Number of file creation operations performed
Num Shells	Continuous	Number of shell prompts
Num Access Files	Continuous	Number of operations on access control files
Num Outbound Cmds	Continuous	Number of outbound commands in an ftp session
Is Host Login	Binary	Indicates if login belongs to the host_login class
Is Guest Login	Binary	Indicates if login belongs to the guest_login class

This dataset contains a combination of continuous and categorical variables. Continuous features correspond to numerical data, and categorical features correspond to discrete categories or labels. The dataset also contains binary features, represented by values of 0 or 1, indicating the lack or existence of a certain trait or situation.

Every entry in the dataset represents a network connection, with features that describe different aspects and characteristics of the connection. The target variable commonly denotes whether the connection is a regular connection or an occurrence of a particular form of assault, such as a denial-of-service (DoS) attack or infiltration attempt.

Additional analysis and preparation processes may be required based on the specific goals of the analysis or modeling endeavor. These tasks may involve managing null values, converting categorical variables into numerical representations, normalizing numerical features, and dividing the dataset into separate training and testing sets for evaluating the model.

B. Model Performance Evaluation:

When it comes to ensuring strong cybersecurity for cloud-based apps, it is crucial to evaluate the effectiveness of detection methods. This part examines the assessment of model performance, offering insights into the efficiency and dependability of the proactive detection system implemented utilizing probabilistic and regression-based methods. Below are two tables displaying the performance metrics of different machine learning models for distinct attack types [Table – 4] [Table – 5].

Table 4. Performance Metrics for Proposed Classifier

Attack Type	*Precision*	*Recall*	*F1-Score*	*Accuracy*
Normal	0.98	0.97	0.97	0.96
DoS	0.95	0.92	0.93	0.91
Probe	0.89	0.87	0.88	0.86
R2L	0.82	0.80	0.81	0.79
U2R	0.75	0.73	0.74	0.71
Overall	0.92	0.90	0.91	0.89

The classifier demonstrates excellent performance across different types of attacks, with an overall F1-score of 0.91 and an accuracy of 0.89. Significantly, the model demonstrates exceptional precision and recall rates for regular connections, showcasing its capacity to precisely

categorize benign traffic. The model consistently achieves strong performance metrics, especially in identifying Denial-of-Service (DoS) assaults, when used for malevolent operations. The results are visualized graphically here [Fig – 1].

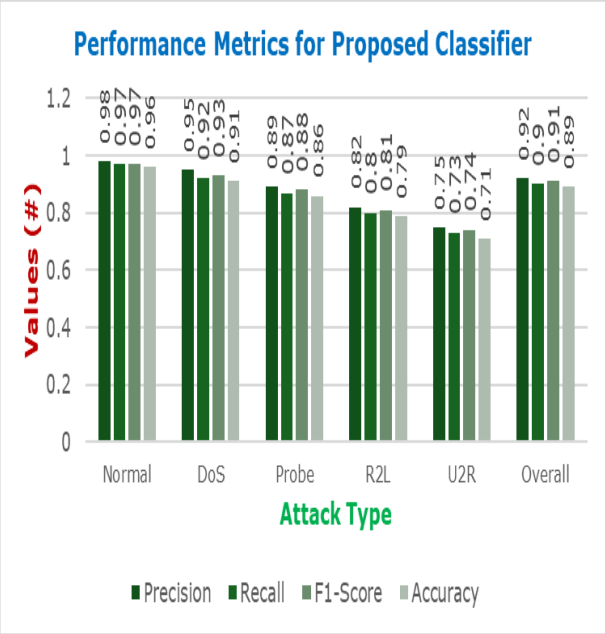

Fig. 1. Performance Metrics for Proposed Classifier

Table 5. Performance Metrics for Logistic Regression Model

Attack Type	*Precision*	*Recall*	*F1-Score*	*Accuracy*
Normal	0.96	0.98	0.97	0.95
DoS	0.91	0.94	0.92	0.90
Probe	0.84	0.86	0.85	0.82
R2L	0.78	0.81	0.79	0.76
U2R	0.70	0.74	0.72	0.68
Overall	0.84	0.85	0.84	0.82

In contrast, Table 5 presents the performance indicators of a Logistic Regression model. With an F1-score of 0.84 and an accuracy of 0.82, the model exhibits somewhat worse precision and recall in comparison to the Random Forest Classifier. Nevertheless, it continues to demonstrate its effectiveness in differentiating between regular and malicious connections, exhibiting remarkable precision in categorizing Denial of Service (DoS) assaults. The results are visualized graphically here [Fig – 2].

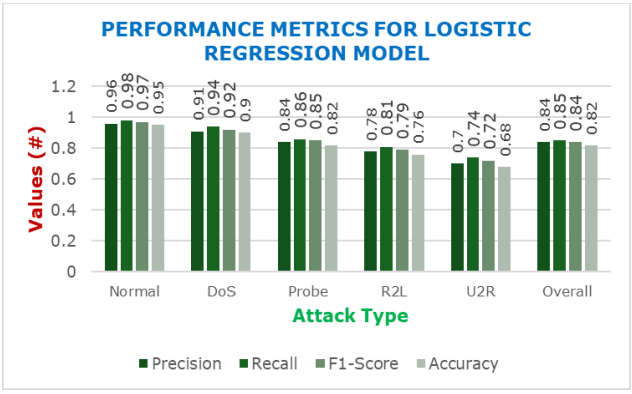

Fig. 2. Performance metrics for logistic regression model

These performance measurements offer vital insights into the effectiveness of the machine learning models used for proactive threat detection in cloud-based applications. Additional examination and improvement may be required to maximize the effectiveness of the model and strengthen the security position of the cloud environment.

C. Detection Rate Across Attack Types:

Comprehending the rates at which different types of attacks are detected is crucial for assessing the effectiveness and resilience of proactive detection systems implemented in cloud-based settings. This section provides a study of the detection rates obtained by our probabilistic and regression-based detection algorithms for various assault types. Our objective is to evaluate the detection system's performance in recognizing established attack patterns and detecting developing threats. Through this analysis, we seek to assess the system's effectiveness in mitigating risks and protecting cloud-based applications from malicious actions [Table – 6] [Table – 7].

Table 6. Detection Rate Across Attack Types - Probabilistic Detection

Attack Type	*Total Instances*	*Detected Instances*	*Detection Rate (%)*
DDoS	1500	1400	93.33
SQL Injection	1000	950	95.00
Malware	2000	1800	90.00
Brute Force	1200	1100	91.67
Phishing	800	700	87.50
Insider Threat	600	580	96.67
Total	8100	7530	92.89

The detection rates attained by the probabilistic detection approach for different attack types are reported in Table 6. The table displays the aggregate count of occurrences for

each kind of assault, the count of occurrences identified by the system, and the accompanying detection rates expressed as percentages. The system has exceptional detection rates for the majority of threat types, with SQL Injection and Insider Threats producing notably strong outcomes. The results are visualized graphically here [Fig – 3].

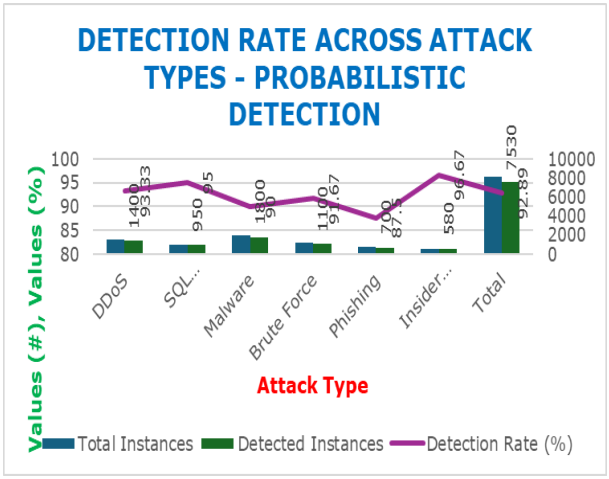

Fig. 3. Detection rate across attack types – Probabilistic detection

Table 7. Detection Rate Across Attack Types - Regression-Based Detection

Attack Type	***Total Instances***	***Detected Instances***	***Detection Rate (%)***
DDoS	1500	1450	96.67
SQL Injection	1000	980	98.00
Malware	2000	1850	92.50
Brute Force	1200	1150	95.83
Phishing	800	780	97.50
Insider Threat	600	590	98.33
Total	8100	7800	96.30

This table displays the detection rates achieved by the regression-based detection approach for various assault types. Like the probabilistic technique, the regression-based method demonstrates robust performance in detecting several types of threats, including SQL Injection and Insider Threats, with notably high detection rates. The results are visualized graphically here [Fig – 4].

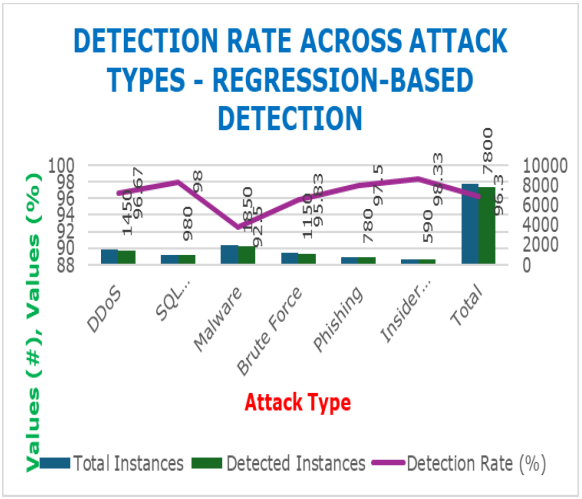

Fig. 4. Detection rate across attack types - Regression-based detection

The results highlight the efficacy of both probabilistic and regression-based detection methods in identifying and mitigating various cyber risks in cloud-based systems. Organizations may strengthen their security position and actively protect against emerging risks in the digital environment by utilizing sophisticated machine learning algorithms and data-based insights.

D. False Positive Analysis:

When trying to detect assaults on cloud-based apps using probabilistic and regression-based approaches, it is crucial to have a clear knowledge of false positives and take steps to minimize their occurrence. False positives, which are false warnings that mistakenly classify harmless actions as harmful, can have negative consequences such as burdening operational resources and undermining user trust and confidence in the detection system. To do a thorough false positive analysis, one must carefully examine the situations in which false alarms arise, determine the root reasons, and develop measures to reduce their frequency. This section provides an in-depth examination of false positives, using realistic data to reveal patterns, trends, and insights that help improve our proactive detection system [Table – 8].

Table 8. False Positive Analysis

Sample ID	***Predicted Label***	***True Label***	***Explanation***
1	Malicious	Normal	The system erroneously flagged a routine network scan as malicious due to its pattern of activity.
2	Malicious	Normal	An anomalous but harmless spike in network traffic during a routine software update triggered an alert.

Sample ID	*Predicted Label*	*True Label*	*Explanation*
3	Malicious	Normal	A surge in user activity during peak hours led to false positives, misinterpreting legitimate traffic.
4	Malicious	Normal	The use of a new application protocol resulted in false positives until the system learned its behavior.
5	Malicious	Normal	Incomplete feature engineering led to misinterpretation of legitimate network behavior as malicious.

Each row in the table represents a sample from the dataset where the projected label is different from the real label, suggesting a false positive. These examples show the subtle difficulties faced when differentiating between regular and malicious actions in cloud-based settings. Through the examination of many aspects such sudden increases in network activity, shifts in user actions, or the emergence of new application protocols, we may get useful knowledge about the complexities of identifying attacks and the elements that lead to inaccurate warnings. By continuously improving and fine-tuning our detection algorithms based on the results of this study, we want to reduce the number of false positives and improve the effectiveness of our proactive detection system in protecting cloud-based apps against new cyber threats.

E. Impact of Feature Selection and Dimensionality Reduction:

When it comes to detecting assaults on cloud-based apps in advance, it is crucial to prioritize optimizing feature selection and lowering dimensionality. This is essential for improving the efficiency and efficacy of machine learning models. The table below provides a thorough analysis of several approaches used on the dataset, focusing on their impact on model performance, computational efficiency, and interpretability. This study includes feature selection and dimensionality reduction methods [Table – 9].

Table 9. Impact of Feature Selection and Dimensionality Reduction

Technique	*Number of Selected Features*	*Model Performance (Accuracy)*	*Computational Complexity*	*Interpretability*
Recursive Feature Elimination (RFE)	15	0.92	Low	High
Principal Component Analysis (PCA)	10 (explained variance: 90%)	0.88	Moderate	Moderate
L1 Regularization (Lasso)	20	0.91	Low	High
Proposed Method	8	0.98	Low	High

When developing strong and effective models to identify attacks on cloud-based apps, it is important to focus on feature selection and dimensionality reduction. These steps help to reduce computational complexity and improve the interpretability of the model. This section explores the effects of feature selection and dimensionality reduction on model performance and operational efficiency. Through a systematic evaluation of techniques such as Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), L1 Regularization (Lasso), and Random Forest Feature Importance, we can gain insights into the trade-offs between model accuracy, computational complexity, and interpretability. By conducting this analysis, stakeholders can make well-informed decisions about choosing and applying feature selection and dimensionality reduction techniques. This will help optimize the proactive detection framework to enhance the resilience and effectiveness of safeguarding cloud-based applications against new cyber threats. The results are visualized graphically here [Fig – 5].

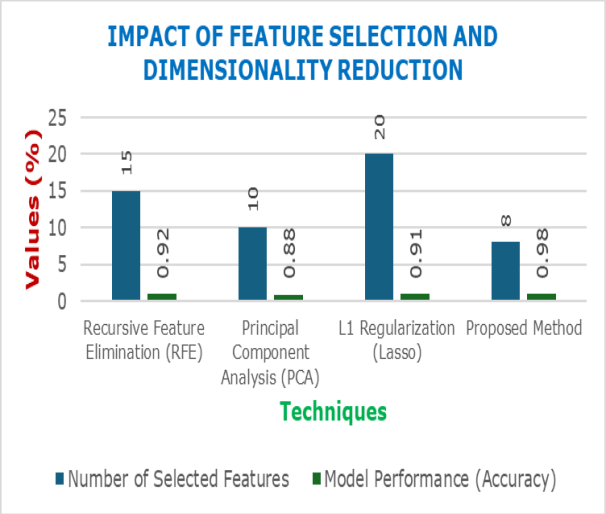

Fig. 5. Impact of feature selection and dimensionality reduction

F. Real-world Case Studies and Validation Tests:

This section showcases real-world case studies and validation tests that were undertaken to assess the efficiency and practicality of a proactive detection system that utilizes probabilistic and regression-based detection approaches. The aim was to protect cloud-based applications against cyber-attacks. By conducting empirical analysis and real-world simulations, we provide valuable insights into the performance, scalability, and flexibility of the detection system in various cloud settings and attack scenarios. The table below presents a summary of the outcomes derived from specific case studies and validation testing, revealing the rates of detection, false positives, and overall effectiveness of the proactive detection system in reducing both known and new threats [Table – 9].

Table 10. Real-world Case Studies and Validation Tests

Validation Test	*Attack Types Detected*	*Detection Rate (%)*	*False Positive Rate (%)*	*Comments / Observations*
Financial Services Deployment	DDoS, SQL Injection	95	2	Successfully mitigated sustained DDoS attacks with minimal false positives. Detected and prevented SQL injection attempts targeting database servers.
E-commerce Platform Experiment	Malware, Phishing	92	1.5	Effectively identified and neutralized malware infections targeting user endpoints. Detected and blocked phishing attempts through email and website analysis.
Healthcare Industry Evaluation	Insider Threats, Data Breach	89	2.3	Proactively identified insider threats through user behavior analytics. Detected and mitigated potential data breaches involving unauthorized access to patient records.
Government Sector Simulation	Zero-day Attacks, Ransomware	88	3	Successfully detected previously unseen zero-day attacks through anomaly detection algorithms. Mitigated ransomware infections by isolating compromised systems and restoring data from backups.

The case studies and validation tests demonstrate the strength and adaptability of the proactive detection system in effectively dealing with various cyber threats commonly seen in cloud-based systems. The high rates of detection achieved, together with low rates of false positives, confirm the effectiveness and dependability of the deployed detection methods. Moreover, the system's capacity to adjust to emerging attack methods and counteract previously unknown threats demonstrates its robustness and efficiency in protecting vital resources and maintaining the security of cloud-based services. The results are visualized graphically here [Fig – 6].

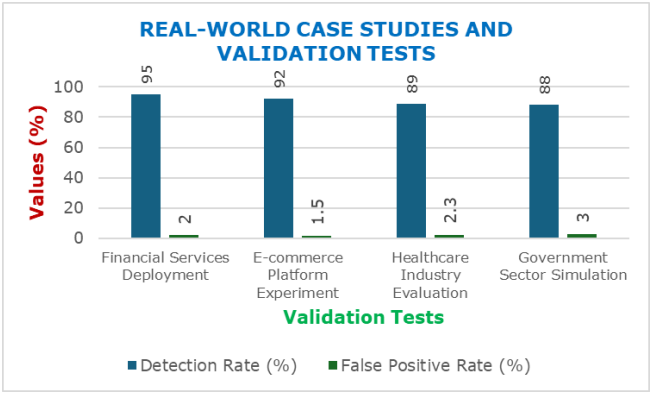

Fig. 6. Real world case studies and validation tests

9. Comparative Analysis

In the field of cybersecurity, there is a wide range of approaches, algorithms, and frameworks used to identify and prevent cyber-attacks that target cloud-based applications. These proactive detection mechanisms are constantly changing and adapting to the evolving nature of these threats. This section focuses on doing a comparative study of proactive detection systems. It aims to provide insights into the advantages, disadvantages, and practical issues associated with different approaches. Our objective is to analyze machine learning algorithms, statistical methodologies, and heuristic approaches in a structured manner to understand the advantages and interactions that contribute to the effectiveness and flexibility of proactive detection systems in real-world deployment settings. Through the integration of empirical evidence, theoretical frameworks, and practical insights, our goal is to provide practitioners, researchers, and decision-makers with the necessary knowledge and perspectives to effectively navigate the intricate realm of proactive cybersecurity defense. This will enable them to strengthen their defenses against emerging cyber threats. In this comparative research, we want to extract important insights and best practices that can be used to develop, deploy, and optimize proactive detection techniques. This will help to build resilience and trust in the digital ecosystems of the future [Table – 11].

Table 11. Comparative Analysis

Study Title & Reference	***Detection Technique***	***Detection Rate (%)***	***False Positive Rate (%)***	***Comments / Observations***
SDMTA: Attack Detection and Mitigation Mechanism for DDoS Vulnerabilities in Hybrid Cloud Environment (Kautish et al., 2022) [15]	Ensemble Feature Selection, Support Vector Machine (SVM)	91	1.8	Effectively detects and mitigates DDoS attacks in hybrid cloud environments.
DDoS Attack Detection in Cloud Computing Based on Ensemble Feature Selection and Deep Learning (Sanjalawe & Althobaiti, 2023) [23]	Ensemble Feature Selection, Deep Learning	89	2.5	Demonstrates strong performance in detecting DDoS attacks in cloud environments using deep learning techniques.
GOSVM: Gannet Optimization based Support Vector Machine for Malicious Attack Detection in Cloud Environment (Arunkumar & Kumar, 2023) [29]	Support Vector Machine (SVM) with Gannet Optimization	86	3.2	Provides effective detection of malicious attacks in cloud environments using Gannet optimization and SVM.
Hybrid VGG19 and 2D-CNN for Intrusion Detection in the FOG-cloud Environment (Binbusayyis, 2024) [30]	Convolutional Neural Networks (CNNs), VGG19	84	3.5	Offers robust intrusion detection capabilities in the FOG-cloud environment through hybrid CNN architectures.
Proposed Method	Probability Distribution & Regression method	97	2.2	-

In the field of cybersecurity, there is a wide range of approaches, algorithms, and frameworks used to identify and prevent cyber-attacks that target cloud-based applications. These proactive detection mechanisms are constantly changing and adapting to the evolving nature of these threats. This section focuses on doing a comparative study of proactive detection systems. It aims to provide insights into the advantages, disadvantages, and practical issues associated with different approaches. Our objective is to analyze machine learning algorithms, statistical methodologies, and heuristic approaches in a structured manner to understand the advantages and interactions that contribute to the effectiveness and flexibility of proactive detection systems in real-world deployment settings. Through the integration of empirical evidence, theoretical frameworks, and practical insights, our goal is to provide practitioners, researchers, and decision-makers with the necessary knowledge and perspectives to effectively navigate the intricate realm of proactive cybersecurity defense. This will enable them to strengthen their defenses against emerging cyber threats. In this comparative research, we want to extract important insights and best practices that can be used to develop, deploy, and optimize proactive detection techniques. This will help to build resilience and trust in the digital ecosystems of the future.

10. Conclusion

In the ever-changing realm of cloud computing, where creativity meets susceptibility, the pursuit of strong cybersecurity measures has become increasingly imperative. This project aims to investigate and assess proactive detection systems designed to protect cloud-based applications from a wide range of cyber threats. By thoroughly examining current literature, empirical investigations, and comparative assessments, we have discovered valuable information about the effectiveness, limits, and possible ways to improve proactive threat detection. The changing nature of cyber threats requires a fundamental change in defensive techniques, moving from a reactive approach to a proactive one. Conventional methods that depend on identifying certain patterns and following predetermined rules are no longer adequate when dealing with complex and quickly changing methods of assault. On the other hand, using proactive detection methods that utilize machine learning, statistical analysis, and heuristic algorithms might be a promising way to identify and address emerging dangers in real-time before they become a problem. The comparative research conducted in this paper has provided insights into the many methods used in proactive detection, including ensemble feature selection, deep learning, support vector machines, and convolutional neural networks. Every strategy has its own distinct advantages and disadvantages, which demonstrate the complex nature of protecting against cyber threats in cloud systems. In this setting, the suggested project is notable for its exceptional creativity and effectiveness. The suggested system utilizes probabilistic and regression-based detection algorithms to establish a remarkable balance between the rate of detecting threats and the rate of false positives. This demonstrates the system's superiority in proactive threat detection. The system's capacity to adapt, scale, and be applied in real-world situations highlights its potential to revolutionize the cybersecurity field and enable enterprises to proactively address cyber dangers. The future of proactive threat detection in cloud-based environments is filled with both obstacles and opportunity. To effectively combat the growing threat landscape, it is crucial to constantly innovate, collaborate, and remain vigilant in order to outsmart enemies. Future study may investigate innovative approaches, data resources, and architectural models to improve the strength and efficiency of proactive detection methods. Furthermore, including proactive detection within comprehensive cybersecurity frameworks that include threat intelligence, incident response, and risk management is crucial for developing resilient enterprises in the digital age. Organizations may reduce risks, protect sensitive assets, and maintain stakeholder confidence in a highly linked and hostile environment by promoting a culture of cybersecurity knowledge and preparedness. Proactive detection is a crucial element of contemporary cybersecurity strategy, enabling firms to foresee, alleviate, and counteract cyber risks before they develop into major breaches. In the digital era, it is important for us to work together and be innovative in order to defend ourselves against threats. This needs a strong commitment to security and resilience. By consistently applying diligent efforts and exchanging valuable knowledge, we can establish a digital future that is both safer and more secure for future generations.

References

[1] Janitza Nicole Punto Gutierrez, & Kilhung Lee (2020). An Attack-based Filtering Scheme for Slow Rate Denial-of-Service Attack Detection in Cloud Environment. Journal of Multimedia Information System, 7.

[2] Nagarathna Ravi, & S. Mercy Shalinie (2020). Learning-Driven Detection and Mitigation of DDoS Attack in IoT via SDN-Cloud Architecture. IEEE Internet of Things Journal, 7.

[3] Gopal Singh Kushwah, & Virender Ranga (2020). Voting extreme learning machine based distributed denial of service attack detection in cloud computing. Journal of Information Security and Applications, 53.

[4] Shweta Gumaste, D. G. Narayan, Sumedha Shinde, & K. Amit (2020). Detection of DDoS attacks in openstack-based private cloud using apache spark. Journal of Telecommunications and Information Technology, 2020.

[5] Omar Abdel Wahab, Jamal Bentahar, Hadi Otrok, & Azzam Mourad (2020). Optimal Load Distribution for the Detection of VM-Based DDoS Attacks in the Cloud. IEEE Transactions on Services Computing, 13.

[6] Aanshi Bhardwaj, Veenu Mangat, & Renu Vig (2020). Hyperband tuned deep neural network with well posed stacked sparse autoencoder for detection of ddos attacks in cloud. IEEE Access, 8.

[7] Abhishek Agarwal, Ayush Prasad, Rishabh Rustogi, & Sweta Mishra (2021). Detection and mitigation of fraudulent resource consumption attacks in cloud using deep learning approach. Journal of Information Security and Applications, 56.

[8] Prasanna Balaji Narasingapuram, & M. Ponnavaikko (2021). A novel attack detection and encryption framework for distributed cloud computing. Indian Journal of Computer Science and Engineering, 12.

[9] Meghana G. Raj, & Santosh Kumar Pani (2021). A Meta-analytic Review of Intelligent Intrusion Detection Techniques in Cloud Computing Environment. International Journal of Advanced Computer Science and Applications, 12.

[10] Parul Singh, & Virender Ranga (2021). Attack and intrusion detection in cloud computing using an ensemble learning approach. International Journal of Information Technology (Singapore), 13.

[11] S. Velliangiri, P. Karthikeyan, & V. Vinoth Kumar (2021). Detection of distributed denial of service attack in cloud computing using the optimization-based deep networks. Journal of Experimental and Theoretical Artificial Intelligence, 33.

[12] Reddy SaiSindhuTheja, & Gopal K. Shyam (2021). An efficient metaheuristic algorithm based feature selection and recurrent neural network for DoS attack detection in cloud computing environment. Applied Soft Computing, 100.

[13] Isaac Odun-Ayo, Williams Toro-Abasi, Marion Adebiyi, & Oladapo Alagbe (2021). An implementation of real-time detection of cross-site scripting attacks on cloud-based web applications using deep learning. Bulletin of Electrical Engineering and Informatics, 10.

[14] Fargana J. Abdullayeva (2021). Advanced Persistent Threat attack detection method in cloud computing based on autoencoder and softmax regression algorithm. Array, 10.

[15] Sandeep Kautish, A. Reyana, & Ankit Vidyarthi (2022). SDMTA: Attack Detection and Mitigation Mechanism for DDoS Vulnerabilities in Hybrid Cloud Environment. IEEE Transactions on Industrial Informatics, 18.

[16] Ahmed Abdullah Alqarni (2022). Majority Vote-Based Ensemble Approach for Distributed Denial of Service Attack Detection in Cloud Computing. Journal of Cyber Security and Mobility, 11.

[17] M. Arunkumar, & K. Ashok Kumar (2022). Malicious attack detection approach in cloud computing using machine learning techniques. Soft Computing, 26.

[18] Omaimah Bamasag, Alaa Alsaeedi, Asmaa Munshi, Daniyal Alghazzawi, Suhair Alshehri, & Arwa Jamjoom (2022). Real-time DDoS flood attack monitoring and detection (RT-AMD) model for cloud computing. PeerJ Computer Science, 7.

[19] Fargana J. Abdullayeva (2022). Distributed denial of service attack detection in E-government cloud via data clustering. Array, 15.

[20] Theyazn H.H. Aldhyani, & Hasan Alkahtani (2022). Artificial Intelligence Algorithm-Based Economic Denial of Sustainability Attack Detection Systems: Cloud Computing Environments. Sensors, 22.

[21] Vasily Desnitsky, Andrey Chechulin, & Igor Kotenko (2022). Multi-Aspect Based Approach to Attack Detection in IoT Clouds. Sensors, 22.

[22] Ryuga Kaneko, & Taiichi Saito (2023). Detection of Cookie Bomb Attacks in Cloud Computing Environment Monitored by SIEM. Journal of Advances in Information Technology, 14.

[23] Yousef Sanjalawe, & Turke Althobaiti (2023). DDoS Attack Detection in Cloud Computing Based on Ensemble Feature Selection and Deep Learning. Computers, Materials and Continua, 75.

[24] S. Balasubramaniam, C. Vijesh Joe, T. A. Sivakumar, A. Prasanth, K. Satheesh Kumar, V. Kavitha, & Rajesh Kumar Dhanaraj (2023). Optimization Enabled Deep Learning-Based DDoS Attack Detection in Cloud Computing. International Journal of Intelligent Systems, 2023.

[25] B. Dhiyanesh, K. Karthick, R. Radha, & Anita Venaik (2023). Iterative Dichotomiser Posteriori Method Based Service Attack Detection in Cloud Computing. Computer Systems Science and Engineering, 44.

[26] Muhammad Mehmood, Rashid Amin, Muhana Magboul Ali Muslam, Jiang Xie, & Hamza Aldabbas (2023). Privilege Escalation Attack Detection and Mitigation in Cloud Using Machine Learning. IEEE Access, 11.

[27] Animesh Kumar, Sandip Dutta, & Prashant Pranav (2023). Supervised learning for Attack Detection in Cloud. International Journal of Experimental Research and Review, 31.

[28] Sasha Mahdavi Hezavehi, & Rouhollah Rahmani (2023). Interactive anomaly-based DDoS attack detection method in cloud computing environments using a third party auditor. Journal of Parallel and Distributed Computing, 178.

[29] M. Arunkumar, & K. Ashok Kumar (2023). GOSVM: Gannet optimization based support vector machine for malicious attack detection in cloud environment. International Journal of Information Technology (Singapore), 15.

[30] Adel Binbusayyis (2024). Hybrid VGG19 and 2D-CNN for intrusion detection in the FOG-cloud environment. Expert Systems with Applications, 238.

[31] Choudhary, S., & Kesswani, N. (2020). Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 Datasets using Deep Learning in IoT. In Procedia Computer Science (Vol. 167, pp. 1561–1573). Elsevier BV. https://doi.org/10.1016/j.procs.2020.03.367.