

Cloud Based Network Threat Analysis and Risk Management Using Log Analysis and Machine Learning (Random Forest)

Mrs. R. Revathi¹, Mr. V. Dickson Iruthayaraj², Mr.S.Sarjeeth³, Mr. P. Selvakumar⁴, Mr. S. Saran⁵

^{1,2}Assistant Professor, Department of IT, V.S.B College of Engineering Technical Campus Coimbatore, Tamilnadu, India

^{3,4,5}UG Scholar, Department Of It, V.S.B College Of Engineering Technical Campus Coimbatore, Tamilnadu, India

Emails: vsbrevathi@gmail.com¹, tamilkumaran15102004@gmail.com², sarjeethsasi@gmail.com³, selvakumar6524@gmail.com⁴, actrevathicse@gmail.com⁵

Abstract:

In the era of modern technologies, introduced the widespread use of cloud computing and other solutions that revolutionized the storage and management of information. Cloud-based network threat identification and risk management using applying Log Analysis and Machine learning is intended to recognize, evaluate, and analyze cybersecurity risks in contemporary cloud settings. As cloud computing becomes more widely used, dynamic, expansive, and dispersed network infrastructures cannot be managed by conventional perimeter-based security measures. The intelligent threat detection system proposed in this research gathers and examines system and network logs produced by cloud resources in order to instantly spot malicious activity. In order to classify network behavior as either normal or abnormal, the model uses machine learning algorithms to extract pertinent data such traffic patterns, protocol usage, access frequency, and temporal behavior. In order to help security teams, prioritize incidents, risk assessment is carried out by allocating weighted scores based on threat severity, asset criticality, and previous behavior. Automated analysis, scalable log ingestion, and visual dashboards for tracking risks and threats are all supported by the architecture. By combining machine learning based random forest detection with log analysis, the suggested approach increases reaction efficiency, decreases false positives, and increases threat visibility. This project shows a realistic, affordable, and expandable solution to cloud security, which makes it appropriate for both real-world deployment scenarios and academic demonstrations. This technology has the potential to revolutionize our approach to cybersecurity and system security.

Keywords: Cloud based network, Threat Analysis, Machine Learning, Cloud Data Security, Risk Management, Random Forest, Log Analysis

1. Introduction

Cloud computing is a massive information technology (IT) revolution that makes it possible for end users to access virtualized and scalable sources with no infrastructure upkeep and costs [1]. They also offer a great deal of flexibility, and the materials are managed by different management associations and made available online using well-known networking protocols, standards, and formats. Vulnerabilities and defects in legacy protocols and underlying technology allow network attackers to get access determined [2]. Improving cloud computing security is an important endeavor makes the advantage of cutting- edge techniques and tools to detect and treat malicious activity. In the constantly

changing field of cybersecurity, proactive defense tactics depend heavily on early threat identification and risk assessment [3]. Consequently, in contemporary cloud-based network infrastructures, this integrated methodology provides a scalable, robust, and analytically rigorous solution for strategic risk management and preemptive threat intelligence into feature relevance [4]. The goal of this study is to promote more investigation into Random Forest's anomaly detection capabilities in the cloud computing space. Log Analysis: Log analysis is the practice of examining computer-generated log events to get knowledge about the functionality and state of IT applications and environments. By examining logs, businesses may understand online user activity,

adhere to security rules, audits, and regulations, and proactively and reactively reduce risks. Using log analysis, you may identify problems before or as they arise and prevent costly delays, wasted time, and other expenses [5]. The steps in the cloud security risk management process are outlined below and are depicted in Figure 1. Random Forest: A machine learning system called Random Forest, which was trademarked by Adele Cutler and Leo Breiman, aggregates the results of multiple decision trees combined to yield a unique outcome. It has become more and more popular for resolving problems with both regression and classification due to its adaptability and simplicity of use. The random forest approach consists of a group of decision trees, A bootstrap sample drawn from a training set, makes up each of them with replacement [6]. A machine learning-based security technique that uses log analysis and Random Forest to automatically identify, categorize, and reduce cyberthreats in cloud environments is cloud-based network threat analysis and risk management [7]. It examines vast amounts of network and system traffic data using the Random Forest algorithm, an ensemble learning technique which creates several decision trees to increase precision and decrease overfitting. Detecting intricate, unidentified attacks and lowering false positives are two areas where this technique excels shown in Figure 1.

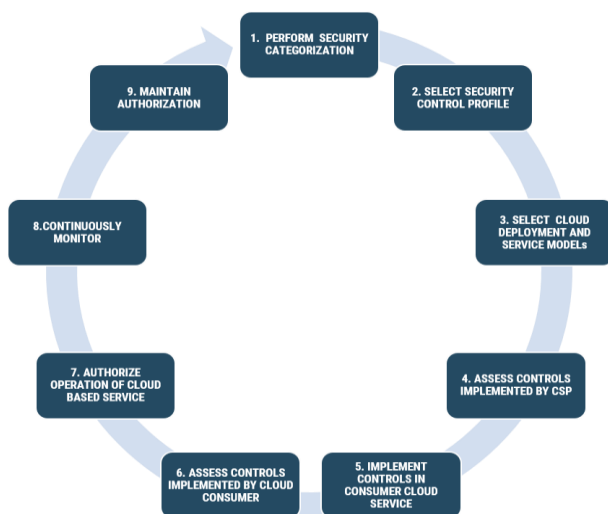


Figure 1 Cloud Security Risk Management Process

Objectives

- To gather network log data generated by the cloud and prepare it for organized analysis. And identifying important signs of compromise by performing feature extraction and selection from log collections.
- To improve total cloud security management by lowering false positives and increasing threat awareness and response efficiency through the integration of automated analysis, scalable log ingestion, and visualization dashboards.
- To evaluate the suggested Random Forest-based machine learning model against conventional detection systems that rely on rules.
- To increase the overall security of cloud data by using automated risk management and sophisticated threat monitoring.
- By facilitating intelligent correlation of events and adaptive reaction techniques, it also enhances forensic analysis, compliance auditing, and proactive risk management. As a result, it is a highly effective and research-relevant solution for contemporary cloud security systems.

The summary of this research is arranged in following order. The broad introduction of the research is explained in Section 1. The limitations of the current approaches for threat analysis and risk management are described in Section 2. The suggested system architecture and operation of the cloud-based network threat analysis and risk management using Log analysis and machine learning method are presented in Section 3. The experimental results for the suggested system using machine learning with random forest are contrasted with those of current systems in Section 4. In Section 5, the conclusion of the research findings is discussed.

2. Literature Survey

The characteristics of different approaches and the networking needs are discussed in this segment. A compilation of pertinent research of threat analysis and risk management is given prior to using this assessment to highlight research gaps and elucidate individual research goals. A Kaggle dataset from the Australian Centre for Cyber Security Cyber Range

Lab that included raw network packets from the UNSW-NB 15 dataset made with the IXIA correct apparatus. An accuracy and false rate of algorithms were compared to those of other ML approaches. In the context of the k-cross validation approach, experimental data demonstrate the superiority of Random Forest over alternate algorithms. A strong machine learning algorithm that can improve cloud computing security by efficiently identifying unusual activity is Random Forest [8]. The danger of attacks concerning the dangers and assaults for cloud computing systems, intrusion detection is done using the most recent dataset (ISOT Cloud Intrusion Dataset) [9]. SVM, random forest, logistic regression, Naïve Bayes, ANN, and KNN are some of the supervised ML algorithms that are used in this methodology to identify and classify disruptions in the cloud environments. Consequently, 98.4% accuracy was assessed for the SVM model. Measures such as accuracy, AUC, F1, precision, and recall are utilized to contrast and examine the evaluation metrics of various ML methods. The outcomes are shown as confusion matrices. The results of this work will also assist the network security manager in reducing instantaneous threats in cloud computing settings. Cloud system security and performance by combining a state-of-the-art Virtual Machine scheduling technique with potent anomaly detection for Kubernetes pods [10]. Scheduling virtual machines in typical cloud model based on short-term need for resources and neglecting long-term and overall used leads to computational inefficiencies. The AdaBoost and Random forest methods utilized in our novel Virtual Machine scheduling solution to reduce the impact on systems that have been installed, limit the number of real computers, and increase actual CPU use of virtual machines simultaneously. For better scheduling decisions, it also uses past resource monitoring data. The Kubernetes pod identification of anomalies framework works in tandem with AdaBoost and Random Forest to identify anomalous user behavior and any crypto mining threats, ensuring a comprehensive security strategy. When occurrences are deemed irregular, they are ultimately categorized into several attack types utilizing DL models including CNN and LSTM. At this point, an unbalanced dataset is resolved via

adaptive synthetic sampling, or ADASYN. According to the experimental findings, this model has higher TPR for the majority of attack events, as well as a faster rate of data preprocessing and possibly a shorter training period. Specifically, multi-target classification accuracy can achieve up to 86% in NSL-KDD dataset and 98% in CIC-IDS2017 dataset at this time [11]. A company's financial risks, ML methods including SVM, RF, ANN, Gaussian Processes, and Adaptive Learning have been applied to Big Data analysis. Using the collateral as an independent variable, credit rating is investigated in relation to data processing. The relationship between credit rating and data processing is examined using collateral as an independent variable. While SVM has the advantage of higher classification accuracy than RF, this study demonstrates that RF technique has advantages over SVM algorithm in terms of speed and operational simplicity. Tests and algorithms are used to validate the predicted model during the execution of the intelligent systems [12]. A comprehensive dataset analysis to describe the respondents' behavior according to three different characteristics: textual spreading, time gap, and time span [13]. Based on machine learning, they put forward two distinct methods: singleton and dual. While the latter employs two independent Random Forest classifiers, one to detect sad participants and another to identify non-depressed individuals, the former uses a single RF classifier with two threshold functions. Similarities in writing, semantics, and text are used to define features in both situations. Time-aware methodology was used in the evaluation, which penalizes late detections and promotes early detections. These findings indicate that a dual approach surpasses the singleton model and can enhance the cutting-edge detection methods by 10%. The issue of trustworthy resource allocation in edge-cloud hybrid settings and examines tools, processes, and techniques that can be utilized to improve the reliability of dispersed systems in a variety of diverse network settings. The poll was structured according to a breakdown of the issue of reliable resource providing into three kinds of methods: application flexibility and remediation, placement of components and system integration; and workload characterization and prediction. In addition to the

survey results, a focused on solving problems assessment of the cutting edge is offered. The essay ends with a synopsis of the issues that have been found and a plan for future study directions [14]. Distributed Denial of Service attacks was among the greatest common ones that compromise cloud performance and cause significant harm. This study uses a new approach called the gradient hybrid leader optimization model to detect DDoS attacks in an efficient manner. Training a Deep Stacked Autoencoder successfully recognizes the attack was the responsibility of this improved technique. The researchers conducted the Deep Maxout Network with the data was augmented using the oversampling process, and features were fused using the overlap coefficient [15]. Additionally, gradient descent and the hybrid leader-based optimization (HLBO) technique are combined to create the suggested GHLBO. A number of performance indicators, including testing accuracy of 90%, true positive rate of 0.907, and true negative rate of 0.917 are used to evaluate and obtained in this approach. To create the dataset, the suggested algorithm for pattern recognition was then fed the behavioral characteristics data. In addition to generating the frequency distribution of each system call, the algorithm also generates a hash value for the list of file names and a hash value for the list of process names based on the SHA256 technique [16]. Lastly, 10-fold cross validation and machine learning techniques are used to assess the created dataset. Comparing the J48 tree classification algorithm to other ML algorithms, it was discovered that it performed well with high detection accuracy. The detection accuracy for a dataset of 225 instances is 98%. The detection accuracy rose with the number of cases in the dataset, reaching a maximum of 99.5 for a dataset with 265 instances. A hybrid approach to enhance threat intelligence that combines ML driven examination of behaviour with anomaly detection [17]. In comparison to more traditional methods like SVM, random forests, and factors of local outlier., the iForest algorithm was evaluated and shows improved performance metrics, including higher detection accuracy and lower rates of false positives and false negatives. Specifically, algorithm of iForest successfully identified anomalies with 0.842 as an

average accuracy and 0.06 as a false rate. Additionally, the algorithm used little network bandwidth and had efficient processing speeds, ranging from 10 to 150 ms. The findings demonstrate that the iForest algorithm improves threat detection accuracy and effortlessly integrates with existing security frameworks, providing a scalable solution for identification of threat in real-time in environments of cloud. The ideal characteristics are used to train a random forest classifier. The idea of a hybrid technique to solve the issue of imbalanced data. Using an ADASYN approach, the minority classes are oversampled, and the majority class is randomly undersampled as necessary. By reducing the false positive rate and increasing the positive rate of truth, this integrated approach has a major impact on every category and raises system activity as a whole. CIC Bell DNS EXF 2021, CIC-DDoS2019, and UNSW-NB15 were the three datasets used to assess the suggested methodology. For these datasets, documented accuracies were 92%, 99% and 98%, in that order. Both classification of multi-class and datasets with individual class results were excellently performed by the IDS based on hybrid feature selection [18]. To overcome these drawbacks in the aforementioned traditional methods, a machine learning algorithm for cloud-based network threat analysis and risk management using log analysis developed and it will have explained in the upcoming section.

3. Proposed System Design

The system design starts with centralized log collection, which involves ingesting server, network, firewall, and cloud activity logs into a scalable cloud processing and storage environment. In order to clean logs and extract pertinent properties including traffic patterns, protocol types, access frequency, and time-based behavior, data preparation and feature extraction are then carried out. Following training, a Random Forest machine learning model is used to categorize network activity as either benign or malevolent. A risk assessment module then assigns weighted scores according to asset criticality and threat intensity after discovery. In order to facilitate effective incident response and ongoing security enhancement, the system also incorporates automated alarms and real-time monitoring

dashboards. The figure 2 explains about the suggested block diagram of the cloud network-based threat analysis and risk management using log analysis and ML based Random Forest (RF) method.

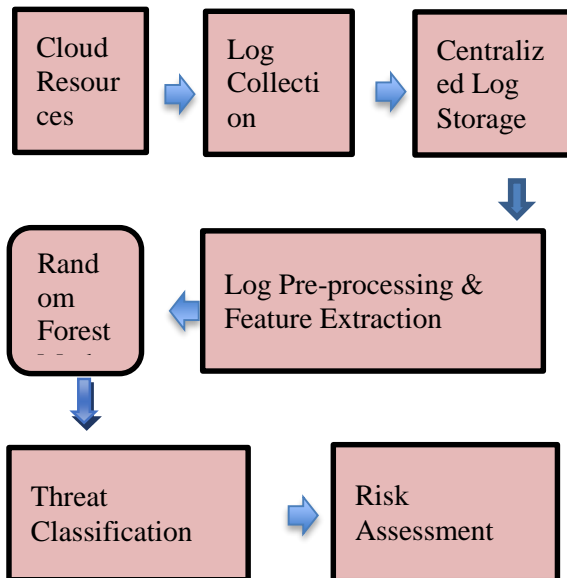


Figure 2 Proposed Block Diagram for Threat Analysis And Risk Management

3.1. Log Analysis

Log analysis is the process of gathering, interpreting, and processing log files in order to transform the information into knowledge that may be used to discover system performance problems and security threats [19]. The essential log ideas for security analysts and investigators are covered in this subject. The goal of threat analysis and log-based risk management is to find security incidents by looking at system, network, and application logs. For monitoring purposes, logs produced by servers, firewalls, cloud platforms, and user activity are gathered and organized. Early detection of possible threats is possible through the analysis of trends like anomalous login attempts, odd traffic spikes, unauthorized access, or questionable protocol usage. Using correlation techniques, one can identify complicated attacks by connecting linked occurrences. By evaluating the impact, likelihood, and seriousness of identified threats, risk management is carried out. Security teams can

prioritize responses, address vulnerabilities, and improve system security overall based on this rating. Random Forest (RF) Algorithm A powerful machine learning method for situations involving both regression and classification is the Random Forest method. It is a component of ensemble learning approaches, which integrate many models to produce more accurate and dependable predictions. Random Forests are widely used due to their versatility, effectiveness, and ease of usage in a range of settings. It considered its scalability and versatility, the Random Forest algorithm is a viable approach to enhancing risk management and data security.

To produce c classifiers:

for $i=1$ to c **do**

The training data should be sampled at random. D with suitable to generate D_i

Generate a root node N , comprising D_i

Call BuildTree (N_i)

end for

BuildTree (N):

if N includes examples of single class **then**

return

else

Choose arbitrarily $x\%$ of the potential splitting characteristics in N

To divide on, pick feature F with the greatest knowledge gain.

Generate f child nodes of N, N_1, \dots, N_f where F has f potential principles (F_1, \dots, F_f)

for $i=1$ to do

Configure the substances of N_i to D_i , where D_i is all examples in N that match F_i

Ask BuildTree (N_i)

end for

end if

Figure 3 Pseudocode for Random Forest method

The probability for every class C in a categorization problem, indicated by the symbol $p(c|L)$. In a regression task, the distribution to be estimated is over the continuous parameter $x \in \mathbb{R}^H$. The two goals

must be met by training random forests for threat identification to recognize and categorize threat-related patches and to utilize the object's scale and placement by regressing those patches. Figure 3 makes the random forest algorithm's pseudo code visible. Indeterminate Regression Frequently, forests show a non-linear mapping $M: RM \rightarrow RK$, which maps a goal forecast y to an example x . Every binary decision tree in the ensemble is trained using a portion of the training data, $\{T_t\}_{t=1}^T$. For this mapping to be obtained, T must be the number of trees. More precisely, the separating roles $\phi(x)$ partition the data into two different subsets, L and R , from which each node in a tree takes a random sample. Prior to evaluating each splitting function, we measure the information gain. The separating operation $\phi^*(x)$ that yields the biggest information gain is used for the L and R subsets. A density method $p(y)$ is estimated for all samples that lie inside this region, and if any of these criteria are satisfied, the density model is used as the region node target. The most straightforward technique for calculating the probability distribution $p(y)$ is to take target and obtain. However, some alternatives are much more complicated, including employing a Gaussian kernel density estimate or nonparametric densities.

4. Implementation Result

The experimental evaluation of the newly proposed system using Log Analysis and ML based Random Forest (LA-RF) algorithm for cloud network based threat analysis and risk management is analyzed and compared with the traditional approaches such as SVM, LSTM, and ANN. The performance evaluation parameters such as accuracy, precision, and recall are used for comparison analysis of the suggested model's findings. The dataset for this research is collected from the online availed repositories, and these obtained dataset is differentiated as 80:20 for training and test dataset.

Table 1 Performance outcome

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
SVM	85.65	86.05	83.42	82.56

LSTM	87.18	84.38	86.37	85.48
ANN	92.36	89.63	88.12	90.15
LA-RF	97.92	92.74	91.26	93.86

For performance comparison, Table 1 shows the overall output results of the suggested Log Analysis and ML based Random Forest method with the current algorithms. The proposed ML-based RF model achieves 97.92% accuracy, 92.74% precision, 91.26% recall, and 93.86% f1 score output based on these findings. Figure 5 presents the result bar chart for both the suggested and current methods quite evident.

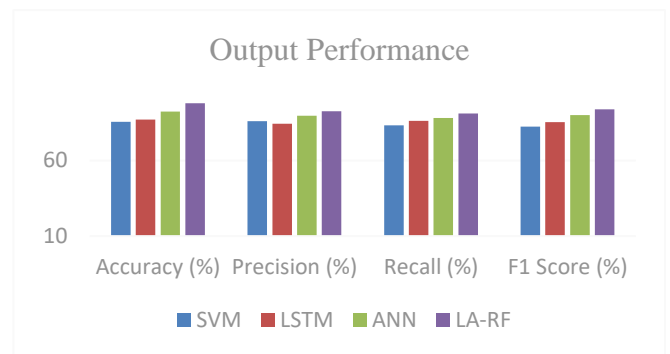


Figure 4 Performance output graph

According to this performance comparison, the suggested Log Analysis and Machine Learning based Random Forest (LA-RF) algorithm for cloud network based threat analysis and risk management performs better than other current algorithms such as SVM, LSTM, ANN shown in Figure 4.

Conclusion

In summary, the suggested system for identifying and managing cloud-based network threats offers a practical way to address the increasing security issues in contemporary cloud settings. A Random Forest-based machine learning model and log analysis are combined in the system to accurately detect anomalous network activity and possible cyber threats. Scalable cloud architecture and automated log ingestion guarantee effective handling of data and real-time observation. Making wise choices and event prioritization are supported by risk assessment

that takes into account historical trends, asset importance, and severity. This method improves danger visibility overall while lowering false positives. Additionally, the framework is scalable, affordable, and flexible enough to accommodate changing infrastructures. Overall, the study shows a clever and useful strategy for enhancing proactive risk management and cloud cybersecurity using ML based Random Forest technique.

References

- [1]. Soni, R., Bhatia, K., & Rajput, N. (2025). A thorough analysis of cloud computing technology: Present, past, and future. In *Recent Advances in Sciences, Engineering, Information Technology & Management* (pp. 137-145). CRC Press.
- [2]. Hamdi, A., Fourati, L., & Ayed, S. (2024). Vulnerabilities and attacks assessments in blockchain 1.0, 2.0 and 3.0: tools, analysis and countermeasures. *International Journal of Information Security*, 23(2), 713-757.
- [3]. Walia, G. K., Kumar, M., & Gill, S. S. (2023). AI-empowered fog/edge resource management for IoT applications: A comprehensive review, research challenges, and future perspectives. *IEEE Communications Surveys & Tutorials*, 26(1), 619-669.
- [4]. Rehman, F., & Hashmi, S. (2023). Enhancing cloud security: A comprehensive framework for real-time detection analysis and cyber threat intelligence sharing. *Advances in Science, Technology and Engineering Systems Journal*, 8(6), 107-119.
- [5]. Paramesha, M., Rane, N., & Rane, J. (2024). Enhancing resilience through generative artificial intelligence such as ChatGPT. Available at SSRN 4832533.
- [6]. Thomas, N. S., & Kaliraj, S. (2024). An improved and optimized random forest-based approach to predict the software faults. *SN Computer Science*, 5(5), 530.
- [7]. Sundaramurthy, S. K., Ravichandran, N., Inaganti, A. C., & Muppalaneni, R. (2025). AI-driven threat detection: leveraging machine learning for real-time cybersecurity in cloud environments. *Artificial Intelligence and Machine Learning Review*, 6(1), 23-43.
- [8]. Gupta, A., & Simon, R. (2024, March). Enhancing security in cloud computing with anomaly detection using random forest. In *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)* (pp. 1-6). IEEE.
- [9]. Sharma, A., & Singh, U. K. (2025). Cloud computing security assurance modelling through risk analysis using machine learning. *International Journal of System Assurance Engineering and Management*, 16(3), 1287-1300.
- [10]. Abarnaa, M., Anandhi, S., Anaswara, J. S., & Shanthini, J. (2025, February). Enhancing cloud container security using random forest and AdaBoost algorithm. In *AIP Conference Proceedings* (Vol. 3204, No. 1, p. 050006). AIP Publishing LLC.
- [11]. Mustafa, A. A., Hussein, H. M., Kadhim, M. M., & Hussein, M. J. (2025). A Hybrid Oversampling Approach for Fraud Detection: Integrating SMOTE-ENN and ADASYN. *International Journal of Safety and Security Engineering*, 15(06), 1243-1250.
- [12]. Wilson, A., & Anwar, M. R. (2024). The future of adaptive machine learning algorithms in high-dimensional data processing. *International Transactions on Artificial Intelligence*, 3(1), 97-107.
- [13]. Nacef, A., Bahroun, S., Khalfallah, A., & Ahmed, S. B. (2023). Features and Supervised Machine Learning Based Method for Singleton Design Pattern Variants Detection. In *ENASE* (pp. 226-237).
- [14]. Merseedi, K. J., & Zeebaree, S. R. (2024). The cloud architectures for distributed multi-cloud computing: a review of hybrid and federated cloud environment. *The Indonesian Journal of Computer Science*, 13(2).
- [15]. Balasubramaniam, S., Vijesh Joe, C., Sivakumar, T. A., Prasanth, A., Satheesh Kumar, K., Kavitha, V., & Dhanaraj, R. K. (2023). Optimization enabled deep learning-based ddos attack detection in cloud computing. *International Journal of Intelligent Systems*, 2023(1), 2039217.

- [16]. Malik, J., Akhunzada, A., Al-Shamayleh, A. S., Zeadally, S., & Almogren, A. (2025). Hybrid deep learning-based threat intelligence framework for Industrial IoT systems. *Journal of Industrial Information Integration*, 45, 100846.
- [17]. Mamidala, V., Yallamelli, A. R. G., Devarajan, M. V., Ganesan, T., Yalla, R. K. M. K., & Sambas, A. (2025). Hybrid anomaly-based cloud security and machine learning-based behavioral analysis for threat intelligence detection in cloud environment. *Cluster Computing*, 28(8), 485.
- [18]. Bakro, M., Kumar, R. R., Husain, M., Ashraf, Z., Ali, A., Yaqoob, S. I., ... & Parveen, N. (2024). Building a cloud-IDS by hybrid bio-inspired feature selection algorithms along with random forest model. *Ieee Access*, 12, 8846-8874.
- [19]. Patil, Y., Solpau, S. S., Umare, S., & Bhosale, S. (2025, May). Log Insight: A Tool for Log Analysis and Threat Detection. In *2025 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC)* (pp. 1-6). IEEE.
- [20]. Jamil, M., & Creutzburg, R. (2025, March). Enhancing cybersecurity in critical infrastructure: utilizing random forest ai model for threat detection. In *Future of Information and Communication Conference* (pp. 388-398). Cham: Springer Nature Switzerland.