Grenze International Journal of Engineering and Technology, June Issue

Zero-Day attacks Detection in Cloud using a Hybrid Machine learning Classifier

Rakesh Raushan[1], Prashant Awasthi[2], Himanshu Tiwari[3] and Alok Kumar[4]
[1-2]Greater Noida Institute of Technology/Department of Computer Science & Engineering, Gautam Buddha Nagar, India
Email: rakeshraushan129@gmail.com, prashantawasthi8858@gmail.com
[3-4]Lloyd Institute of Engineering & Technology/Department of Computer Science & Engineering, Gautam Buddha Nagar, India
Email: him.tiwarics57@gmail.com, alokcs9@gmail.com

***Abstract*— Cloud computing has become an essential part of the modern IT ecosystem in present time due to its flexibility, openness, accessibility and availability for all. And because of this feature has made it vulnerable to several security threats, including Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks. Signature and rule-based Intrusion detection system were not able to detect the new and unknown attacks that could be launched against a cloud environment. This research uses machine learning algorithms such as support vector machines (SVM), K-nearest neighbors (KNN), and C4.5 Decision Trees to identify how well specific machine learning algorithms can detect and mitigate security threats. In this study the NSL-KDD Dataset was used. After evaluating the performance of individual classifiers, we ensemble Support Vector Machine, K-nearest neighbors (KNN) and C4.5 Decision Tree. This Hybrid approach has been trained and tested with the above dataset after feature selection SelectKBest and ANOVA F-Test, achieving high accuracy, higher Precision, recall, and F1-score than the individual classifier. The model also shows a low false-positive rate and false-negative rate. This shows that combining the strength of C4.5 with the classification power of SVM and KNN yields a more reliable and generalizable intrusion detection mechanism.**

***Index Terms*— Intrusion Detection System, Distributive Denial of Services (DDOS), Support Vector Machine, C4.5 Decision Tree, K-Nearest Neighbours, SelectKBest, ANOVA F-Test.**

I. INTRODUCTION

Cloud computing revolutionized the entire IT industry by delivering computing resources in an on-demand manner that is scalable, elastic and cost-efficient. Cloud computing is an online platform which offers service by the internet to accommodate various applications such as IaaS, PaaS, and SaaS [1]. By enabling these deployments to be virtualized, a plurality of virtual machines may be instantiated to run on a single physical server, increasing resource utilization and operational efficiency [6]. The advancements brought by this technology have led to widespread use in government, healthcare, education, and business sectors [8].
However, this cloud-computing architecture does have a large attack surface due to its ability for many customers to use the infrastructure at once (multi-tenant), its ability for remote access to all resources, and its dynamic resource allocation capabilities.[20]. As more cloud computing is utilised and, ultimately, integrated with the IoT, huge amounts of data will be transmitted between all parties and create increased security risks for devices/companies/users. A DDoS Attack is one of the most destructive threats [9][8].

Grenze ID: 01.GIJET.12.2.108_20

The various forms of attacks involve overwhelming an organization's network services with a flood of unwanted incoming traffic from multiple compromised computers, thereby rendering the services inaccessible to legitimate users. The complexity associated with the increasing number of attack vectors indicate that some traditional network protection measures such as firewalls and signature-based Intrusion Detection Systems (IDS) are inadequate alone to fully protect the enterprise from these types of attacks [5].

Researchers have been exploring new ML techniques in the last several years as potential solutions for enhancing the efficacy of intrusion detection systems within cloud-based environments (clouds). A main advantage of using ML models is their ability to recognize anomalies in network activity patterns based on an examination of historical data from previous network activities [2][3]. Though there are many machine learning algorithms available, many algorithms do not perform well on different types of attack signatures and rapidly changing network conditions. To address this issue, multiple hybrid models have been proposed which use a combination of different algorithms to improve the ability to accurately detect alarms and reduce the number of false alarms issued [7][8].

This article presents a new hybrid approach to detecting Cybersecurity threats in Cloud environments using an Intrusion Detection System (IDS). Based on the C4 model, the hybrid model utilizes a combination of three different machine learning algorithms (Decision Trees, K-Nearest Neighbors, and Support Vector Machines) to capitalize on the strengths of each of these methods [5][6]. The NSL-KDD dataset, with its good mixture of normal and attack traffic, is used for this analysis. To ensure valid results, we implemented a cross-validation procedure up to 10 folds in order to better account for evaluation bias [13]. To increase the effectiveness of the classification, the C4.5 integrated and SVM hybrid models were used to increase classification accuracy through the combination of rule-based decision-making and margin-based learning.

This approach allows for accurate monitoring and diagnosing of malicious network traffic while providing a low number of false positives and false negatives [19][20]. Our goal is to build a Comprehensive, Scalable and Adaptable Intrusion Detection System (IDS) for Cloud Environments, capable of continuously adapting to evolving threats.

II. Literature Review

The research being reviewed shows that the use of machine learning to improve computer security has steadily increased. The novel approach will be based on multiple machine learning techniques and uses feature selection methods such as ANOVA F-tests and Select K Best by Hari Krishna et al [3].

Olasehinde investigated how ML can be adapted through the use of auto-encoders, reinforcement learning, and unsupervised learning to identify zero-day threats, highlighting that real-time learning is critical due to the constantly changing nature of clouds [5]. Similarly, Dorothy and colleagues developed an AI-generated IDS that included both Random Forest and Isolation Forest algorithms in order to have a highly accurate and recall rate by using a hybrid supervised/unsupervised approach [4].

TABLE I: Comparison of Different Machine Learning Techniques used, their accuracy, Key contribution and Limitation

Author(s)	ML Techniques Used	Accuracy / Results	Key Contributions / Limitations
Hari Krishna et al.	Ensemble SVM (Bagging & Boosting), ANOVA, SelectKBest	High accuracy, robust detection	Effective feature selection; reduced false positives; focused on cloud-specific IDS design
Olasehinde et al.	Autoencoders, Reinforcement Learning (RL), Unsupervised ML	Outperformed static models	Adaptive ML for zero-day threat detection; strong in real-time learning
Dorothy et al.	Random Forest, Isolation Forest	95% accuracy, 96% recall	AI-driven hybrid system with real-time response; handles unknown threats effectively
Attou et al.	Random Forest, Feature Engineering	99.99% (NSL-KDD), 98.3% (BoT-IoT)	Emphasis on preprocessing; excellent results with with tree-based models

They examined how to enhance Random Forest classifiers via Feature Engineering using NSL-KDD and BoT-IoT datasets, producing Detection Accuracies of up to 99.99%. In their findings, they presented Pre-processing Techniques of Feature Reduction & Normalization that work to increase the effectiveness of Random Forest Classifiers [1]. Mahendar and Chatrapati suggested a hybrid solution combining the dimensionality reduction approach with a dynamic rule-based engine for detecting complex attack types such as Distributed Denial of Service (DDoS) [11][14]. Additionally, Random Forest (RF) was applied to the hybrid solution to achieve an accuracy close to 99% [2]. Numerous authors have developed models combining existing algorithms together (e.g., Decision Trees aggregating with Support Vector Machines (SVM) or K-Nearest Neighbours (K-NN)),

allowing for improved interpretability/performance trade-offs. The NSL-KDD data set remains a widely-used benchmark for these types of studies due to its appropriately structured dataset compared to the original KDD'99 dataset [6][8]. Hybrid Machine Learning demonstrates significant potential as solutions for cloud-based intrusion detection systems. However, there are still challenges present, Including False Positive Rates, scalability problems and difficulty adapting to unknown attacks [8][10][12]

III. METHODOLOGY

The proposed intrusion detection model is designed to identification and spotting cyber security threats and other network-based attacks within cloud environments using ensembling support vector machine and C4.5 Decision tree of machine learning classifier. The methodology comprises several sequential phases as shown in fig.1.

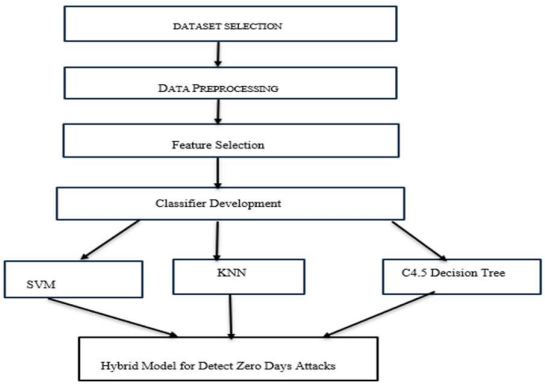

Fig 1: Phases of Methodology

A. Dataset Selection

The NSL-KDD dataset [2] is chosen for the study by refining the limitation of redundant records and class imbalances. It consists of 41 features and a class label identifying the type of normal and attack network connection. Attacks are categorized as: Distributed Denial of Services, User to Root(U2R), Remote to Local(R2L), and Probe.

Although NSL-KDD dataset is relatively old, but it is still widely used as a benchmark dataset for evaluating intrusion detection systems because it removes redundant records present in the original KDD'99 dataset and provides balanced training and testing samples. Therefore, it allows fair comparison with previous studies.

B. Data Preprocessing

Data preprocessing is used for the transformation of raw data into a suitable format for machine learning models, which includes Label Encoding, Normalization, and Noise Removal. Label Encoding converts the Categorical features into numerical values, Normalisation normalises the Continuous features to ensure that the all inputs come under the range [0,1] using Min-Max Scaling. Similarly, Noise Removal eliminates the duplicate and incomplete records for improving training data quality.

C. Feature Selection

Feature Selection is used to enhance model accuracy and reduce computational cost. It uses ANOVA F-test to determine statistical relationships between features and the target variable and SelectKBest that selects the top k features based on the ANOVA F-value.

D. Classifier Development

Various Machine Learning classifiers are selected according to their strength, and each classifier is trained on selected features. Also, each feature is cross-validated up to 10-fold to ensure that the model is reliable, generalises well and avoids overfitting.

- K-Nearest Neighbours distance-based classifier that adapts well to non-linear attack boundaries.
- Support Vector Machines (SVMs) are suitable for high-dimensional data and are capable of handling overlapping classes.
- C4.5 Decision Tree performs well on categorical data and offers transparency in decision rules.

E. Experimental Setup

The dataset was divided into training and testing sets using an 80:20 split. To ensure robustness, 10-fold cross-validation was applied during model training. The SVM classifier was implemented with an RBF kernel, KNN used k = 5 neighbors, and the C4.5 decision tree was constructed using information gain as the splitting criterion.

F. Hybrid Model Construction

A hybrid ensemble approach is proposed to avoid the limitations of Individual models. The SVM and C4.5 Decision Tree are integrated in a parallel voting system. Each classifier predicts the outputs, and the final decision is made through the majority voting system. It combines the interpretability of Decision Trees with robust margin classification of SVM.

G. Evaluation Matrices

The effectiveness of the proposed model is evaluated using standard performance metrics like accuracy, Precision, Recall, F1-Score, False Positive rate and False Negative Rate. This will provide a holistic view of model performance under both detection efficacy and misclassification risk.

IV. Result and Discussion

This section presents the experimental results using the above-discussed (in the methodology section) metrics by assessing 10-fold cross-validation to ensure generalisation and mitigate overfitting. The result is evaluated on individual supervised classifiers (SVM, KNN, C4.5 Decision Tree) and the proposed hybrid model.

A. Performance of Individual Classifiers

The performance of individual supervised classifiers (SVM, KNN, C4.5 Decision Tree) is shown in the following Table 2.SVM shows the highest accuracy among all due to its strong margin optimization capability in high-dimensional spaces. However, C4.5 Decision Tree has better interpretability, and KNN shows robustness for localised patterns.

Table II: Performance of Individual Supervised Classifiers

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FPR (%)	FNR (%)
C4.5	91.3	90.8	91.5	91.1	4.7	8.5
KNN (k=5)	88.6	86.9	87.3	87.1	6.8	12.7
SVM	93.2	92.5	92.9	92.7	3.9	7.1

B. Hybrid Model Performance (C4.5 + SVM)

The proposed hybrid model, which integrates C4.5 and SVM via majority voting, achieved superior performance as shown in Fig 2. As per the following pie chart illustrated in Fig. 2, the hybrid system achieves a 95.4% accuracy, as well as reducing both false positive and false negative rates. This shows that combining the strength of C4.5 with the classification power of SVM yields a more reliable and generalizable intrusion detection mechanism.

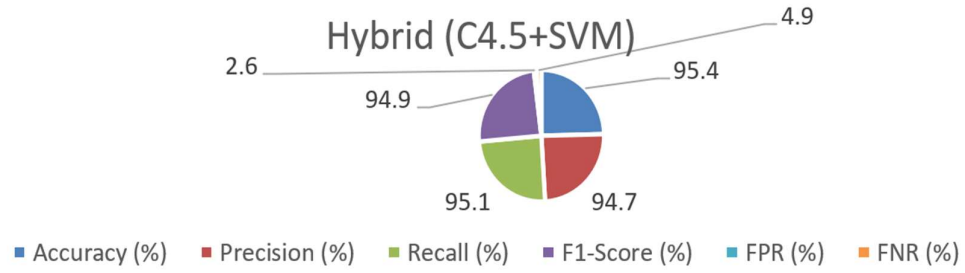

Fig.2: Pie Chart to show to show the performance of Hybrid Model

C. Comparison of Performance of different Model

Fig. 3 illustrate the performance of different machine learning model based on the different performance metrices. The figure shows that the hybrid model achieves better performance metrices as compared to all individual machine learning classifier.

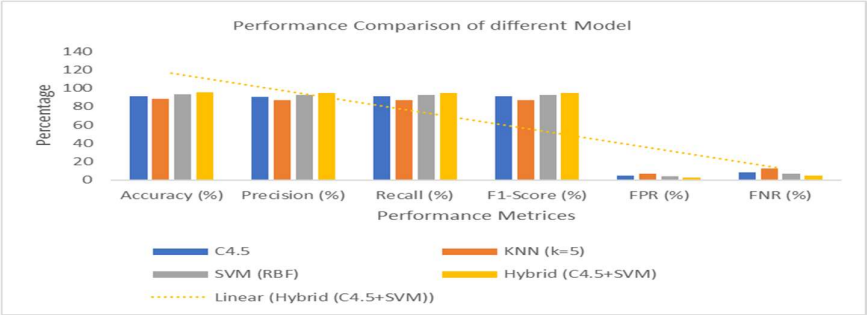

Fig. 3: Comparison diagram of Different models between Different Performance metrics

D. Discussion

These results demonstrate how successful Hybrid Machine Learning Methods are at detecting intrusions using cloud computing resources. Depending on which combination of Synthetic Preceding and Layered features were used, it was found that hybrid methods can help improve overall detection accuracy while minimizing the number of false alerts experienced. This is especially important as more and more organizations utilize Cloud Computing and rely on the availability of rapidly deployed alerts to mitigate risks associated with Intrusions. Additionally, utilizing Feature Selection Techniques aided in reducing Data Dimensions, leading to faster Model Training and increased Detection Accuracy.

The proposed hybrid model is computationally efficient because it combines lightweight machine learning algorithms instead of computationally intensive deep learning models. The use of feature selection methods such as ANOVA and SelectKBest further reduces dimensionality, thereby lowering computational cost and improving training speed. This contributes to sustainable AI by minimizing resource consumption.

Future work may also compare the proposed hybrid model with deep learning approaches such as CNN and LSTM to further validate performance.

V. Conclusion

The growth of cloud computing as part of the foundation for modern information technology has raised security issues concerning advanced hacker techniques. The use of older styles of intrusion detection techniques will not function well on the complex, large, and viewed as networked applications that represent the unique nature of Cloud Computing and are particularly vulnerable to new and unknown vulnerabilities and multi-step types of attacks. To combat these issues, a combination of decision support tools using machine learning, including C4.5 decision trees and support vector machines (SVM), was proposed as a hybrid, or fusion approach, for building a secure environment within cloud services.

The NSL-KDD dataset was used to evaluate the effectiveness of the system and it closely mimics real cloud traffic in the real world. Through experimentation, the framework used techniques such as SelectKBest and ANOVA F-tests to identify relevant attributes from the dataset. Thus, increasing accuracy, while decreasing the need for processing power. The Experimental results indicate that the hybrid model had higher accuracy, precision, recall and f1-score values than individual classifiers with 95.4%, 94.7%, 95.1% and 94.9% , respectively. The low number of false positives, as well as false negatives, produced by this model make it suitable for real-time application in the cloud. The research has proven that implementing multiple machine-learning algorithms (in an ensemble configuration) has been of significant benefit for intrusion detection and increasing the quality of classification.

The present method is scalable and intelligent; it can detect many different kinds of attack types, as well as defense methods, in Cloud-based systems.

References

[1] H. Attou, A. Guezzaz, S. Benkirane, M. Azrour, and Y. Farhaoui, “Cloud-Based Intrusion Detection Approach Using Machine Learning Techniques,” Big Data Mining and Analytics, vol. 6, no. 3, pp. 311–320, 2023.

[2] A. Mahendar and K. S. Chatrapati, “Detection and Prevention of Cyber Attacks on Cloud-Based Data Centers using Machine Learning,” Int. J. Comput. Digit. Syst., vol. 12, no. 1, pp. 1063–1070, 2022.

[3] R. H. Krishna, P. V. Bhaskar, P. Narahari, M. Nalluri, and M. R. Mallapureddy, “Intrusion Detection System on Cloud Computing Using Ensemble SVM,” IJFMR, vol. 6, no. 2, pp. 1–6, 2024.

[4] D. B. Dorothy et al., “AI-Driven Threat Intelligence in Cloud Computing: Detecting and Responding to Cyber Attacks,” in Proc. Int. Conf. on Intelligent Algorithms for Computational Intelligence Systems (IACIS), IEEE, 2024.
[5] T. Olasehinde, “Adaptive Machine Learning Algorithms for Zero-Day Threat Identification in Cloud Systems,” ResearchGate, 2024.
[6] I. M. Selim and R. A. Sadek, “An Effective Fog-aware NIDS for DDoS Attack Detection,” IAENG Int. J. Comput. Sci., 2025.
[7] B. Havano and A. Dobush, “Enhancing Host Intrusion Detection Systems for Linux Based Network Operating Systems,” Visnyk of Lviv Polytechnic National University, 2025. [Online]. Available: https://science.lpnu.ua
[8] S. R. Harsini and D. Houshangi, “Designing a Network Intrusion Detection System Based on Machine Learning for Software-Defined Networks,” PowerTech J., 2025. [Online]. Available: https://paperet.com
[9] M. A. Almarhabi, “The Role of Machine Learning in Enhancing Cloud Computing Security: Critical Review,” IJACST, vol. 11, no. 4, 2025.
[10] A. HAMDOUCHI and I. Ali, “Empowering IoT Security: Deploying TinyML Ensemble Techniques for Cyberattack Detection,” Scientific African, 2025.
[11] K. Oliullah, M. Whaiduzzaman, M. J. N. Mahi, and T. Jan, “A Machine Learning Based Authentication and Intrusion Detection Scheme for IoT in Fog Environment,” PLOS ONE, 2025. DOI: 10.1371/journal.pone.0323954
[12] B. Swathi et al., “Machine Learning Techniques in Cloud Based Intrusion Detection,” in Lecture Notes in Networks and Systems, Springer, 2025.
[13] J. K. Medhi, “Reliable, Secure and Safe Deep Learning for Intrusion Detection in Cloud Computing,” OhioLINK Electronic Theses and Dissertations Center, 2025.
[14] G. Karamchand, “AI-Optimized NFV Security in Cloud Infrastructure,” Int. J. Humanities and Information Technology, vol. 11, no. 2, 2025.
[15] F. Kandah, T. Mendis, L. Medury, and H. Sherawat, “Navigating IoT Security: Architectures, Emerging Threats, and Adaptive Countermeasures,” in IEEE Conf. on IoT Systems, 2025.
[16] D. Kalimumbalo et al., “SecFedMDM-1: Federated Malware Detection Model for Interconnected Cloud,” in IEEE Int. Conf. Cloud Computing, 2025.
[17] Y. Zeng and G. Liu, “Joint SFC Routing Update and Elastic Resource Configuration in Edge Cloud,” IEEE Internet of Things J., 2025.
[18] A. Setha, “A Review of Intrusion Detection Schemes in Fog Computing Using ML and DL,” Int. J. of Advanced Scientific Computing, vol. 9, no. 2, 2025.
[19] A. Binthiya and S. Ravindran, “An Effective Classification Using Enhanced Flame Seeker Optimization with CNN for IDS,” Wireless Networks, Springer, 2025.
[20] P. Pranav and S. Dutta, “A Systematic Literature Review on Intrusion Detection Techniques in Cloud Computing,” Discover Computing, Springer, 2025.