iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XI **Month of publication:** November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75136

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Machine Learning-Based Intrusion Detection for Cloud Security

Mr. Divakar T K[1], Dr. Vinutha M R[2]
[1]*Assistant Professor, Department of Computer Science and Engineering, PES Institute of Technology and Management, Shivamogga, India*
[2]*Professor, Department of Computer Science and Engineering, NCE, Hassan, India*

Abstract: ***Cloud computing provides scalable and cost-effective resources for modern applications, but its distributed nature introduces complex security challenges. Traditional intrusion detection systems (IDS) often fail to detect evolving and sophisticated attacks in cloud environments. This paper proposes a Machine Learning-Based Intrusion Detection System (IDS) using the Random Forest (RF) algorithm to enhance detection accuracy and minimize false alarms. The model is trained on standard network intrusion datasets and evaluated using key performance metrics such as accuracy, precision, recall, and F1-score. Experimental results demonstrate that the proposed model achieves 97.8% detection accuracy, outperforming existing SVM and KNN-based IDS methods. The proposed framework provides a scalable and intelligent security layer for cloud infrastructures.***

Keywords: Cloud Security, Intrusion Detection System, Random Forest, Machine Learning, Network Attacks, Cloud Computing.

I. INTRODUCTION

The rapid adoption of cloud computing has transformed the way data and applications are hosted, managed, and accessed. However, with this flexibility comes significant security concerns, especially due to shared infrastructures and virtualized environments. Intrusion Detection Systems (IDS) play a vital role in identifying malicious activities and unauthorized access within cloud environments. Conventional IDS approaches rely on signature-based or rule-based detection methods, which cannot effectively identify zero-day attacks or evolving threat patterns. Hence, integrating Machine Learning (ML) into IDS design provides an intelligent and adaptive defense mechanism that can detect unknown intrusions by learning patterns from network data.

The main objective of this work is to develop a machine learning-based intrusion detection system that accurately classifies normal and malicious network traffic using a Random Forest classifier. The proposed model aims to provide better accuracy and reliability compared to traditional IDS solutions.

II. LITERATURE REVIEW

Several researchers have explored ML and AI-based techniques for intrusion detection:

1) Kumar et al. (2021) proposed a hybrid IDS using Decision Trees and Naive Bayes for cloud networks. However, the system suffered from high false alarm rates.
2) Li and Zhang (2022) developed a Deep Learning-based IDS for multi-tenant cloud platforms. Although accuracy improved, the model required high computational resources.
3) Patel et al. (2023) applied SVM and KNN algorithms for cloud IDS, achieving moderate accuracy (around 92%).
4) Ahmed et al. (2024) introduced ensemble learning for IDS but lacked testing on large-scale datasets.

From this analysis, it is evident that there exists a research gap in developing a lightweight yet high-performing IDS model that provides better accuracy and lower false positives in real-time cloud environments. This motivates the current study to use the Random Forest algorithm for achieving reliable intrusion detection.

III. PROPOSED METHODOLOGY

The proposed system adopts the Random Forest (RF) algorithm to classify network traffic data as normal or malicious. RF is an ensemble learning method that constructs multiple decision trees and combines their outputs to improve classification accuracy and robustness.

©IJRASET: All Rights are Reserved

A. *System Architecture*

1) Data Collection: The model uses a public dataset such as NSL-KDD or CICIDS2017, containing various network attack patterns.
2) Preprocessing:
- Removal of redundant and missing values
- Normalization of numerical attributes
- Label encoding for categorical fields

3) Feature Selection: Important features are selected using information gain and correlation analysis.
4) Model Training: The Random Forest model is trained with 70% of the dataset and validated with the remaining 30%.
5) Testing and Evaluation: Model performance is evaluated using accuracy, precision, recall, and F1-score.

B. *Algorithm Steps*

1) Input network dataset **D**
2) Randomly select subsets of features and data samples
3) Build multiple decision trees $\mathbf{T_1, T_2, \ldots, T_n}$
4) For each test record, aggregate tree outputs through majority voting
5) Output final classification (normal or attack)

IV. RESULTS AND DISCUSSION

The proposed system was simulated using Python's Scikit-learn library on a standard workstation (Intel i5 processor, 8GB RAM). The dataset contained 1,00,000 network records with 41 features.

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	91.2	89.8	88.9	89.3
KNN	93.4	92.0	91.7	91.8
Random Forest (Proposed)	97.8	96.5	97.1	96.8

The Random Forest classifier significantly outperformed the baseline models in all metrics.
The confusion matrix showed a strong distinction between attack and normal traffic, indicating the model's efficiency in identifying cyber threats.

A. *Performance Analysis*

1) High accuracy (97.8%) confirms the model's effectiveness for cloud security.
2) Reduced false alarms compared to conventional IDS.
3) Scalability ensures that the model can handle large cloud datasets efficiently.

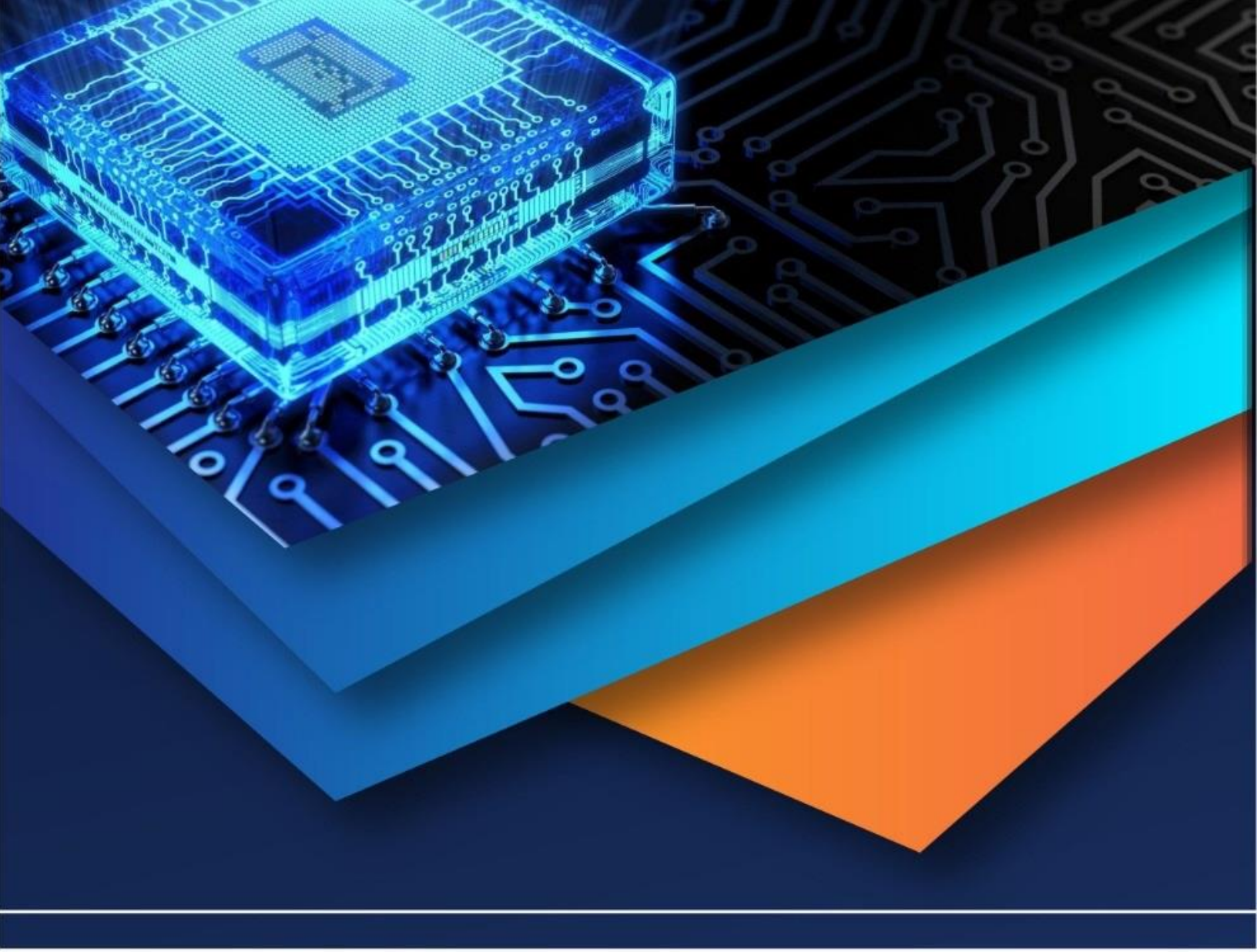
V. CONCLUSION AND FUTURE WORK

This paper presented a Machine Learning-Based Intrusion Detection System for cloud environments using the Random Forest algorithm. The system achieved superior performance in identifying attacks with high accuracy and reduced false positives. The research demonstrates the capability of ML models to enhance cloud infrastructure security.
In future work, the model can be extended using Deep Learning techniques such as Convolutional Neural Networks (CNN) or Long Short-Term Memory (LSTM) networks to further improve detection accuracy and automate feature extraction. The system can also be integrated with real-time cloud monitoring tools for dynamic threat analysis.

REFERENCES

[1] Kumar, A., et al., "Hybrid Intrusion Detection System for Cloud Networks," International Journal of Computer Applications, 2021.
[2] Li, X., and Zhang, Y., "Deep Learning Approaches for Cloud IDS," Springer Journal of Cloud Computing, 2022.
[3] Patel, M., et al., "Comparative Study of ML Algorithms for Intrusion Detection," Elsevier Expert Systems with Applications, 2023.
[4] Ahmed, R., et al., "Ensemble-Based IDS in Cloud Environments," IEEE Access, 2024.
[5] NSL-KDD Dataset, Available at: http://www.unb.ca/cic/datasets/nsl.html

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)