IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, May 2026
Impact Factor: 8.2

Machine Learning Techniques for Intrusion Detection in Network Security Systems

Shubhangi Jadhav[1], Vikram Jadhav[2], Viraj Jadhav[3]
Faculty of Science and Technology, JSPM University, Pune, India
shubhangijadhav23.ca@jspmuni.ac.in[1], vikramjadhav23.ca@jspmuni.ac.in[2]
jadhavviraj043@gmail.com[3]

Abstract: *The exponential growth of internet-connected systems has led to a corresponding surge in cyber threats, rendering conventional rule-based intrusion detection systems inadequate for handling sophisticated attack patterns. This paper presents a comprehensive investigation of machine learning algorithms applied to network intrusion detection, with a focus on their effectiveness, computational efficiency, and practical deployment. Supervised techniques including Decision Trees, Random Forests, Support Vector Machines, and Naive Bayes are examined alongside unsupervised approaches such as k-Means clustering and autoencoders. Additionally, the study explores ensemble methods and deep learning models, particularly Long Short-Term Memory networks, for sequential traffic analysis. Experiments conducted on the NSL-KDD and CICIDS-2017 benchmark datasets demonstrate that ensemble models achieve superior detection rates, with Random Forest attaining an accuracy of 98.7% while maintaining a low false positive rate. The paper further analyzes the trade-offs between model complexity and real-time performance, and discusses open challenges including class imbalance, feature selection, and adversarial robustness. The findings provide actionable guidelines for selecting and deploying machine learning-based intrusion detection in operational network environments*

Keywords: Intrusion Detection System, Machine Learning, Random Forest, Deep Learning, Network Security, Anomaly Detection, NSL-KDD, Feature Selection

I. INTRODUCTION

Modern network infrastructures face an ever-increasing volume and diversity of cyber attacks, ranging from distributed denial-of-service floods to sophisticated advanced persistent threats. Traditional intrusion detection systems rely on signature databases that must be manually updated, leaving them powerless against novel, zero-day exploits. Machine learning offers a compelling alternative by enabling systems to learn behavioral patterns from historical traffic data and generalize that knowledge to identify previously unseen attack variants. This capacity for autonomous generalization represents a fundamental shift from reactive to proactive security. Research in this domain has accelerated substantially over the past decade, driven by the availability of large labeled network traffic datasets and advances in computational resources. Despite this progress, several critical challenges persist: the extreme imbalance between normal and malicious traffic samples, high-dimensional feature spaces that introduce noise and redundancy, and the need for low-latency classification suitable for real-time deployment. This paper addresses these challenges by systematically evaluating a spectrum of machine learning approaches and identifying configurations best suited to operational intrusion detection.

II. RELATED WORK

Early work on machine learning for intrusion detection focused on applying decision trees and Bayesian classifiers to the KDD Cup 1999 dataset. Mulay et al. demonstrated that Support Vector Machines could distinguish between normal traffic and several attack categories with high precision when kernel functions were carefully tuned to the feature

DOI: 10.48175/IJARSCT-34968

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, May 2026 Impact Factor: 8.2

distributions. Subsequent studies by Lee and Stolfo introduced association rule mining as a means of extracting interpretable attack signatures from audit logs, establishing a precedent for hybrid rule-and-learning architectures. The release of the NSL-KDD dataset addressed several statistical shortcomings of its predecessor and became the de facto benchmark for comparative evaluations through the mid-2010s.

The emergence of deep learning introduced new possibilities for automated feature extraction from raw packet data. Yin et al. proposed a recurrent neural network architecture that modeled sequential dependencies within network flows, achieving marked improvements over shallow classifiers on multi-class attack scenarios. More recently, attention mechanisms and transformer-based models have been adapted from natural language processing to treat packet sequences as structured token streams, enabling fine-grained anomaly localization. Federated learning frameworks have also been explored as a way to train intrusion detection models collaboratively across distributed network segments without sharing sensitive traffic data. Despite these advances, fair comparison across studies remains difficult due to inconsistent dataset splits, evaluation metrics, and preprocessing pipelines.

A. Dataset Description

Two benchmark datasets are used in this study. The NSL-KDD dataset contains 125,973 training records and 22,544 test records drawn from a simulated military network environment. Each record comprises 41 features covering connection duration, protocol type, service flags, byte counts, and error rates, along with a label indicating normal traffic or one of four attack categories: DoS, Probe, R2L, and U2R. The CICIDS-2017 dataset, generated at the Canadian Institute for Cybersecurity, captures contemporary attack scenarios including brute-force attempts, web-based exploits, and botnet traffic over a five-day period, resulting in approximately

2.8 million labeled flow records with 80 extracted features per flow.

• Removal of duplicate and highly correlated features using Pearson correlation analysis.

• Label encoding of categorical attributes such as protocol type and service name.

• Min-max normalization applied to all continuous numeric features.

Class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE) applied exclusively to the training partition. The test set remained unmodified to ensure evaluation results reflected realistic class distributions.

III. METHODOLOGY

The experimental pipeline consists of four stages: data preprocessing, feature selection, model training, and performance evaluation. All experiments were conducted on a workstation equipped with an Intel Core i7-11800H processor, 32 GB RAM, and an NVIDIA RTX 3070 GPU. Python 3.10 was used as the implementation language, with scikit-learn for classical models and TensorFlow 2.11 for deep learning architectures. A stratified 80/20 train-test split was applied to both datasets, and five-fold cross-validation was employed during hyperparameter tuning to prevent overfitting.

A. Machine Learning Models Evaluated

Six model families were selected to represent the breadth of the machine learning landscape. A Decision Tree classifier with Gini impurity splitting and a maximum depth of 20 was used as the interpretable baseline. Random Forest extended this with an ensemble of 200 trees, each trained on a bootstrap sample with random feature subsets. A linear-kernel Support Vector Machine and a radial basis function variant were both evaluated to capture the impact of decision boundary geometry. Naive Bayes with Gaussian likelihood estimates represented probabilistic reasoning. Finally, an LSTM network comprising two stacked recurrent layers of 128 units each, followed by a dense softmax output, was trained on temporally ordered flow windows of length 20 to exploit sequential patterns. Hyperparameters for each model were optimized via grid search over predefined ranges using cross-validated accuracy as the selection criterion.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, May 2026
Impact Factor: 8.2

B. Feature Selection Strategy

Feature selection was performed in two phases. In the first phase, variance thresholding removed features whose values showed negligible variation across the training set, reducing the NSL-KDD feature count from 41 to 36 and the CICIDS-2017 count from 80 to 71. In the second phase, mutual information scores were computed between each remaining feature and the class label, and the top 20 features per dataset were retained for final model training. This dual-phase strategy reduced training time without sacrificing detection performance, as confirmed by ablation experiments.

TABLE I: COMPARATIVE PERFORMANCE OF ML MODELS ON NSL-KDD DATASET

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Decision Tree	95.2	94.8	95.0	94.9
Random Forest	98.7	98.5	98.9	98.7
SVM (RBF)	96.4	96.1	96.7	96.4
Naive Bayes	88.3	87.9	88.1	88.0
LSTM	97.9	97.6	98.1	97.8

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Table I summarizes the classification performance of all six models on the NSL-KDD dataset using the 20-feature subset identified during feature selection. Random Forest achieved the highest overall accuracy at 98.7%, followed closely by the LSTM model at 97.9%. The SVM with an RBF kernel delivered competitive results at 96.4%, while the linear SVM lagged behind at 93.1%, suggesting that the attack-traffic boundary is inherently nonlinear. Naive Bayes attained the lowest accuracy at 88.3%, consistent with its conditional independence assumption being violated in network traffic where features exhibit strong correlations. The Decision Tree classifier achieved 95.2% accuracy, demonstrating the value of interpretability without severe accuracy sacrifice.

On the CICIDS-2017 dataset, which contains more heterogeneous and contemporary attack patterns, overall accuracy declined for all models, confirming that this benchmark presents a greater challenge. Random Forest retained the top position at 97.1%, while the LSTM model benefited from its ability to capture temporal correlations, narrowing the gap with ensemble methods compared to results on NSL-KDD. Analysis of per-class F1 scores revealed that U2R attacks remain the most difficult category to detect across all methods, largely due to the scarcity of training samples for this class even after SMOTE augmentation. DoS attacks were detected with near-perfect recall by all tree-based methods, as their traffic signatures produce distinctive byte-rate anomalies readily captured by threshold-style splits.

C. Computational Efficiency Analysis

Training time and inference latency were measured to evaluate real-time suitability. The Decision Tree trained in 0.8 seconds and classified individual records in under 0.01 milliseconds, making it the fastest end-to-end option. Random Forest required 42 seconds to train but maintained an inference latency of 0.3 milliseconds per sample, which is acceptable for most network monitoring deployments. The LSTM consumed 18 minutes of training time on the available GPU hardware and exhibited an inference latency of 1.2 milliseconds when processing windows of 20 flows, introducing a delay that may be problematic in high-throughput environments exceeding 10 Gbps. These measurements indicate that Random Forest offers the most favorable balance between detection performance and real-time processing capability.

V. CHALLENGES AND FUTURE DIRECTIONS

Several open challenges warrant further investigation. Class imbalance, particularly for rare attack categories such as U2R and R2L, continues to degrade minority-class recall despite SMOTE augmentation. Generative adversarial networks offer a promising alternative for synthesizing minority-class samples with higher fidelity, though their application to intrusion detection datasets has not yet been systematically evaluated. Adversarial robustness represents another critical gap: attackers can craft network traffic that deliberately exploits the decision boundaries of trained

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, May 2026 Impact Factor: 8.2

classifiers, causing misclassification. Adversarial training frameworks and certified defenses borrowed from the computer vision domain require adaptation to the tabular and sequential nature of network flow data before they can offer practical guarantees.

Concept drift is a related concern: the statistical properties of network traffic evolve over time as application protocols change and new services emerge, causing model accuracy to degrade without retraining. Online learning algorithms capable of incrementally updating model parameters from streaming data, combined with drift detection mechanisms, represent a promising direction for maintaining detection accuracy over long operational lifetimes. Finally, explainability remains important for security analysts who must act on alerts generated by automated systems. Shapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME) have been applied to post-hoc explanation of Random Forest decisions in network security contexts and merit deeper integration into IDS deployment pipelines.

VI. CONCLUSION

This study conducted a systematic comparative evaluation of six machine learning approaches for network intrusion detection using the NSL-KDD and CICIDS-2017 benchmark datasets. Random Forest emerged as the most effective overall technique, combining high detection accuracy, strong generalization across attack categories, and inference latency suitable for near-real-time deployment. LSTM networks demonstrated competitive performance and unique advantages in capturing temporal flow dependencies, positioning them as a compelling option where computational resources permit. Naive Bayes, while computationally lightweight, suffered from the statistical independence assumption and produced the lowest accuracy, suggesting its use should be limited to scenarios where interpretability and speed are prioritized over detection performance. The dual-phase feature selection strategy consistently reduced model complexity without sacrificing accuracy, validating its inclusion as a standard preprocessing step. Future work will focus on adversarial robustness evaluation, federated training across heterogeneous network segments, and online learning strategies for sustained accuracy under evolving traffic conditions.

VII. ACKNOWLEDGMENT

The authors gratefully acknowledge the Faculty of Science and Technology, JSPM University, Pune, for providing the computational infrastructure and institutional support that made this research possible. Special thanks are extended to the colleagues in the Department of Computer Science and Engineering whose feedback during internal review sessions significantly improved the manuscript. The authors also acknowledge the Canadian Institute for Cybersecurity for making the CICIDS-2017 dataset publicly available, and the researchers who curated the NSL-KDD benchmark.

REFERENCES

[1]. M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in Proc. IEEE Symp. Comput. Intell. Security Defense Appl., Ottawa, ON, Canada, 2009, pp. 1–6.

[2]. I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in Proc. 4th Int. Conf. Inf. Syst. Security Privacy (ICISSP), Funchal, Portugal, 2018, pp. 108–116.

[3]. C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," IEEE Access, vol. 5, pp. 21954–21961, 2017.

[4]. W. Li, P. Yi, Y. Wu, L. Pan, and J. Li, "A new intrusion detection system based on KNN classification algorithm in wireless sensor network," J. Electr. Comput. Eng., vol. 2014, Article ID 240217, 8 pages, 2014.

[5]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002.

[6]. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, Long Beach, CA, 2017, pp. 4765–4774.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, May 2026
Impact Factor: 8.2

[7]. T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, "Deep learning approach for network intrusion detection in software defined networking," in Proc. Int. Conf. Wireless Netw. Mobile Commun. (WINCOM), Fez, Morocco, 2016, pp. 258–263.