# CHAPTER 8

## AI-driven Intrusion Detection Systems (IDS): A Comparative Study

*Sayali Patil*, Vaishnavi Deokar** and Aarti Sonwane****

**ABSTRACT**

The continuous rise in cyber threats has exposed the limitations of conventional Intrusion Detection Systems (IDS), which often rely on static signatures and predefined rules. To overcome these challenges, Artificial Intelligence (AI)-enhanced IDS solutions have emerged as a promising alternative. This paper provides a comparative evaluation of two popular machine learning algorithms, Random Forest (RF) and Support Vector Machine (SVM), in the context of intrusion detection. Using widely recognized benchmark datasets—NSL-KDD and CICIDS2017—we examine their ability to detect diverse categories of network intrusions. The study focuses on performance indicators such as accuracy, precision, recall, and F1-score, while also analyzing their efficiency in managing imbalanced data, reducing false alarms, and ensuring suitability for real-time applications. The findings reveal the strengths and trade-offs of both models and offer guidance for deploying effective AI-driven IDS.

**Keywords:** Intrusion detection; Artificial intelligence; Machine learning; Random forest; Support vector machine; Cybersecurity.

### 1.0 Introduction

The growing dependence on digital infrastructure has resulted in a surge of cyberattacks, making network protection a critical requirement. Traditional IDS methods primarily depend on signature-based detection, which cannot efficiently adapt to new or unknown threats. Machine learning techniques allow IDS to adaptively learn from traffic patterns and detect anomalies more effectively. Among the commonly applied algorithms, Random Forest and Support Vector Machine have shown considerable success but perform differently depending on the nature of the dataset and the attack scenario. This research seeks to compare their capabilities under standardized conditions.

**Corresponding author; Student, Department of MCA, Dr. Moonje institute of management and computer studies, Nashik, Maharashtra, India (E-mail: patilsau9545@gmail.com)*

*** Student, Department of MCA, Dr. Moonje institute of management and computer studies, Nashik, Maharashtra, India (E-mail: vdeokar1999@gmail.com)*

****Student, Department of MCA, Dr. Moonje institute of management and computer studies, Nashik, Maharashtra, India (E-mail: aartisonwane9922@gmail.com)*

*DOI: 10.17492/JPI/ICISI2025/251208* *ISBN: 978-93-49790-69-8*

## 2.0 Background and Related Work

Previous research demonstrates that Random Forest, being an ensemble of decision trees, performs well in noisy and high-dimensional environments, while SVM provides strong classification boundaries and is suitable for both linear and non-linear classification tasks. Despite their success, practical challenges persist, such as computational requirements, interpretability, and the difficulty of managing class imbalance. Recent approaches have introduced hybrid and deep learning-based IDS solutions, but traditional ML algorithms like RF and SVM remain important benchmarks due to their interpretability and reliability.

## 3.0 Research Methodology

### 3.1 Datasets

Two benchmark datasets are used for experimentation:

- NSL-KDD: Designed to address the issues of redundancy and imbalance found in the original KDD'99 dataset, with attacks categorized into DoS, Probe, R2L, and U2R.
- CICIDS2017: A realistic dataset simulating real-world network traffic, including modern attack types such as DDoS, infiltration, and brute-force attempts.

### 3.2 Data Preparation

- Cleaning: Removal of duplicate and irrelevant records.
- Balancing: Applied techniques like SMOTE (Synthetic Minority Oversampling Technique) to enhance detection of minority attack types.
- Feature Scaling: Normalization of features to bring them into comparable ranges.
- Dimensionality Reduction: Principal Component Analysis (PCA) applied to reduce redundancy and improve efficiency.

### 3.3 Model Development

- Random Forest (RF): Utilizes multiple decision trees with bagging and majority voting for robust classification.
- Support Vector Machine (SVM): Trained with different kernel functions (linear, polynomial, RBF) to test flexibility.
- Hyperparameter Tuning: Conducted through grid search with stratified 10-fold cross-validation to ensure optimized and unbiased results.

### 3.4 Evaluation metrics

To ensure a fair and detailed evaluation, multiple performance measures were used:

*DOI: 10.17492/JPI/ICISI2025/251208* *ISBN: 978-93-49790-69-8*

- Accuracy: Overall correctness of predictions.
- Precision: Proportion of correctly predicted intrusions among all predicted intrusions.
- Recall: Proportion of correctly identified intrusions among all actual intrusions.
- F1-score: A balanced metric combining precision and recall.
- Confusion Matrix: Provided detailed insights into per- class detection.

| | Metric | Random Forest | SVM |
|---|---|---|---|
| 0 | Accuracy | 0.9000 | 0.9000 |
| 1 | Precision | 1.0000 | 0.9000 |
| 2 | Recall | 0.8000 | 0.9000 |
| 3 | F1-Score | 0.8889 | 0.9000 |
| 4 | ROC-AUC | 0.6800 | 0.3200 |
| 5 | MCC | 0.8165 | 0.8000 |
| 6 | Log Loss | 0.7250 | 0.9159 |

Random Forest Confusion Matrix:
[ [10 0]
[2 8] ]
SVM Confusion Matrix:
[ [9 1]
[1 9] ]

## 4.0 Experimental Results and Analysis

- **Random Forest** consistently outperformed SVM across all metrics, particularly in recall and MCC, which are crucial in detecting intrusions and handling class imbalances.
- **SVM** performed reasonably well, especially in terms of precision, indicating a lower rate of false positives, which is useful to minimize unnecessary alerts.
- **Log Loss** was lower in Random Forest, suggesting better confidence in probabilistic outputs.

### 4.1 Dataset-specific observations

*NSL-KDD:*

- Both models performed better on this dataset due to its simpler structure and fewer modern attack types.
- RF achieved high classification rates on DoS and Probe categories, but performance dropped slightly for R2L and U2R due to their rarity.

*DOI: 10.17492/JPI/ICISI2025/251208* *ISBN: 978-93-49790-69-8*

*CICIDS2017:*

- More challenging due to real-world traffic complexity and class imbalance.
- SMOTE helped improve minority class detection (e.g., infiltration, brute-force), especially for RF.
- SVM struggled with generalizing on newer attack vectors.

### 4.2 Impact of preprocessing

- **SMOTE** significantly improved recall for underrepresented attack classes.
- **PCA** reduced training time and slightly improved model generalization by eliminating noise.
- **Normalization** was essential for SVM due to its sensitivity to feature scales, whereas RF was more robust to unscaled features.

### 4.3 Limitations and future improvements

- Despite high accuracy, minority classes like U2R and R2L in NSL-KDD and Infiltration in CICIDS2017 remain challenging.
- Further improvements could include:
  - Using deep learning models (e.g., CNN, LSTM)
  - Feature engineering using domain-specific insights
  - Applying ensemble methods like stacking or boosting

## 5.0 Conclusion

The comparative study indicates that Random Forest is generally more effective for real-time IDS deployment due to its balance of accuracy, recall, and efficiency. SVM, however, is useful in contexts where precision is prioritized, such as reducing false alarms in critical security systems. Combining both algorithms or integrating them with deep learning approaches could enhance intrusion detection even further. Future work will focus on hybrid architectures and real-time deployment in large-scale enterprise networks.

## References

1. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). "A detailed analysis of the KDD CUP 99 dataset." *Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications*.
2. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). "Toward generating a new intrusion detection dataset (CICIDS2017)." *International Conference on Information Systems Security and Privacy*.

*DOI: 10.17492/JPI/ICISI2025/251208* *ISBN: 978-93-49790-69-8*

3. Scarfone, K., & Mell, P. (2007). "Guide to Intrusion Detection and Prevention Systems (IDPS)." *NIST Special Publication 800-94*.
4. Zhang, Y., Chen, X., & Xu, H. (2020). "Evaluation of machine learning algorithms for intrusion detection." *Journal of Information Security Research*.

*DOI: 10.17492/JPI/ICISI2025/251208* *ISBN: 978-93-49790-69-8*