Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 03 Issue: 11 | Nov-2025
ISSN (O) 3107-6696

AI in Cybersecurity: Intrusion Detection using Machine Learning

Smriti Thakur[1], Dr. Yatu Rani[2]

[1,2]Dr. Akhilesh Das gupta Institute of Professional Studies, New Delhi, Delhi, India.
[1]smritithakur2552@gmail.com, [2]yaturani@gmail.com

ABSTRACT - The rapid increase in cyberattacks across modern digital infrastructures has highlighted the urgent need for intelligent and adaptive security solutions. Traditional intrusion detection systems (IDS), though widely deployed, struggle to detect novel, evasive, and complex threats due to their dependence on predefined signatures and static rules.

Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL), has emerged as a powerful alternative for intrusion detection by learning from historical attack patterns and modeling anomalous behaviors. This research paper provides a structured overview of AI-driven intrusion detection techniques, focusing on signature-based, anomaly-based, and hybrid ML models. The paper examines their mechanisms, performance strengths, limitations, and practical applicability across different network environments.

Furthermore, it evaluates commonly used datasets such as NSL-KDD and CICIDS2017 that form the foundation of IDS research. The study also identifies challenges in deploying ML-based IDS, including interpretability issues, scalability constraints, adversarial vulnerabilities, and false positive rates.

Keywords: Intrusion detection, machine learning, cybersecurity, anomaly detection, deep learning.

ABBREVIATIONS

IDS – Intrusion Detection System
AI – Artificial Intelligence
ML – Machine Learning
DL – Deep Learning
DoS – Denial of Service
DDoS – Distributed Denial of Service
SVM – Support Vector Machine
LSTM – Long Short-Term Memory
XAI – Explainable Artificial Intelligence
NSL-KDD – Network Security Laboratory–KDD Dataset

1. INTRODUCTION

Cybersecurity has become a critical priority in an era where digital connectivity permeates every sector—ranging from finance and healthcare to cloud computing, IoT ecosystems, and enterprise infrastructures. The surge in cyberattacks, including ransomware, phishing, DDoS attacks, and botnets, places organizations at constant risk. Traditional defense systems such as firewalls and antivirus software provide perimeter-based protection but are insufficient against sophisticated attacks that evolve rapidly and exploit unknown vulnerabilities.

Intrusion Detection Systems (IDS) are designed to identify unauthorized, malicious, or anomalous activities within a network. Conventional IDS approaches—primarily signature-based and anomaly-based—face limitations in scalability, adaptability, and zero-day threat detection. These constraints have driven the integration of Artificial Intelligence (AI) into IDS design.

Machine Learning (ML) techniques can analyze large amounts of network traffic data to discover patterns, classify malicious behavior, and detect deviations indicative of intrusions. Classical ML models like Decision Trees, Random Forest, and SVM offer strong detection capabilities, whereas deep learning architectures such as Autoencoders and LSTM networks provide enhanced adaptability for complex attack behaviors. Benchmark datasets like NSL-KDD and CICIDS2017 have enabled systematic research and evaluation of these models.

Despite their promise, ML-driven IDS face several challenges involving computational complexity, data imbalance, interpretability, and adversarial robustness. This paper provides a structured examination of AI-driven IDS approaches, critically evaluating their effectiveness and outlining challenges and future research directions.

2. CYBERSECURITY THREAT LANDSCAPE

Modern networks face a diverse range of threats that challenge the capabilities of traditional IDS. Cyberattacks have evolved from simple rule-based exploits to advanced multi-stage intrusions capable of bypassing conventional defenses.

A significant portion of attacks involves Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks, which

overload network resources, causing service disruptions. Remote-to-Local (R2L) and User-to-Root (U2R) attacks exploit vulnerabilities to gain unauthorized privileges. Contemporary threat vectors include botnets, brute-force attacks, infiltration attempts, and web-based exploits that target application-level weaknesses.

Zero-day attacks pose an especially severe threat because they exploit unknown vulnerabilities, rendering signature-based IDS ineffective. Attackers increasingly use stealthy and evasive techniques, such as slow-rate DDoS or encrypted traffic exploitation, making detection more challenging. As attack patterns become more dynamic and polymorphic, there is a clear need for intelligent systems that can learn from evolving data and detect anomalies beyond predefined rules.

These complexities underscore the necessity for AI-driven IDS that can accurately analyze, classify, and adapt to emerging cyber threats in real time.

3. TAXONOMY OF AI-DRIVEN INTRUSION DETECTION TECHNIQUES

AI-based IDS techniques can be categorized based on their detection methodologies, learning strategies, and underlying algorithms. The taxonomy presented here outlines the primary approaches used in modern IDS research.

Signature-based ML models rely on learning known attack patterns. While highly accurate for familiar threats, they fail against novel attacks. Decision Trees and Random Forest classifiers are often used in this category due to their interpretability and strong classification capabilities.

Anomaly-based ML models detect deviations from normal network behavior. Methods such as SVMs, Autoencoders, and clustering algorithms model typical network traffic and identify abnormal variations. These systems excel at detecting previously unseen intrusions but may suffer from high false positive rates.

Deep Learning-based IDS leverage architectures like LSTM networks to analyze sequential traffic patterns and Autoencoders for unsupervised anomaly detection. These approaches capture complex temporal and non-linear relationships but require intensive computational resources.

Hybrid IDS models combine multiple techniques—such as integrating anomaly detection with supervised classification or layering ML with rule-based systems—to enhance detection accuracy and reduce false alarms. Hybrid models are increasingly used due to their ability to balance performance, efficiency, and adaptability.

Datasets such as NSL-KDD offer structured features ideal for classical ML models, while CICIDS2017 provides realistic traffic flows suitable for deep learning approaches. Together, these datasets support comprehensive evaluation of IDS performance across both classical and modern attack scenarios.

4. CRITICAL EVALUATION OF ML-BASED IDS

Each ML model used for intrusion detection has distinct benefits and limitations depending on dataset characteristics, attack types, and network environments. Random Forest models typically achieve high accuracy and low false positives on structured datasets like NSL-KDD, making them suitable for enterprise systems with well-defined traffic patterns. Decision Trees provide transparency but may lack the robustness required for complex traffic conditions.

SVM classifiers offer strong performance in distinguishing attack categories but struggle with scalability when handling high-dimensional data. Autoencoders excel at detecting unknown attacks through reconstruction errors but can be sensitive to noise and require careful tuning. LSTM-based models effectively analyze sequential traffic, capturing multi-stage intrusions patterns, but their training times and resource requirements remain significant barriers to real-time deployment.

A key evaluation challenge lies in balancing accuracy with false positive rates. Anomaly-based systems may detect zero-day attacks but often generate frequent false alarms, reducing their practical usability. Deep learning models provide adaptability but lack interpretability, creating difficulty in understanding their decision-making processes and reducing trust among cybersecurity analysts.

5. FUTURE RESEARCH DIRECTIONS

Future research on AI-driven IDS is expected to focus on several key areas. Explainable AI (XAI) will play a significant role in enhancing transparency and analyst trust. Developing interpretable deep learning models can bridge the gap between accuracy and explainability.

Federated Learning (FL) offers promising directions for decentralized IDS by enabling collaborative model training

without sharing raw data. However, research is required to address gradient leakage and communication overhead.

Adversarial robustness is another emerging focus area, as ML models remain vulnerable to evasion attacks, where adversaries craft inputs to mislead detection systems. Strengthening ML-based IDS against adversarial manipulation is essential for reliable deployment.

Real-time IDS deployment requires more efficient models capable of processing high-volume data streams with low latency. Research into lightweight deep learning architectures and hardware-accelerated inference will be crucial for supporting real-time detection.

Hybrid AI frameworks combining multiple ML techniques are expected to dominate the next generation of IDS, offering adaptive, resilient, and scalable protection against evolving threats.

6. CHALLENGES AND OPEN ISSUES

The integration of AI-based intrusion detection faces several challenges. Data imbalance is persistent in IDS datasets, where normal traffic significantly outweighs malicious samples, leading to biased model training. Scalability remains another concern, as enterprise and IoT networks generate massive traffic that must be analyzed in real time.

Interpretability is a major open issue, particularly for deep learning models. Analysts require transparency when evaluating ML-driven alerts and ensuring compliance with regulatory frameworks. Additionally, ensuring robustness against adversarial attacks poses a significant challenge, as attackers continue to develop techniques to exploit ML model vulnerabilities.

Other challenges include dataset limitations, encrypted traffic analysis, integration with existing security architectures, and the high computational cost of training advanced ML models. These open issues highlight the need for future research focused on secure, explainable, and scalable IDS solutions.

7. CASE STUDIES

AI-based IDS solutions have seen practical implementation across various sectors. Enterprise networks often deploy Random Forest–based systems for fast and reliable traffic classification, benefiting from their efficiency and interpretability. Cloud service providers increasingly use deep learning techniques to detect sophisticated threats within large-scale distributed systems.

In IoT environments, anomaly detection using lightweight ML models has improved security for resource-constrained devices. Autoencoder-based models are frequently used in industrial ICS/SCADA networks to detect operational anomalies.

Cybersecurity research labs globally rely on CICIDS2017 and NSL-KDD datasets to benchmark IDS performance under realistic attack conditions. These case studies demonstrate that while AI-driven IDS are highly effective, their practical deployment requires careful balancing of accuracy, efficiency, and interpretability.

8. CONCLUSION

AI-driven intrusion detection represents a significant advancement in modern cybersecurity, offering the ability to analyze complex behaviors, detect unknown threats, and adapt to evolving attack patterns. This paper provided a structured analysis of ML and DL techniques for intrusion detection, covering signature-based, anomaly-based, and hybrid approaches. It also examined dataset contributions, model performance, and limitations.

Although ML-based IDS show strong potential, challenges such as high computational requirements, false positive rates, interpretability issues, and adversarial vulnerabilities must be addressed. Future research should focus on explainable models, federated learning, adversarial defense strategies, and real-time lightweight architectures.

AI-enabled IDS have the potential to redefine cybersecurity by providing intelligent, scalable, and adaptive protection mechanisms. Continued research and innovation will be pivotal to ensuring robust and trustworthy intrusion detection capabilities.

REFERENCES

1. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009).
   A detailed analysis of the KDD CUP 99 dataset. *IEEE Symposium on Computational Intelligence for Security and Defense Applications.*
2. Revathi, S., & Malathi, A. (2013).
   A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion

detection. *International Journal of Computer Applications, 68*(25), 1–6.

3. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP).*
4. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444.
5. Hindy, H., Brosset, D., Bayne, E., Seeam, A., Tachtatzis, C., Atkinson, R., & Bellekens, X. (2019). A taxonomy and survey of intrusion detection system design techniques, network threats and datasets. *IEEE Communications Surveys & Tutorials, 21*(1), 163–204.
6. Lee, W., & Stolfo, S. (1998). Data mining approaches for intrusion detection. *Proceedings of the 7th USENIX Security Symposium.*
7. Canadian Institute for Cybersecurity (CIC). CICIDS2017 Dataset. University of New Brunswick. Available: https://www.unb.ca/cic/datasets/ids-2017.html
8. Dhanabal, L., & Shantharajah, S. P. (2015). A study on NSL-KDD dataset for intrusion detection system based on classification algorithms. *International Journal of Advanced Research in Computer and Communication Engineering, 4*(6).
9. Kim, J., Kim, J., Thu, H. L. T., & Kim, H. (2016). Long short-term memory recurrent neural network classifier for intrusion detection. *IEEE International Conference on Platform Technology and Service (PlatCon).*
10. Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence, 2*(1), 41–50.
11. Zhang, J., Li, C., & Manikopoulos, C. (2018). Anomaly detection using Autoencoders for intrusion detection. *IEEE International Conference on Cyber Security.*
12. Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. *IEEE Symposium on Security and Privacy.*
13. Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security, 28*(1–2), 18–28.
14. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176.
15. Li, W., Yi, P., Wu, Y., Pan, L., & Li, J. (2017). A hybrid deep learning model for anomaly detection. *IEEE International Conference on Applied System Innovation.*