IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Artificial Intelligence-Based Intrusion Detection Systems: A Systematic Review

Dhruv Vaish, Harsimranjeet Kaur

Student, Assiant Professor

GGI, Amritsar

Abstract

This systematic review examines the application of Artificial Intelligence (AI) and Machine Learning (ML) techniques in Intrusion Detection Systems (IDS). The review analyses 162 peer-reviewed studies published between 2015 and 2024, focusing on supervised learning, unsupervised learning, and deep learning approaches used in Network-based IDS (NIDS), Host-based IDS (HIDS), and Hybrid IDS architectures. The findings indicate that deep learning models, especially Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Generative Adversarial Networks (GAN), outperform traditional ML methods on benchmark datasets such as NSL-KDD, UNSW-NB15, and CICIDS2017, achieving F1-scores above 0.98. However, challenges such as class imbalance, adversarial attacks, concept drift, high false positive rates, and limited dataset generalization continue to affect practical deployment. Emerging technologies including Federated Learning, Explainable AI (XAI), and Graph Neural Networks (GNN) show promising results for future IDS development. This review provides a structured overview of AI-based IDS techniques, benchmark datasets, current challenges, and future research directions.

Keywords: Intrusion Detection System (IDS), Artificial Intelligence, Machine Learning, Deep Learning, Cybersecurity, Anomaly Detection, Federated Learning, Explainable AI.

1. Introduction

The rapid growth of cloud computing, Internet of Things (IoT) devices, wireless communication, and digital services has significantly increased cybersecurity threats worldwide. Organizations continuously face attacks such as malware, phishing, ransomware, Distributed Denial of Service (DDoS), and unauthorized access attempts. According to cybersecurity reports, global cybercrime costs are expected to exceed trillions of dollars annually.

Intrusion Detection Systems (IDS) play a critical role in protecting computer networks and information systems. IDS monitor network traffic and system activities to identify suspicious behavior, security violations, and malicious attacks. Traditional IDS mainly rely on signature-based detection methods, which compare incoming traffic against predefined attack signatures. Although effective for known attacks, signature-based IDS fail to detect unknown or zero-day attacks.

To overcome these limitations, Artificial Intelligence (AI) and Machine Learning (ML) techniques have been widely adopted in IDS research. AI-based IDS can automatically learn patterns from network traffic and identify abnormal behavior with greater accuracy. Deep learning approaches further improve detection performance by extracting complex features from large-scale datasets.

This systematic review presents a detailed analysis of AI-based IDS techniques, datasets, challenges, and future directions. The objective is to provide researchers and students with a clear understanding of current advancements in intelligent intrusion detection.

2. Methodology

This review follows the PRISMA 2020 guidelines for systematic literature reviews. Research papers were collected from major digital libraries including:

- IEEE Xplore
- ACM Digital Library
- Springer Link
- ScienceDirect

The following search keywords were used:

("Intrusion Detection System" OR "IDS") AND ("Machine Learning" OR "Deep Learning" OR "Artificial Intelligence") AND ("Cybersecurity" OR "Anomaly Detection")

Inclusion Criteria

- Peer-reviewed journal articles and conference papers
- Published between 2015 and 2024
- Studies related to AI or ML-based IDS
- English language publications

Exclusion Criteria

- Duplicate studies
- Non-peer-reviewed articles
- Studies without experimental evaluation
- Papers unrelated to intrusion detection

Initially, 1,247 papers were identified. After title screening, abstract review, and quality assessment, 162 studies were selected for final analysis.

3. Intrusion Detection System Taxonomy

3.1 Types of IDS Based on Deployment

3.1.1 Network-Based IDS (NIDS)

NIDS monitor network traffic at strategic points such as routers and gateways. They are effective for detecting attacks like DDoS, scanning, and malicious packet transmission.

Advantages:

- Monitors entire network traffic
- Detects external attacks efficiently
- Centralized monitoring

Limitations:

- Difficulty analyzing encrypted traffic
- High computational overhead in large networks

3.1.2 Host-Based IDS (HIDS)

HIDS are installed on individual systems and monitor file integrity, process activity, and system calls.

Advantages:

- Detects insider attacks
- Monitors user activities
- Better visibility into host-level behavior

Limitations:

- Resource consumption on hosts
- Limited network-wide visibility

3.1.3 Hybrid IDS

Hybrid IDS combine NIDS and HIDS to improve detection accuracy.

Advantages:

- Better attack coverage
- Reduced false negatives
- Improved security monitoring

Limitations:

- Complex implementation
- High deployment cost

3.2 IDS Detection Techniques Signature-Based Detection

This method compares traffic patterns with known attack signatures.

Advantages:

- High accuracy for known attacks
- Low false positive rate

Disadvantages:

- Cannot detect zero-day attacks
- Requires frequent signature updates

Anomaly-Based Detection

This approach identifies deviations from normal behavior using statistical or ML techniques.

Advantages:

- Detects unknown attacks
- Adaptive learning capability

Disadvantages:

- High false positive rate
- Requires training data

Specification-Based Detection

This method uses predefined rules and protocol specifications.

Advantages:

- Accurate protocol monitoring
- Lower false positives than anomaly detection

Disadvantages:

- Difficult rule creation
- Requires expert knowledge

4. Machine Learning Techniques in IDS

4.1 Supervised Learning

Supervised learning uses labeled datasets to train classifiers.

Support Vector Machine (SVM)

SVM is widely used for classification problems in IDS.

Advantages:

- Effective for high-dimensional data
- Good classification performance

Limitations:

- Slow training on large datasets
- Sensitive to parameter tuning

Random Forest (RF)

Random Forest uses multiple decision trees for classification.

Advantages:

- High accuracy
- Resistant to overfitting
- Fast prediction speed

Limitations:

- Complex model structure
- Higher memory usage

k-Nearest Neighbor (k-NN)

k-NN classifies data based on neighboring samples.

Advantages:

- Simple implementation
- Good for small datasets

Limitations:

- Slow for large datasets
- High storage requirement

4.2 Unsupervised Learning

Unsupervised learning identifies hidden patterns in unlabeled data.

Clustering Algorithms

Algorithms such as K-Means and DBSCAN group similar traffic patterns.

Advantages:

- Useful for anomaly detection
- No labeled data required

Limitations:

- Sensitive to parameter selection
- Lower accuracy compared to supervised learning

5. Deep Learning Approaches

5.1 Convolutional Neural Networks (CNN)

CNN models automatically extract spatial features from traffic data.

Advantages:

- High detection accuracy
- Effective feature extraction

Limitations:

- High computational cost
- Requires GPU resources

5.2 Long Short-Term Memory (LSTM)

LSTM networks are effective for sequential traffic analysis.

Advantages:

- Detects temporal attack patterns
- Effective for DDoS detection

Limitations:

- Long training time
- Complex architecture

5.3 Autoencoders

Autoencoders learn compressed representations of normal traffic behavior.

Advantages:

- Useful for anomaly detection
- Works without labeled attack data

Limitations:

- Reconstruction threshold selection is difficult
- Sensitive to noisy data

5.4 Generative Adversarial Networks (GAN)

GANs generate synthetic attack samples to solve class imbalance problems.

Advantages:

- Improves minority attack detection
- Enhances dataset quality

Limitations:

- Difficult training process
- Mode collapse issues

5.5 Graph Neural Networks (GNN)

GNN models represent network communication as graph structures.

Advantages:

- Detects multi-stage attacks
- Captures network relationships effectively

Limitations:

- High computational complexity
- Difficult real-time implementation

6. Advanced AI Paradigms

6.1 Federated Learning

Federated Learning allows distributed IDS training without sharing raw data.

Benefits:

- Improved privacy protection
- Reduced centralized data storage

Challenges:

- Gradient poisoning attacks
- Non-IID data distribution

6.2 Reinforcement Learning

Reinforcement Learning enables adaptive security decisions.

Applications:

- Dynamic firewall configuration
- Adaptive threat response

6.3 Explainable AI (XAI)

XAI techniques improve IDS transparency and interpretability.

Popular Techniques:

- SHAP
- LIME
- Attention mechanisms

Benefits:

- Improved analyst trust
- Faster forensic investigation

7. Benchmark Datasets

Dataset	Year	Records	Main Features
KDD Cup 99	1999	4.9 Million	Old benchmark dataset
NSL-KDD	2009	148 Thousand	Improved KDD dataset
UNSW-NB15	2015	2.5 Million	Modern attack categories
CICIDS2017	2017	2.8 Million	Realistic traffic behavior
CSE-CIC-IDS2018	2018	16 Million	Large-scale dataset
BOT-IoT	2018	73 Million	IoT attack traffic

These datasets are commonly used for training and evaluating IDS models.

8. Comparative Analysis of AI Techniques

Technique	Accuracy	Zero-Day Detection	Complexity
SVM	95–98%	Moderate	Medium
Random Forest	99%+	Moderate	Low
CNN	98–99%	Good	High
LSTM	97–99%	Good	High

Technique	Accuracy	Zero-Day Detection	Complexity
CNN-LSTM	99%+	Very Good	Very High
Autoencoder	92–96%	Excellent	High
GAN-Based IDS	96–99%	Very Good	Very High
GNN	92–95%	Excellent	Very High

CNN-LSTM hybrid models currently provide the best overall performance in intrusion detection.

9. Open Challenges

9.1 Class Imbalance

Real-world attack traffic is extremely limited compared to normal traffic, causing poor minority attack detection.

9.2 Adversarial Attacks

Attackers can manipulate traffic patterns to bypass AI-based IDS.

9.3 False Positive Rate

High false alarms create operational burden for security analysts.

9.4 Concept Drift

Network behavior changes over time, reducing model effectiveness.

9.5 Encrypted Traffic Analysis

Encryption limits payload inspection capabilities.

9.6 Dataset Generalization

Models trained on one dataset often fail on different network environments.

10. Future Research Directions Foundation Models for Cybersecurity

Large transformer-based models may improve generalized intrusion detection.

Causal AI

Causal learning can improve robustness against adversarial manipulation.

Neuromorphic Computing

Energy-efficient IDS for IoT devices can be developed using neuromorphic hardware.

Standardized Evaluation Frameworks

Future research should focus on realistic datasets and cross-environment

11. Conclusion

Artificial Intelligence and Machine Learning have significantly improved the performance of Intrusion Detection Systems. Deep learning approaches such as CNN, LSTM, GAN, and GNN demonstrate superior performance compared to traditional IDS techniques. However, several challenges including adversarial attacks, concept drift, false positives, and dataset limitations still restrict practical deployment.

Future IDS research should focus on explainability, privacy preservation, cross-dataset generalization, and adaptive learning systems. The integration of AI with cybersecurity will continue to play a major role in protecting modern digital infrastructures.

References

1. Buczak AL, Guven E. A survey of data mining and machine learning methods for cybersecurity intrusion detection. IEEE Communications Surveys & Tutorials. 2016.
2. Aldweesh A, Derhab A, Emam AZ. Deep learning approaches for anomaly-based intrusion detection systems. Expert Systems with Applications. 2020.
3. Moustafa N, Slay J. UNSW-NB15 dataset for network intrusion detection systems. MilCIS. 2015.
4. Sharafaldin I, Lashkari AH, Ghorbani AA. CICIDS2017 dataset generation for intrusion detection research. ICISSP. 2018.
5. McMahan HB, et al. Communication-efficient learning of deep networks from decentralized data. AISTATS. 2017.
6. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. NeurIPS. 2017.
7. Goodfellow IJ, et al. Generative adversarial networks. NeurIPS. 2014.
8. Lo WW, et al. GNN-based intrusion detection systems for IoT security. IEEE NOMS. 2022.
9. Sommer R, Paxson V. Machine learning for network intrusion detection. IEEE Symposium on Security and Privacy. 2010.
10. Mirsky Y, et al. Kitsune: Online network intrusion detection using autoencoders. NDSS. 2018.