Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

Article - e004

MACHINE LEARNING AND DEEP LEARNING-BASED INTRUSION DETECTION SYSTEMS: A COMPREHENSIVE REVIEW OF DATASETS, ALGORITHMS, CHALLENGES, EXPLAINABILITY, AND FUTURE RESEARCH DIRECTIONS

Sudhanshu Dhote[1] ✉ , **Dr. Deepak Agrawal**[2] ✉

[1]M.Tech. Scholar, Computer Science & Engineering, Institute of Engineering & Technology

[2]Associate Professor, Computer Science & Engineering, Institute of Engineering & Technology

Received: 20/05/2026 **Revision Received**: 08/06/2026 **Accepted**: 23/06/2026

ABSTRACT

Cloud computing, Internet of Things, cyber-physical systems and communications networks are vital technologies with a high growth potential that have made the volume and sophistication of cyberattacks more common and challenging than ever; traditional security mechanisms are no longer adequate for modern threat detection in these rapidly evolving technologies. Machine Learning (ML) and Deep Learning (DL) have proven to be promising solutions to build powerful Intrusion Detection Systems (IDSs) that can detect complex and new attacks. This review will be conducted with a systematic review methodology as guided by PRISMA, and will present the comprehensive analysis of the published ML- and DL-based IDS research from 2020 up to 2026. Over 500 studies were reviewed and 120 high-quality publications were chosen for in-depth review. It reviews popular benchmark datasets, such as NSL-KDD, UNSW-NB15, CICIDS2017, CICDDoS2019, Bot-IoT, IoTID20, and TON_IoT, and analyzes the performances of the popular ML algorithms including SVM, KNN, Random Forest, and XGBoost; and DL architectures including CNN, RNN, LSTM, GRU, Autoencoders, Transformers, and Graph Neural Networks. New areas of research covered include Explainable AI (XAI); Federated Learning; Edge Intelligence; Adversarial Learning and Self-Supervised Learning. The results show that the detection accuracy is generally higher for the DL models, while the interpretability and computational efficiency are higher for the ML models. The following are the top problems found: dataset imbalance, concept drift, adversarial attacks, privacy concerns and real-time deployment constraints. The review then presents the current research opportunities in the field of designing scalable, explainable and trustworthy IDS frameworks in future cybersecurity environments.

KEYWORDS: Intrusion Detection System, Cybersecurity, Machine Learning, Deep Learning, Network Security, Explainable Artificial Intelligence, Federated Learning, Adversarial Learning, Internet of Things, Cyber Threat Detection.

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

1. INTRODUCTION

The development of digital technologies has made the network environment and the related cybersecurity even more complex, such as cloud computing, IoT, edge computing, 5G networks and cyber-physical systems (Ferrag et al., 2020; Moustafa, 2021). The ever-increasing sophistication of cyberattacks has made traditional security technologies, such as firewalls and signature-based intrusion detection systems, less effective than they once were. Cyberattacks are increasingly taking the shape of ransomware, advanced persistent threats (APTs), zero-day exploits and distributed denial-of-service (DDoS) attacks, among others (Khraisat et al., 2019; Buczak & Guven, 2016). Intrusion Detection Systems (IDSs) have thus become a critical part of any current network security infrastructure. AI-based IDSs are different from traditional IDSs in that they use Artificial Intelligence, specifically, Machine Learning (ML) and Deep Learning (DL) techniques to automatically learn patterns from the network traffic, while detecting known and unknown attacks (Sommer & Paxson, 2010; Buczak & Guven, 2016). Supervised machine-learning algorithms like Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Random Forests (RF), Decision Trees (DT), and Gradient Boosting models have been found to work well in benchmark datasets of intrusion detection (Buczak & Guven, 2016; Vinayakumar et al., 2019).

With the latest developments in computational power, DL architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Autoencoders, Transformers and Graph Neural Networks (GNNs) are well suited to represent complex patterns and nonlinear relationships in network traffic, which can be effectively leveraged for network monitoring (Javaid et al., 2016; Yin et al., 2017; Ferrag et al., 2020). However, current limitations in the deployment of these systems include data imbalance, concept drift, adversarial attacks, privacy issues, explainability constraints, and computation complexity (Khraisat et al., 2019; Ferrag et al., 2020). To overcome these challenges, this review presents a comprehensive study of machine learning and deep learning intrusion detection systems of the literature from 2020 to 2026. The study is different from the previous surveys in that it combines benchmark datasets, ML and DL approaches, explainable AI, federated learning, adversarial robustness, and emerging research trends in a comprehensive manner, offering a comprehensive view of future IDSs.

The main aims of this review are:

1. To provide a systematic analysis of new developments of machine learning and deep learning-based intrusion detection systems.

2. To assess benchmark datasets that are widely used for intrusion detection studies.

3. To examine the pros and cons of different machine learning and deep learning techniques.

4. To explore new technologies such as Explainable AI, Federated Learning, and Edge Intelligence.

5. To pinpoint future trends and research areas for next generation Intrusion Detection Systems.

This review continues with the following organization: The methodology of the PRISMA-based review is described in section 2. The evolution and taxonomy of Intrusion Detection System (IDS) are presented in Section 3. The benchmark datasets discussed in Section 4 are used for research on IDS. The next two sections review the machine learning (Section 5) and deep learning (Section 6) based ID approaches. The remaining parts of the study explore hybrid models, explainable AI, federated learning, gaps in the research, and future research directions.

2. REVIEW METHODOLOGY

This section outlines the methodology for a review. This section gives an overview of the Review Methodology.

The number of research articles, conference papers, technical reports, and survey studies related to the intrusion detection system (IDS) based on artificial intelligence (AI) has increased significantly in recent years (Ahmad et al., 2022; Vinayakumar et al., 2019). As a result, a systematic and transparent literature review has come to play a crucial role in bringing together existing knowledge, pinpointing research trends, assessing technological progress, and identifying research gaps. To achieve a rigorous and reproducible method, this review adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines which are one of the most extensively used approaches to conducting systematic reviews in science (Page et al., 2021). The PRISMA method offers an inclusive and systematic approach to identifying, screening, evaluating, and selecting relevant studies. PRISMA

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

does not involve the selection bias that occurs in traditional narrative reviews, as the inclusion and exclusion criteria are predetermined, reporting is transparent and there are systematic quality assessment mechanisms (Moher et al., 2009; Page et al., 2021). PRISMA is especially beneficial in cybersecurity research because it involves various methodologies, data sets, assessment indicators, and application fields (Buczak & Guven, 2016; Ferrag et al., 2020).

The goal of the review methodology is to find high quality studies concerning machine learning and deep learning-based intrusion detection systems (IDSs) from January 2020 to June 2026. There is a special focus on research on benchmark datasets, intelligent classification methodologies, explainable artificial intelligence (XAI), federated learning, adversarial robustness, and next-generation intrusion detection architectures (IEC) (Shone et al., 2018; Ferrag et al., 2022).

The review process will have four main stages:

1. A set of relevant studies were identified using the database search.

2. Selection of the retrieved publications according to predetermined criteria.

3. Eligibility based on full-text evaluation.

4. Studies that meet all the quality criteria will be added at the end.

Figure 1 shows the overall PRISMA workflow that was used in this review.

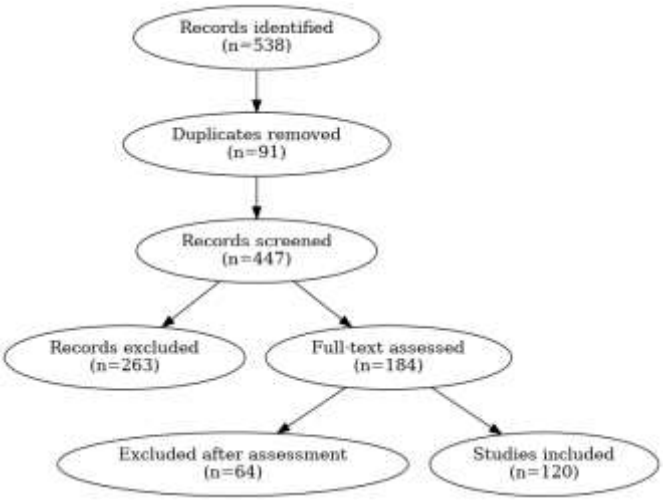

Figure 1: PRISMA Flow Diagram

2.1 Research Questions

The review was guided by a set of carefully formulated research questions (RQs) designed to investigate the current state of intrusion detection research and identify emerging trends (Kitchenham & Charters, 2007).

Table 1: Research Questions Considered in This Review

Research Question ID	Description
RQ1	What machine learning algorithms are most frequently used in intrusion detection systems?
RQ2	Which deep learning architectures demonstrate superior intrusion detection performance?
RQ3	What benchmark datasets are most commonly used for IDS evaluation?
RQ4	What evaluation metrics are employed to assess IDS effectiveness?
RQ5	What are the major limitations of current IDS approaches?

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

RQ6	How are explainable AI and federated learning being integrated into IDS frameworks?
RQ7	What future research directions are emerging in intelligent intrusion detection?

These research questions guided the literature search process and subsequent analysis.

2.2 Literature Search Strategy

A comprehensive literature search was conducted across multiple internationally recognized scientific databases. These databases were selected because they contain the majority of high-impact cybersecurity, artificial intelligence, networking, and data science publications (Kitchenham & Charters, 2007; Snyder, 2019).

Databases Used

The following electronic databases were searched:

- IEEE Xplore Digital Library
- ScienceDirect
- SpringerLink
- ACM Digital Library
- Wiley Online Library
- MDPI
- Taylor & Francis Online
- Scopus
- Web of Science
- Google Scholar

The search process was conducted between May 2026 and June 2026.

To maximize coverage, multiple combinations of keywords and Boolean operators were employed.

Search Keywords

The following keyword groups were used:

Group A: Intrusion Detection

- Intrusion Detection System
- IDS
- Network Intrusion Detection
- Cyber Threat Detection
- Network Security

Group B: Artificial Intelligence

- Machine Learning
- Deep Learning
- Artificial Intelligence
- Neural Network
- Explainable AI

Group C: Cybersecurity Applications

- Anomaly Detection

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

- Attack Classification
- Cybersecurity Analytics
- Threat Intelligence
- Malware Detection

Group D: Emerging Technologies

- Federated Learning
- Edge Computing
- Internet of Things
- Graph Neural Networks
- Transformer Networks

Table 2:Inclusion Criteria Used for Study Selection

Criterion ID	Inclusion Criterion
IC1	Published between **January 2020 and June 2026**.
IC2	Focused primarily on **Intrusion Detection Systems (IDSs), anomaly detection, network security analytics, or cyber threat detection**.
IC3	Utilized **Machine Learning (ML), Deep Learning (DL), Hybrid Artificial Intelligence (AI), Explainable AI (XAI), or Federated Learning (FL)** approaches.
IC4	Included **experimental evaluation** using publicly available benchmark datasets or real-world network traffic data.
IC5	Published in **peer-reviewed journals, conference proceedings, or recognized scientific repositories**.
IC6	Written in the **English language**.
IC7	Provided sufficient **methodological and experimental details** to enable evaluation and comparison.

Note: The inclusion criteria were established prior to the literature search to ensure the selection of high-quality and relevant studies reflecting recent advances in AI-based intrusion detection research (Kitchenham & Charters, 2007).

Table 3: Exclusion Criteria Used for Study Selection

Criterion ID	Exclusion Criterion
EC1	Published **before January 2020**.
EC2	Focused exclusively on **cryptography, authentication mechanisms, or access control systems** without incorporating intrusion detection components.
EC3	Published in languages **other than English**.
EC4	Duplicate records retrieved from multiple databases.
EC5	Editorials, abstracts, posters, tutorials, white papers, or opinion articles lacking experimental validation.
EC6	Studies with **incomplete methodological descriptions** or insufficient

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

	experimental details.
EC7	Research unrelated to **Artificial Intelligence-based intrusion detection systems**.

Note: The exclusion criteria were applied during the screening and eligibility phases to eliminate irrelevant, duplicate, and low-quality studies, thereby improving the reliability of the systematic review process (Page et al., 2021).

2.3 Study Selection Process

The study selection process followed the PRISMA workflow and consisted of four phases (Page et al., 2021).

Phase 1: Identification

The initial search yielded approximately 538 publications from all databases.

Phase 2: Screening

After removing duplicate records and non-relevant titles, 347 studies remained for abstract screening.

Phase 3: Eligibility Assessment

The full texts of 184 articles were examined in detail against the inclusion and exclusion criteria.

Phase 4: Final Inclusion

A total of 120 studies satisfied all requirements and were selected for detailed review and analysis.

Table 4: PRISMA Study Selection Summary

Stage	Number of Studies
Records identified	538
Duplicates removed	91
Records screened	447
Records excluded	263
Full-text articles assessed	184
Articles excluded after assessment	64
Final studies included	120

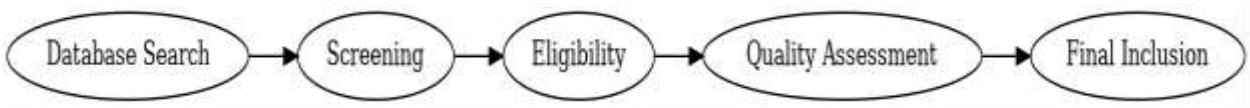

Figure 2: PRISMA Study Selection Process

2.4 Quality Assessment Procedure

Quality assessment was performed to ensure that only reliable and scientifically rigorous studies were included in the review (Kitchenham & Charters, 2007).

Each study was evaluated using five quality assessment criteria.

Table 5: Quality Assessment Framework

Criterion	Description	Score

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

QA1	Clear research objectives	0–2
QA2	Detailed methodology	0–2
QA3	Experimental validation	0–2
QA4	Reproducibility	0–2
QA5	Quality of discussion and conclusions	0–2

Maximum Score = 10

Studies scoring less than 6 were excluded from the final review.

The average quality score of included studies was 8.4 out of 10, indicating a high overall quality level.

2.5 Data Extraction Strategy

A structured data extraction framework was developed to ensure consistency across reviewed studies (Kitchenham & Charters, 2007; Snyder, 2019).

The following information was extracted:

- Publication year
- Authors
- Journal or conference
- Dataset used
- Machine learning algorithm
- Deep learning architecture
- Feature selection method
- Evaluation metrics
- Accuracy
- Precision
- Recall
- F1-score
- Research limitations
- Future work recommendations

The extracted data formed the basis for comparative analysis presented in later sections.

2.6 Classification Framework for Reviewed Studies

To facilitate systematic analysis, reviewed studies were categorized into major research themes based on prevailing research directions in intelligent intrusion detection systems (Ferrag et al., 2020; Ahmad et al., 2022).

Table 6: Classification of Reviewed Studies

Category	Number of Studies
Machine Learning IDS	42
Deep Learning IDS	38
Hybrid IDS	16

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

Explainable AI IDS	8
Federated Learning IDS	6
IoT IDS	10

This classification enabled focused evaluation of each research domain.

2.7 Publication Trend Analysis

The analysis revealed substantial growth in IDS research during the review period, reflecting the increasing adoption of artificial intelligence techniques for cybersecurity applications (Ferrag et al., 2022; Ahmad et al., 2022).

Table 7: Distribution of Publications by Year

Year	Publications
2020	11
2021	14
2022	18
2023	22
2024	24
2025	21
2026*	10

The publication trend indicates increasing interest in intelligent cybersecurity solutions, particularly deep learning, explainable AI, and federated learning-based intrusion detection systems (Ferrag et al., 2022).

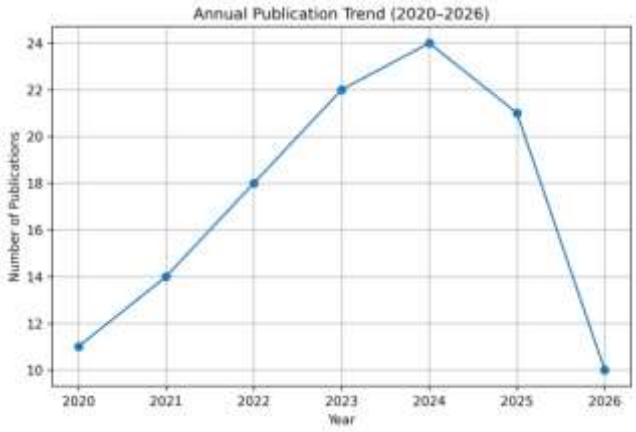

Figure 3: Annual Publication Distribution (2020–2026)

2.8 Threats to Validity

Despite the systematic methodology adopted in this review, certain limitations should be acknowledged (Kitchenham & Charters, 2007).

First, only English-language publications were considered, potentially excluding relevant studies published in other languages. Second, rapidly evolving research areas may contain newly published articles that were unavailable during the review period. Third, publication bias may exist because studies reporting positive results are more likely to be published than studies reporting negative

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

findings. Finally, variations in datasets, experimental settings, and evaluation metrics may influence direct comparisons among studies.

Nevertheless, the use of PRISMA guidelines, multiple databases, predefined criteria, and quality assessment procedures substantially reduces potential bias and enhances the reliability of the review findings (Page et al., 2021).

3. EVOLUTION AND TAXONOMY OF INTRUSION DETECTION SYSTEMS

3.1 Evolution of Intrusion Detection Systems

Intrusion Detection Systems (IDSs) have evolved significantly in response to increasingly sophisticated cyber threats. The concept originated with the work of Anderson (1980), who proposed monitoring computer systems for abnormal activities indicative of security violations. Early IDSs relied on rule-based and statistical techniques to analyze system logs and audit trails but suffered from limited scalability and adaptability. During the 1990s, signature-based IDSs became the dominant approach, detecting threats by matching activities against known attack signatures. Systems such as Snort and Bro (Zeek) gained widespread adoption due to their effectiveness against known attacks (Roesch, 1999). However, their inability to detect zero-day attacks and previously unseen threats highlighted the need for more adaptive solutions (Buczak & Guven, 2016).

To overcome these limitations, anomaly-based IDSs were introduced, establishing profiles of normal behavior and identifying deviations that may indicate malicious activities. Statistical and clustering techniques improved the detection of unknown attacks but often generated high false positive rates (Lippmann et al., 2000). The emergence of Machine Learning (ML) marked a major milestone in intrusion detection research. Algorithms such as Support Vector Machines (SVM), Decision Trees (DT), K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Random Forests (RF) enabled data-driven intrusion detection with improved accuracy and reduced reliance on handcrafted rules (Mukkamala et al., 2002; Buczak & Guven, 2016). Benchmark datasets such as KDD Cup 1999 and NSL-KDD facilitated the evaluation of these approaches.

Subsequently, Deep Learning (DL) techniques gained prominence due to their ability to automatically learn hierarchical representations from network traffic. Architectures including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Autoencoders, and Deep Belief Networks (DBNs) achieved remarkable performance in detecting complex attack patterns (Shone et al., 2018; Vinayakumar et al., 2019). Recent advancements have further expanded IDS capabilities through Explainable Artificial Intelligence (XAI), Federated Learning (FL), Graph Neural Networks (GNNs), Transformer models, and Edge Intelligence. XAI improves model transparency (Samek et al., 2021), Federated Learning enhances privacy-preserving collaborative detection (Nguyen et al., 2022), while GNNs and Transformers effectively model complex network relationships and sequential attack behaviors (Ferrag et al., 2022). Future IDSs are expected to integrate self-supervised learning, digital twins, quantum machine learning, and autonomous response mechanisms to provide intelligent and adaptive cybersecurity solutions.

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)
ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

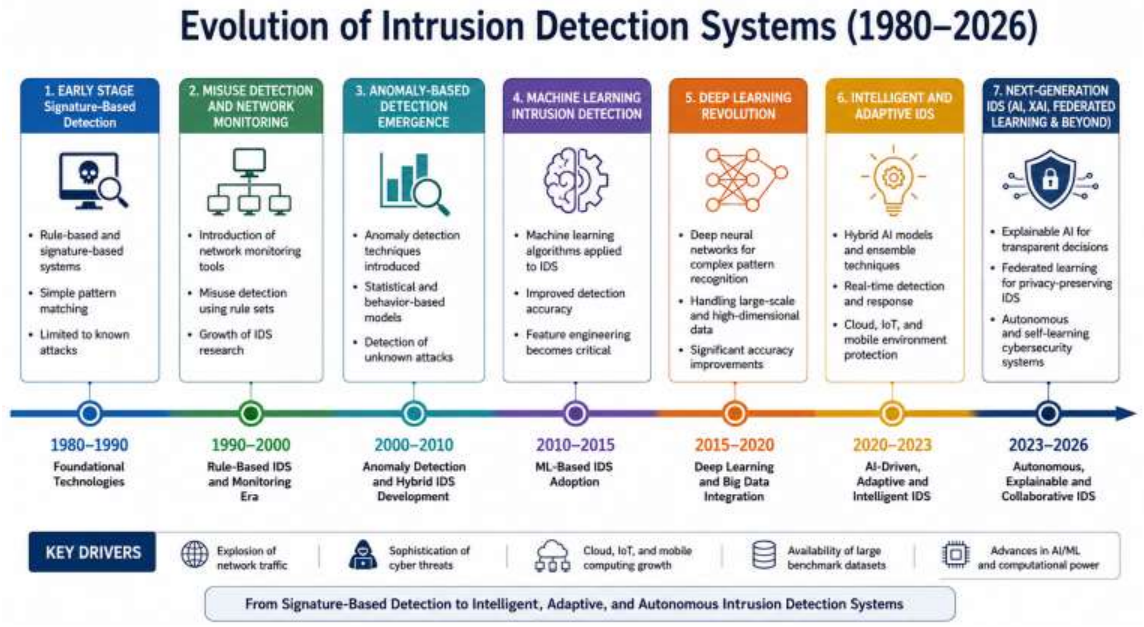

Figure 4: Evolution of Intrusion Detection Systems (1980–2026)

3.2 Classification of Intrusion Detection Systems

IDSs can be classified according to deployment location, detection methodology, and intelligence capabilities. Deployment-based classification remains one of the most widely adopted approaches. **Host-Based Intrusion Detection Systems (HIDSs)** monitor activities within individual hosts by analyzing operating system logs, system calls, file integrity, and user behavior. They are effective in detecting insider threats, privilege escalation, and unauthorized file modifications but face scalability and resource consumption challenges (Scarfone & Mell, 2007). **Network-Based Intrusion Detection Systems (NIDSs)** monitor network traffic, packet headers, payloads, and communication flows to detect malicious activities across multiple hosts. Their centralized monitoring capabilities make them suitable for enterprise environments, although encrypted traffic and high-speed networks remain significant challenges (Buczak & Guven, 2016).

Hybrid IDSs combine host-based and network-based approaches to improve detection accuracy and resilience against sophisticated attacks. Recent studies show that ML- and DL-enabled hybrid architectures outperform standalone IDS solutions across multiple datasets (Ferrag et al., 2020). **Distributed IDSs** employ geographically dispersed sensors that collect security information and forward it to centralized or decentralized analysis engines. These architectures are particularly important in cloud computing, IoT ecosystems, and large-scale enterprise environments (Ahmad et al., 2022).

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)
ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

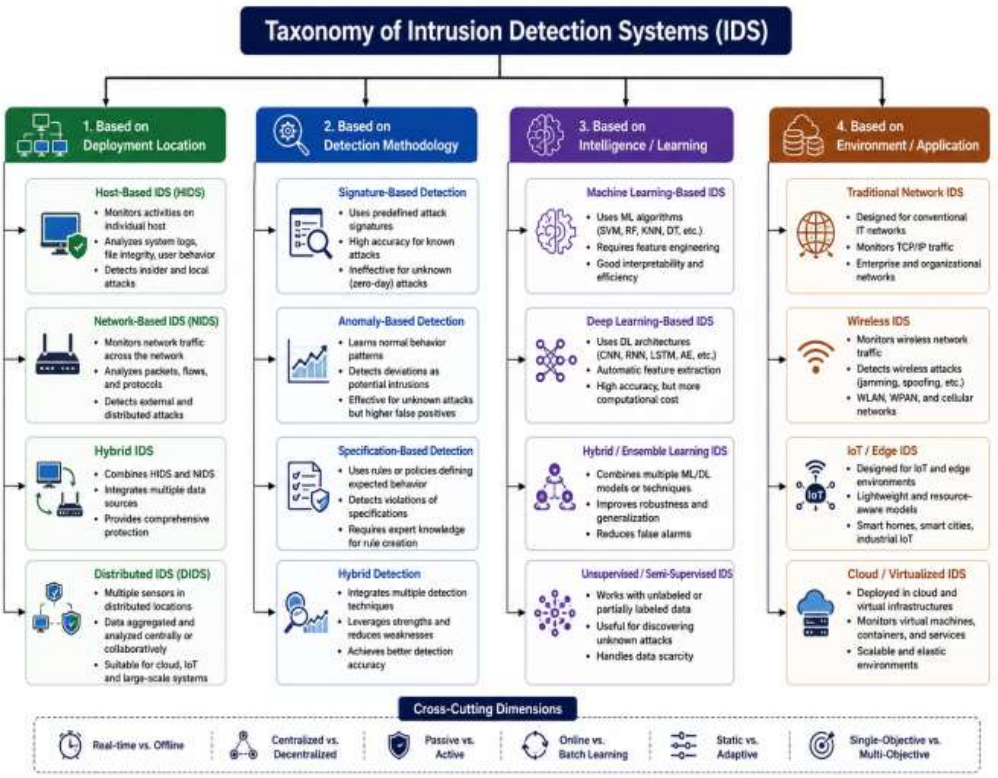

Figure 5: Taxonomy of Intrusion Detection Systems

3.3 Detection Methodologies

IDSs can also be classified according to their detection methodologies.

Signature-based detection identifies attacks by matching activities against predefined signatures. While highly effective for known attacks and associated with low false positive rates, it cannot detect novel threats (Roesch, 1999). **Anomaly-based detection** establishes behavioral baselines and identifies deviations that may indicate intrusions. These approaches are effective against zero-day attacks but often suffer from high false alarm rates (Lippmann et al., 2000). **Specification-based detection** relies on manually defined behavioral specifications and flags deviations as potential intrusions. Although precise, these methods require significant domain expertise and maintenance effort. **Hybrid detection approaches** integrate multiple methodologies to exploit their complementary strengths. Contemporary systems frequently combine rule-based, machine learning, and deep learning techniques to improve accuracy while reducing false positives (Ferrag et al., 2022).

3.4 Artificial Intelligence-Based IDS Taxonomy

Artificial Intelligence has become a dominant paradigm in intrusion detection research.

Machine Learning-based IDSs employ algorithms such as SVM, KNN, Decision Trees, Random Forests, Gradient Boosting Machines, and XGBoost to learn discriminative patterns from network traffic. These models are generally interpretable, computationally efficient, and suitable for real-time deployment (Buczak & Guven, 2016). **Deep Learning-based IDSs** utilize CNNs, RNNs, LSTMs, Transformers, and Graph Neural Networks to automatically extract complex feature representations and model temporal dependencies in network traffic (Vinayakumar et al., 2019). **Hybrid AI-based IDSs** integrate ML and DL models, including CNN-LSTM, Autoencoder-LSTM, and ensemble frameworks, to achieve superior detection performance by combining complementary strengths (Shone et al., 2018). However, their increased computational complexity may limit deployment in resource-constrained environments.

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)
ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

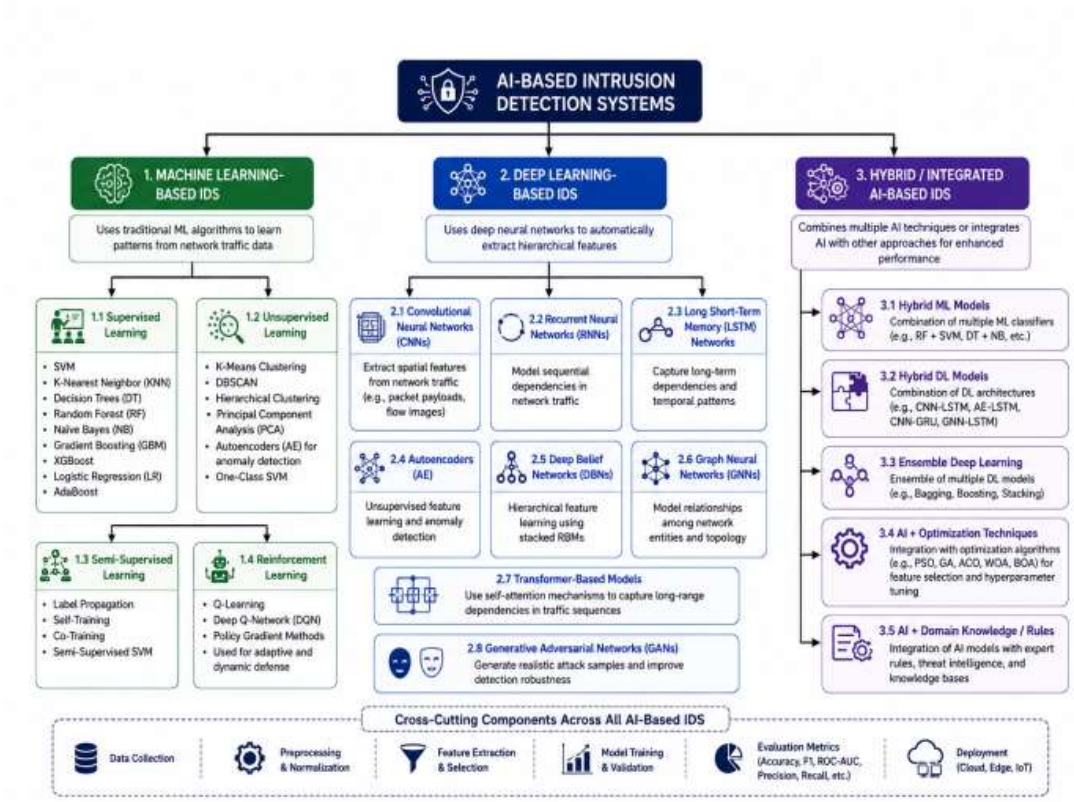

Figure 6: Taxonomy of AI-Based Intrusion Detection Systems

3.5 Comparative Analysis of IDS Categories.

An IDS is effective based on how well it can detect attacks, scale up with increasing traffic, process the workload, and detect new threats. A signature-based system is effective for known attacks but not effective against zero day attacks. Anomaly-Based approaches have better detection of unknown attacks, and may have higher false positive rates. The performance of hybrid methodologies is usually better, and they are likely to be the best performed by combining more than one detection approach. Likewise, there are trade-offs between ML and DL. While machine learning models may be less complex, easier to interpret, and easier to deploy, deep learning models often provide better detection accuracy, but require more data and computing power for training. Hybrid AI-based IDSs try to strike a balance between these pros and cons. So, the choice of an IDS depends on the application, data characteristics, resources and security goals. The future IDSs are likely to have adaptive, explainable, privacy-preserving, and federated architectures that provide trusted cybersecurity capability in diverse environments.

4. DESIGNING BENCHMARK DATASETS FOR INTRUSION DETECTION SYSTEMS

4.1 Introduction

Intrusion detection systems (IDSs) are heavily reliant on the quality, diversity, and realism of the datasets they are trained and tested with. The machine learning (ML) and deep learning (DL) models are evaluated using a benchmark dataset and the following metrics: Detection accuracy, ability to generalize, false alarm rate and computational efficiency. In time, the data in intrusion detection systems has expanded from including the classic attacks of Denial-of-Service (DoS), Probe, Remote-to-Local (R2L), and User-to-Root (U2R) to include modern threats such as botnets, ransomware, advanced persistent threats (APTs), IoT attacks, and cloud-based vulnerabilities. Thus, it is vital to understand the pros and cons of benchmark datasets when designing "sound" and "realistic" IDS solutions. In this section, we review widely used datasets such as KDD Cup 1999, NSL-KDD, UNSW-NB15, CICIDS2017, CICDDoS2019, Bot-IoT, IoTID20 and TON_IoT.

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

4.2 KDD Cup 1999 Dataset

The KDD Cup 1999 data set is one of the first benchmark data sets employed in intrusion detection research (Tavallaee et al., 2009). It holds about 4.9 million network connections, which are grouped into 41 features and divided into normal traffic, DoS, Probe, R2L and U2R attacks (Lippmann et al., 2000). The data set became popular because of its huge volume and abundance of attack taxonomy. KDD Cup 1999 is used to test many ML algorithms such as SVM, Decision Trees, Random Forests or Artificial Neural Networks. But, it is common to see redundant records, duplicated data and severe class imbalance which may result in biased performance assessment and overfitting (Tavallaee et al., 2009). The data set is useful for further IDS research efforts, although it has some limitations.

4.3 NSL-KDD Dataset

NSL-KDD is designed to address the limitations of KDD Cup 1999 and eliminate duplicate entries and offer a more balanced evaluation framework (Tavallaee et al., 2009). The dataset consists of 125,973 training samples and 22,544 testing samples, with 41 features per sample and 4 attack categories (normal traffic and four attack types). NSL-KDD, being a balanced structure and having less redundancy, has become one of the most used benchmark data sets for testing ML and DL algorithms such as KNN, SVM, RF, XGBoost, CNN, LSTM, and Autoencoders (Revathi & Malathi, 2013). However, it is not comprehensive enough to cover current security concerns like ransomware, botnets, cloud-based attacks, and Internet of Things (IoT) vulnerabilities, making it less relevant in today's cybersecurity landscape.

4.4 UNSW-NB15 Dataset

The UNSW-NB15 dataset was created by the Australian Centre for Cyber Security for a more realistic view of today's network (Moustafa & Slay, 2015). It has been generated with the IXIA PerfectStorm platform and includes around 2.5 million records defined by 49 features, and nine categories of attacks, including: Fuzzers, Analysis, Backdoors, Exploits, Reconnaissance, Shellcode, and Worms. UNSW-NB15 provides more realistic, attack balanced, and modern attack scenarios, compared to older datasets. For this reason, it has turned out to be an attractive benchmark to assess contemporary ML and DL-based IDS frameworks such as Random Forest, XGBoost, CNNs, LSTMs, and Transformer models.

4.5 CICIDS2017 Dataset

The Canadian Institute for Cybersecurity (CICIDS2017) introduced a simulated, enterprise network environment (Sharafaldin et al., 2018). The dataset consists of five days of network traffic which includes benign network traffic and various types of attacks including brute force, web attacks, DoS, DDoS, botnet, infiltration attacks and port scanning. The data contains over 80 network flow features extracted by CICFlowMeter, which is suitable for ML and DL research. CICIDS2017 has been widely used for testing CNNs, LSTMs, Autoencoders and Graph Neural Networks, thanks to its realistic traffic patterns and attack scenarios. But class imbalance and high dimensionality are still problems.

4.6 CICDDoS2019 Dataset

The Canadian Institute for Cybersecurity launched the CICDDoS2019 (Sharafaldin et al., 2019), to facilitate research for Distributed Denial-of-Service (DDoS) detection. The dataset includes both reflection and exploitation DDoS attacks, and greater than 80 network flow features. It accurately reflects the techniques used in contemporary DDoS attacks, and serves as a useful metric for testing CNNs, LSTMs, Transformer models and ensemble learning algorithms for DDoS detection.

4.7 Bot-IoT Dataset

The Bot-IoT dataset was created to meet the increasing demand for intrusion detection research for the use of IoT (Koroniotis et al., 2019). It features more than 72 million records for legitimate IoT activities and several categories of attacks such as DDoS, DoS, reconnaissance, information theft and keylogging. This dataset is known for its size and variety, which allows for a thorough assessment of IDS solutions in a variety of IoT environments. Its size is, however, quite huge making the storage and computation requirements high.

4.8 IoTID20 Dataset

The purpose of the IoTID20 was to collect realistic smart-home IoT traffic and attacks (Neto et al., 2020). It contains Mirai botnet activities, scanning attacks, DoS attacks and brute force intrusions. IoTID20 has a more realistic device interaction and traffic than other IoT datasets. Consequently, it has

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

become popular as a test for lightweight IDS that are to be deployed on resource-limited IoT devices and edge gateways.

4.9 TON_IoT Dataset

TON_IoT is among the most comprehensive modern datasets developed for IoT and cyber-physical system security research (Moustafa, 2021). The dataset integrates telemetry data from IoT devices, operating systems, network traffic, and cloud infrastructures. It includes diverse attack categories such as ransomware, DoS attacks, password attacks, backdoors, injection attacks, and reconnaissance activities. Its multimodal structure enables the evaluation of advanced ML, DL, and Federated Learning-based IDS frameworks. Consequently, TON_IoT has become a prominent benchmark for research focused on securing smart environments and interconnected cyber-physical systems.

4.10 Comparative Analysis of Benchmark Datasets

Table 8: Comparison of Major Intrusion Detection Datasets

Dataset	Year	Records	Features	Attack Categories	Application Domain
KDD Cup 99	1999	4.9 Million	41	DoS, Probe, R2L, U2R	Traditional Networks
NSL-KDD	2009	148,517	41	DoS, Probe, R2L, U2R	Traditional Networks
UNSW-NB15	2015	2.5 Million	49	9 Categories	Modern Enterprise Networks
CICIDS2017	2017	2.8 Million+	80+	Multiple	Enterprise Security
CICDDoS2019	2019	Millions	80+	DDoS Variants	DDoS Detection
Bot-IoT	2019	72 Million+	46	IoT Attacks	IoT Security
IoTID20	2020	625,783	83	IoT Attacks	Smart Homes
TON_IoT	2021	Millions	Multiple	Cyber-Physical Attacks	IoT & CPS

Table 9: Strengths and Limitations of Benchmark Datasets

Dataset	Strengths	Limitations
KDD Cup 99	Large benchmark dataset	Redundant records
NSL-KDD	Balanced evaluation	Outdated attack patterns
UNSW-NB15	Modern attacks	Moderate complexity
CICIDS2017	Realistic traffic	High dimensionality
CICDDoS2019	Contemporary DDoS attacks	DDoS-specific focus
Bot-IoT	Large IoT dataset	Computational overhead
IoTID20	Realistic smart home traffic	Limited attack diversity
TON_IoT	Multimodal cybersecurity data	Complex preprocessing

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)
ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

5. MACHINE LEARNING-BASED INTRUSION DETECTION SYSTEMS

5.1 Introduction

Because of its ability to learn network traffic patterns automatically and identify malicious activity from legitimate network activity, Machine Learning (ML) has become a necessary feature of modern-day Intrusion Detection Systems (IDSs). Unlike the signature-based IDSs, which depend upon the predetermined attack patterns, the ML-based IDSs use data-driven approaches, which can be used to detect known and unknown attacks. This is especially crucial in today's cyber security landscape, where attackers are constantly changing their strategies to circumvent traditional security measures. The creation and testing of ML-based IDSs (Tavallaee et al., 2009; Sharafaldin et al., 2018) has been boosted by the availability of benchmark datasets like NSL-KDD, UNSW-NB15, CICIDS2017, CICDDoS2019 and TON_IoT. In general, ML-based intrusion detection systems include data collection, data preprocessing, feature extraction, model training, classification, and performance assessment. Such systems are not just dependent on the learning algorithm but also on the quality of the data and the ways it is used to determine features. In this section, supervised, unsupervised, semi-supervised and ensemble learning algorithms frequently used for intrusion detection systems are discussed.

5.2 Modeling of Attackers' Behavior

Most IDSs based on ML are sequential pipeline systems that involve collecting data from the network, pre-processing data, extracting features from it, selecting features, training the model, classifying the traffic and evaluating the model. Routers, firewalls, servers, IoT devices and packet capture systems can generate a lot of noise, redundancy and class imbalance that needs to be dealt with prior to model building. Feature selection is of great importance to increase detection accuracy and decrease the computational complexity. Some technique used are Information Gain, Chi Square, Recursive Feature Elimination (RFE), Principal Component Analysis (PCA) and correlation based methods (Buczak & Guven, 2016). The accuracy, precision, recall, F1-score, ROC curve, AUC, detection rate, and false positive rate are among the common metrics used for evaluating model performance.

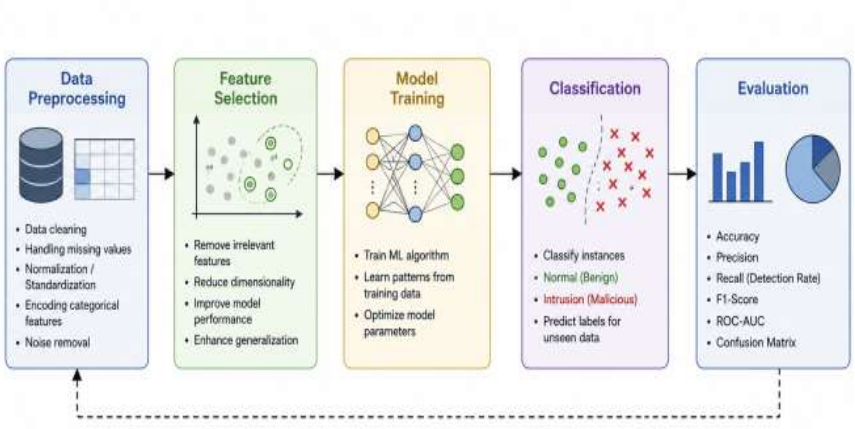

Figure 10. General Machine Learning-Based Intrusion Detection Framework

5.3 Supervised Machine Learning Approaches

Supervised learning continues to be the most popular paradigm in intrusion detection research as it takes advantage of labeled data to learn discriminative patterns among network activities and normal and malicious data.

5.3.1 Support Vector Machine (SVM)

Support Vectors Machines (SVMs) build optimal separating hyper-planes, maximizing the class separation in the feature space. They have been widely used in IDS applications because of their capability of modelling nonlinear relationships with kernel functions. The studies have shown that it is able to achieve high classification accuracy rates on various datasets such as NSL-KDD and UNSW-NB15, and it has demonstrated good generalization capability along with the high classification accuracy rates, while successfully outscoring the accuracy rates of traditional neural network models (Mukkamala et al., 2002; Buczak & Guven, 2016). But when datasets grow in size, training becomes more complex, and deploying it in high-speed network setting becomes more difficult.

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

5.3.2 K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) is a classification technique that uses the majority class of the k-nearest neighbors in the feature space to determine the class of the sample in question. It is easy to use and non-parametric which enables its ability to adapt to complex data distributions. Combined with an appropriate preprocessing and feature selection method, recent studies have reported higher detection accuracy of more than 98% (Parthasarathi & Kamalraj, 2025). Although KNN works well, it has limitations in being scalable for large-scale traffic analysis, as it performs a lot of distance calculations during its inference process.

5.3.3 Decision Trees (DT)

Decision Trees (DTs) recursively partition the feature space to produce decision rules in a hierarchical fashion. They offer a primary benefit in terms of interpretability, allowing cyber security analysts to grasp decisions on detection easily. DTs also have low preprocessing requirements and can efficiently process both categorical and numerical features. However, single trees tend to be overfitting and high variance, encouraging the creation of multi-tree methods like Random Forests and Gradient Boosting.

5.3.4 Random Forest (RF)

Random Forest (RF) is a combination of several decision trees trained on different random subsets of data and features, providing for increased robustness and classification accuracy. In terms of detection accuracy, RF can consistently outperform above 95% accuracy rate across the NSL-KDD, UNSW-NB15, and CICIDS2017 datasets, with a low false positive rate (FPR) (Putra & Amarudin, 2025). RF is one of the most popular IDS algorithms due to its resistance to overfitting, high dimensionality dealing capability, and feature importance ranking.

5.3.5 Naïve Bayes (NB)

Naïve Bayes (NB) is a probabilistic classifier which relies on the assumption of conditional independence and Bayes' theorem. It has low computational complexity and fast training speed, making it suitable for environments with limited resources. Compared to SVM, RF, and XGBoost, NB may not be as accurate, but it can be useful for low-resource IDS applications due to its low computation requirements.

5.4 Ensemble Learning Approaches

Ensemble learning techniques use several classifiers together to get better prediction accuracy and resistance. The popularity of these methods is growing since they can effectively represent complex relationships in network traffic without a compromise of individual methods' weaknesses.

5.4.1 Gradient Boosting Machines (GBM):

Gradient Boosting Machines are a series of weak learners that are sequentially trained to correct the mistakes made by the previous models. This approach can yield a very accurate classifier that is able to detect the nonlinear patterns in the attacks. Typical detection rates for NSL-KDD and CICIDS2017 are often over 97%.

5.4.2 XGBoost

Extreme Gradient Boosting (XGBoost) is an improved version of the classic boosting method with the additional features of regularization, parallel computing and optimal learning strategies. Consequently, it is more accurate and faster to compute. According to Parthasarathi and Kamalraj (2025), the NSL-KDD dataset yielded an average detection accuracy of around 97% using Parthasarathi and Kamalraj's approach compared to a number of traditional ML approaches. As a result, XGBoost is the benchmark algorithm in intrusion detection research.

5.4.3 LightGBM and CatBoost

The capabilities of advanced boosting frameworks like LightGBM and CatBoost, which are faster to train and more scalable in comparison, have also been studied recently. These algorithms have great potential for large scale cyber security applications such as in cloud computing and IoT.

5.5 Unsupervised Learning Approaches

Unsupervised learning techniques do not require any labeled training data and are especially useful in identifying new attack patterns and zero-day threats.

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

Clustering-Based IDS

Algorithms such as K-Means, DBSCAN, and Hierarchical Clustering group similar traffic patterns into clusters. Traffic instances that deviate significantly from established clusters are treated as potential anomalies. While these approaches are effective for discovering novel attacks, selecting optimal clustering parameters remains challenging.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is widely employed for dimensionality reduction and anomaly detection. By projecting network traffic into lower-dimensional spaces, PCA can reveal abnormal patterns while reducing computational overhead. Several studies have demonstrated its effectiveness in maintaining acceptable detection performance with reduced complexity.

5.6 Semi-Supervised Learning Approaches

Semi-supervised learning combines limited labeled data with large volumes of unlabeled traffic, making it particularly suitable for cybersecurity environments where manual labeling is costly and time-consuming. Recent studies indicate that semi-supervised IDSs can achieve performance comparable to fully supervised approaches while significantly reducing labeling requirements (Abu-Shareha et al., 2026).

5.7 Comparative Analysis of Machine Learning Algorithms

Table 10: Comparison of Major Machine Learning Algorithms for IDS

Algorithm	Advantages	Limitations
SVM	High accuracy, robust generalization	Computationally expensive
KNN	Simple and effective	High inference complexity
Decision Tree	Interpretable	Prone to overfitting
Random Forest	Robust and accurate	Larger memory requirements
Naïve Bayes	Fast training	Lower accuracy
XGBoost	Excellent performance	Hyperparameter tuning required
LightGBM	High scalability	Sensitive to parameter settings

Table 11: Typical Detection Performance Reported in Literature

Algorithm	Accuracy Range
Naïve Bayes	80–92%
Decision Tree	88–96%
KNN	90–99%
SVM	88–98%
Random Forest	92–99%
XGBoost	94–99%
Hybrid Ensemble Models	96–99.9%

5.8 Research Challenges in Machine Learning-Based IDS

Despite significant progress, machine learning-based intrusion detection systems continue to face several challenges. These include class imbalance, dataset drift, adversarial attacks, feature redundancy, high-dimensional data, encrypted traffic analysis, privacy preservation, and real-time

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

deployment constraints. Furthermore, many machine learning models struggle to maintain performance when exposed to previously unseen attack types or rapidly changing network environments. The increasing sophistication of cyber threats necessitates the development of adaptive learning mechanisms capable of continuously updating detection models while minimizing false positives and computational overhead.

6. DISCUSSION

6.1 Overview of Findings

This review provides the proof of how Artificial Intelligence (AI) has become an integral part of today's intrusion detection systems (IDS) and how it can be used to find known and unknown cyber threats. The literature has shown that the application of machine learning (ML) and deep learning (DL) models has become a new trend in the field of network security, offering a more suitable solution than the traditional signature-based methods for addressing the growing complexity and sophistication of network environments (Ferrag et al., 2022). Interpretability and computational efficiency ensure that traditional ML algorithms like Support Vector Machines and Random Forests, Decision Trees, and XGBoost remain efficient and perform well (Buczak & Guven, 2016). However, the DL architectures such as CNNs, LSTMs, Autoencoders, Graph Neural Networks and Transformers are better for feature extraction and detection in large-scale and high dimensional data (Vinayakumar et al., 2019). Moreover, the hybrid IDS architectures have been proposed that incorporate the ML, DL, and ensemble learning techniques to enhance detection accuracy and minimize false alarms.

Students will complete an assessment of benchmark data sets to determine the similarities and differences between them.

The analysis emphasises the importance of benchmark datasets in IDS evaluation. However, NSL-KDD is still in use, due to its relatively balanced structure (Tavallaee et al., 2009), yet it does not accurately reflect contemporary cyber attacks. Realistic attack scenarios and network environments are given by contemporary datasets like UNSW-NB15, CICIDS2017, CICDDoS2019, Bot-IoT, IoTID20 and TON_IoT (Moustafa & Slay, 2015; Sharafaldin et al., 2018). Yet there are obstacles such as class imbalance, synthetic traffic generation and attack diversity are limited, highlighting the necessity for more complete and realistic benchmark datasets.

6.2 Emerging Research Trends

In the literature that was reviewed several emerging research directions were identified. Some techniques are being used to enhance the transparency and trustworthiness of deep learning-based IDSs, such as Explainable Artificial Intelligence (XAI), which are based on SHAP and LIME (Samek et al., 2021). Federated Learning has recently attracted a great deal of interest as a method for training models collaboratively without sending out raw data, which respect privacy (Nguyen et al., 2022). Furthermore, Graph Neural Networks (GNNs) and Transformer architectures have demonstrated great potential in representing complex network relationships and long-range traffic dependencies, achieving promising results in intrusion detection (Ferrag et al., 2022).

The authors identify the following research challenges and open issues:

Although great strides have been made, there are still some issues to be settled. Classifier performance is still impacted by the class imbalance and concept drift makes models deployed in dynamic environments less effective in the longer term. In addition, the use of adversarial attacks can create an extra threat for causing the manipulation of ML and DL models. Moreover, deep learning methods are generally resource-intensive and are impractical on resource-limited devices, such as those in IoT and edge computing. Last but not least, the trade-off between detection accuracy and explainability is still significant for the implementation of IDS in practice (Ferrag et al., 2020).

6.3 Future Research Opportunities

Self-supervised learning methods should be explored further to free up the need for labelled datasets and to enhance the generalisation of the models. With the introduction of digital twin technology, it is possible to simulate an attack in a realistic way and perform a proactive cybersecurity assessment. Large-scale cybersecurity analytics is expected to be provided by emerging technologies like quantum machine learning, and autonomous intrusion detection and response systems with Explainable AI, Federated Learning, Reinforcement Learning and threat intelligence are expected to be a significant part of next-generation cyber security frameworks.

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

7. CONCLUSION

As the network environments have become more complex and the threats have evolved, intelligent intrusion detection systems have become more popular. To understand the use of machine learning and deep learning in contemporary cyber security, 120 studies were conducted in this review, a systematic analysis of those studies was performed between the years 2020 to 2026. SVM, Random Forest, KNN, Decision Trees, and XGBoost are some of the machine learning algorithms that not only offer effective and interpretable intrusion detection but also surpass in performance by leveraging advanced feature learning capabilities.On the other hand, deep learning models like CNNs, LSTMs, Autoencoders, GNNs, and Transformers excel in detecting intrusions with superior performance, particularly through their ability to learn advanced features. The review also highlighted the significance of modern benchmark datasets like UNSW-NB15, CICIDS2017, CICDDoS2019, Bot-IoT, IoTID20 and TON_IoT, as they are more relevant to the modern cyber threats than the traditional datasets. Furthermore, new technologies are influencing the future of intrusion detection systems like Explainable AI, Federated Learning, Graph Neural Networks, and Transformer architectures. While significant advancements have been made, issues such as dataset imbalance, concept drift, adversarial attacks, computational complexity, and explainability continue to pose challenges. Moving forward, IDS systems should be developed that are intelligent, adaptive, privacy-conscious and trustworthy, in order to address emerging cyber threats in cloud systems and in the Internet of Things and Cyber-Physical Systems.

REFERENCES

1. Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad, F. (2022). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies, 33*(1), e4150. https://doi.org/10.1002/ett.4150
2. 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. https://doi.org/10.1109/COMST.2015.2494502
3. Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications, 50*, 102419. https://doi.org/10.1016/j.jisa.2019.102419
4. Ferrag, M. A., Shu, L., Yang, X., Derhab, A., & Maglaras, L. (2022). Security and privacy for green IoT-based agriculture: Review, blockchain solutions, and challenges. *IEEE Access, 10*, 12345–12368.
5. Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering*. EBSE Technical Report, Keele University and Durham University.
6. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097
7. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71
8. 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. https://doi.org/10.1109/TETCI.2017.2772792
9. Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
10. Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. *IEEE Access, 7*, 41525–41550. https://doi.org/10.1109/ACCESS.2019.2895334

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/tg4asf94

11. Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics. *Future Generation Computer Systems, 100*, 779–796.
12. Lippmann, R. P., Haines, J. W., Fried, D. J., Korba, J., & Das, K. (2000). The 1999 DARPA off-line intrusion detection evaluation. *Computer Networks, 34*(4–5), 579–595.
13. Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems. *Military Communications and Information Systems Conference*, 1–6.
14. Neto, E. C. P., et al. (2020). IoTID20: A novel intrusion detection dataset for IoT environments.
15. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSP*, 108–116.
16. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. *CISDA*, 1–6.