International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 5, September-October 2025||

DOI:10.15662/IJARCST.2025.0805003

AI-Based Early Detection of Cyber Attacks using Network Traffic Analysis – A Comprehensive Review

Bhavana B R, Karthik Rajesh Shet, Kavana A R

Assistant Professor, Department of M.C.A, Surana College (Autonomous), Kengeri, Bangalore, India

PG Student, Department of M.C.A, Surana College (Autonomous), Kengeri, Bangalore, India

PG Student, Department of M.C.A, Surana College (Autonomous), Kengeri, Bangalore, India

ABSTRACT: Cyberattacks have dramatically increased in recent years, exposing the flaws in traditional intrusion detection systems (IDS) that rely on static signatures or flimsy anomaly models. Conventional techniques have a high false positive rate and struggle to detect novel, polymorphic, or multi-stage attacks. Artificial intelligence (AI) has developed into a powerful enabler of data-driven and adaptive network defines in order to get around these restrictions. ML, DL, RL, and GNNs are used by AI-powered IDS to automatically extract features from data, analyse complex traffic patterns, and identify minute irregularities that could indicate malicious activity. This review compiles research from 2023 to 2025 that critically assesses hybrid ensembles, supervised and unsupervised frameworks, and novel concepts like explainable and federated learning. Comparative experiments show that AI-based IDS perform with greater accuracy and flexibility on various datasets, but issues persist around dataset imbalance, scalability, adversarial robustness, and interpretability. Industrial applications are examined in enterprise networks, cloud systems, Internet of Things (IoT) environments, and real-time monitoring. Future directions are also pointed out that can influence resilient, trustworthy, and scalable IDS for future cybersecurity.

KEYWORDS: Cybersecurity, AI, Anomaly Detection, Intrusion Detection, Deep Learning.

I. INTRODUCTION

Cybersecurity has emerged as the backbone of contemporary society as digital technologies envelop virtually every aspect of human existence. The expansion of high-speed internet, cloud computing, 5G networks, and the Internet of Things (IoT) has enabled an interconnected world where billions of devices are constantly interacting and exchanging data. Although this digital revolution fuels innovation and drives economic growth, it also enlarges the attack surface for adversaries. Recent estimates place global cybercrime damages at over USD 10 trillion per year by 2025, one of the biggest threats facing digital economies and national security [1]. Activity spans from ransomware attacks on critical infrastructure to highly developed advanced persistent threats (APTs) that lie undetected for months in enterprise networks.

Intrusion Detection Systems (IDS) have long been frontline defences for cybersecurity. They observe and inspect network traffic or system activity to identify evidence of malicious activity. Conventional IDS are divided into two types: signature-based systems, which recognize threats through the comparison of traffic against known attack signatures, and anomaly-based systems, which create baselines of "normal" operation and alert when they are broken [2]. Signature-based IDS are effective against known attacks but do not identify new exploits or polymorphic malware. Conversely, anomaly-based IDS in theory can identify zero-day threats but are infamous for producing excessive false positives and tend to overwhelm security analysts [3].

As multi-step intrusions, polymorphic malware, and encrypted traffic have become prevalent, legacy IDS strategies become less and less sufficient. For example, in Software-Defined Networks (SDNs) and IoT environments, existing systems are not capable of scaling to accommodate diversity and speed of data flows [4], [6]. Additionally, attackers are now more commonly using AI themselves to create evasion methods, which further obscures detection [5]. These facts make a compelling case for the need for smart IDS solutions that can adapt, scale, and be resilient to changing threats.

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 5, September-October 2025||

DOI:10.15662/IJARCST.2025.0805003

New computational techniques are leading profound shifts in how we identify threats, and security systems become more adaptive and responsive. IDS powered by AI can be taught to recognize intricate patterns of normal and attack behaviour directly from information, allowing subtle anomalies to be discovered that are undetectable to rule-based systems. Machine Learning (ML) algorithms, including decision trees, random forests, and support vector machines, were among the initial ones to be used for intrusion detection [5], [8]. They, though, are feature-engineering dependent and do not perform well with high-dimensional, non-linear patterns. Deep Learning (DL) avoids these drawbacks by learning hierarchies of features from raw traffic automatically. Architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have shown significant success in identifying distributed denial-of-service (DDoS) attacks, botnets, and insider threats [3], [9], [10]. Additionally, hybrid models integrating CNNs with LSTMs or BiLSTMs are able to recognize both spatial and temporal features, greatly enhancing detection accuracy [8].

Recent developments have taken IDS beyond supervised learning. Reinforcement Learning (RL) enables IDS agents to learn adaptive policies in interaction with network environments, allowing for the flexibility to react to unknown threats [7], [14]. Graph Neural Networks (GNNs) have also become popular since they model the communication patterns in graphs whose nodes represent hosts and edges represent traffic flows. This enables the identification of coordinated attacks and lateral mobility in enterprise networks [12], [32]. In the meantime, federated learning has been proposed to handle privacy issues by training IDS models jointly in distributed devices without centralising sensitive traffic information [11], [21].

These advances notwithstanding, there are a number of challenges involved. One significant challenge is dataset imbalance, where malicious samples are rare with respect to normal traffic. This results in biased models that misclassify infrequent but high-value zero-day attacks [13], [16]. Further, most IDS models are tested against static offline datasets like CIC-IDS2017 or UNSW-NB15, which, although good benchmarks, do not accurately reflect the dynamics of real-time traffic on the cloud and IoT [19]. False positives still dog AI-powered IDS, as anomaly models tend to confuse legitimate unusual activity with malicious activity [5]. Scalability is also a problem: applying computationally intensive deep models to real-time enterprise or IoT environments means trading off detection accuracy against latency and computational cost [9], [24]. Moreover, AI-powered IDS themselves are susceptible to adversarial attacks, where an attacker prepares inputs intended to deceive detection models [25].

The other significant barrier is the explainability. Deep IDS are largely black boxes, and it is challenging to interpret why a specific alert was generated. This discourages trust in automated systems and makes it harder to integrate them into security operations centres (SOCs) [26], [31]. Researchers are now investigating explainable AI (XAI) techniques like LIME, SHAP, and attention visualisation to overcome this gap [26]. Lastly, data privacy issues create challenges, particularly when training IDS on sensitive traffic logs. IDS solutions that are federated and blockchain-based are being explored to improve both trust and privacy in collaborative defence [19], [22].

By combining recent developments, this review offers a current reference for MCA students, researchers, and practitioners interested in learning how AI can benefit early intrusion detection using network traffic analysis. The rest of the paper is organised as follows: Section 2 summaries conventional and AI-driven IDS methods; Section 3 investigates selected AI methods used in intrusion detection; Section 4 addresses cross-domain applications; Section 5 summaries challenges and limitations; Section 6 discusses future directions; and Section 7 concludes with final thoughts.

A. Procedure for Choosing Papers

In order to maintain relevance and quality, this review took a systematic approach when it came to choosing papers. Digital libraries like IEEE Explore, SpringerLink, ScienceDirect, ACM Digital Library, MDPI, and arXiv were queried for studies from the period 2023-2025. Keywords used were "AI-based intrusion detection," "network traffic analysis," "deep learning IDS," "federated learning," and "IoT intrusion detection." Initially, the search yielded about 86 papers. Screening of titles and abstracts narrowed it down to 56 papers. A full-text evaluation was then performed, concentrating on methods, datasets (e.g., UNSW-NB15, CIC-IDS2017, Bot-IoT), and performance metrics. High-impact journals and well-known conferences were prioritised. Last, 32 papers were selected, discussing machine learning, deep learning, reinforcement learning, graph neural networks, and hybrid models. A systematic approach allowed the review to demonstrate current, diverse, and high-quality trends in research.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 5, September-October 2025||

DOI:10.15662/IJARCST.2025.0805003

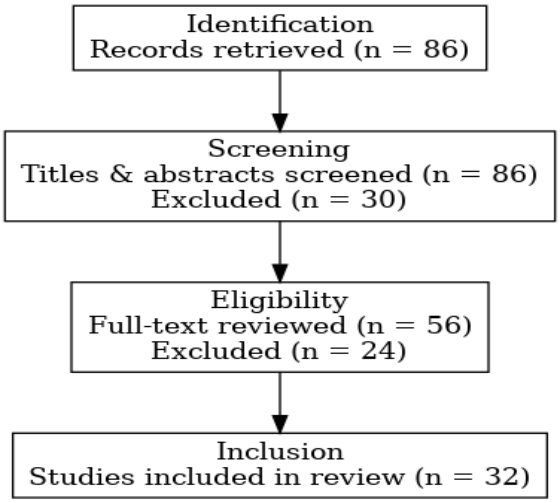

Fig 1. Procedure for Selecting Papers

II. BACKGROUND AND LITERATURE REVIEW

Intrusion detection has come a long way since its beginnings in the late 1980s, yet the root motivation is still the same: to detect and react to malicious activity before it is able to erode system integrity. Older methods, based on fixed rule-sets and superficial statistical models, have served a vital historical purpose in protecting networks. But as attacks have become more sophisticated, so too has the requirement for intelligent and adaptive approaches. This chapter summaries conventional IDS paradigms, discusses how machine learning and deep learning are being applied to intrusion detection, and looks at more recent developments like reinforcement learning, graph-based detection, and federated learning.

A. Conventional IDS Paradigms

Early IDS implementations were classified under two categories: signature-based and anomaly-based systems. Signature-based IDS work by matching incoming traffic with a database of known attack signatures. Though effective at detecting previously catalogued exploits, such systems are ineffective against zero-day attacks and necessitate frequent manual updates to remain effective [2].

Anomaly-based IDS, on the other hand, try to recognise deviations from a learned baseline of "normal" behaviour. Theoretically, this can enable the identification of new threats. Practically, anomaly-based IDS tend to produce high rates of false positives, since normal but atypical behaviour—such as an unexpected spike in traffic because of software updates—is detected as malicious [3]. Research consistently demonstrates that false positives remain one of the most serious impediments to practical use of anomaly-based IDS [5].

Hybrid approaches that integrate signatures and anomalies have been presented as an intermediary solution, allowing for increased coverage and improved adaptability. For instance, Mahmoud et al. [14] introduced a reinforcement learning-based anomaly module combined with a traditional signature engine for adaptive detection. However, hybrid approaches based on earlier paradigms are not enough for contemporary threat landscape, encompassing polymorphic malware, advanced persistent threats (APTs), and coordinated IoT-based botnets.

B. Machine Learning in Intrusion Detection

The use of machine learning (ML) in IDS was a paradigm shift. Traditional supervised ML models like decision trees, support vector machines (SVM), k-nearest neighbours (KNN), and random forests have shown promising performance when trained on benchmark intrusion detection datasets [5], [8]. These algorithms learn decision boundaries for benign and malicious traffic by examining features like packet sizes, flow duration, and port numbers.

As an example, Hassan et al. [5] compared ML classifiers in IDS and pointed out that random forests are always better than simpler classifiers because they can capture high-order feature interactions. Likewise, Sharma and Kim [8] proposed a hybrid CNN–BiLSTM IDS but also compared conventional ML approaches, commenting that ensembles

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 5, September-October 2025||

DOI:10.15662/IJARCST.2025.0805003

like random forests and gradient boosting are still competitive, especially when computational complexity is sought to be minimised.

Unsupervised ML, i.e., clustering algorithms like k-means and density-based spatial clustering (DBSCAN), have been used when there is no labeled data. Although powerful for outlier detection, they are parameter-sensitive and tend to yield inconsistent results in high-dimensional spaces [16].

Though they offer advantages, ML-based IDS have some major drawbacks. Feature engineering is necessary for them, wherein the professionals manually choose traffic attributes, which may be error-prone and time-consuming. In addition, concept drift—i.e., changes in patterns of normal traffic over time—can reduce model precision [16]. These issues prompted the use of deep learning.

C. Deep Learning for Intrusion Detection

Deep learning (DL) transformed IDS research by enabling automatic extraction of features from raw network traffic data. Instead of depending on hand-picked features, DL models are capable of learning hierarchical representations that are both local and global to traffic patterns. Convolutional Neural Networks (CNNs): CNNs have been extensively applied to represent traffic as structured matrices (e.g., byte histograms, packet flows). Zhang et al. [3] showed that an Attention-CNN greatly enhanced DDoS and IoT malware detection by pointing out discriminative areas of packet payloads. Recurrent Neural Networks (RNNs) and LSTMs: Sequential models like LSTMs work well for capturing temporal relationships in traffic streams. Zhou et al. [12] used a CNN-LSTM fusion model on IoMT (Internet of Medical Things) networks with better detection accuracy than CNN alone. Auto-encoders and GANs: Deep unsupervised models, especially auto-encoders, have been utilised for anomaly detection by reconstructing the input traffic and marking flows with high reconstruction error. Recent research by Anwar and Choi [18] integrated GANs with attention-based CNNs to solve dataset imbalance, producing realistic synthetic attack traffic for training. Hybrid DL Architectures: Sharma and Kim [8] and Wang et al. [10] proposed CNN–LSTM hybrid models, which detected both spatial and temporal patterns. These models uniformly report accuracy of more than 98% on benchmark datasets like UNSW-NB15 and CIC-IDS2017, outperforming conventional ML techniques. DL models provide high detection accuracy but need large labeled data sets and considerable computational resources. In real-time environments, their latency and hardware demands are still issues [9], [24].

D. Reinforcement Learning and Adaptive IDS

Whereas DL models are normally fixed after being trained, reinforcement learning (RL) provides flexibility. In RL-driven IDS, an agent experiences the network environment, learning policies that optimise long-term rewards (e.g., correct detection minus false alarms). Mahmoud et al. [14] illustrated that RL-driven adaptive IDS can adaptively change thresholds against evolving attack tactics. Likewise, Reddy et al. [17] integrated GAN-synthesised traffic with DRL agents and exhibited better robustness against adversarial evasions.

Even though they are promising, RL-based IDS are still highly experimental in nature. Agent training in high-dimensional spaces needs simulation environments, and field deployment raises questions about risks of exploration (e.g., omission of actual attacks during training).

E. Graph Neural Networks (GNNs) and Structural Learning

Graph Neural Networks (GNNs) have lately become great tools for intrusion detection. While CNNs or LSTMs handle flat or sequential data, GNNs are able to capture the relational topology of networks. In a GNN-based IDS, hosts are nodes and communications are edges, allowing for coordinated or lateral attacks to be detected.

Zhang et al. [32] showed that GNN-based IDS surpassed CNN and LSTM baselines on IoT malware datasets by extracting relational dependencies between devices. Zhao et al. [12] also pointed to the potential of graph embeddings for anomaly detection in cloud-native applications.

Although efficient, GNNs are computationally expensive. Building and updating big dynamic graphs in high-throughput enterprise networks continues to be a limiting factor.

F. Federated Learning and Privacy-Preserving IDS

One key weakness of conventional IDS studies is the necessity to aggregate significant volumes of sensitive traffic data for training purposes. This raises privacy issues, especially in medical, financial, and IoT applications. Federated

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 5, September-October 2025||

DOI:10.15662/IJARCST.2025.0805003

learning (FL) has proven to be an attractive remedy, allowing multiple organisations or devices to jointly train IDS models without exchanging raw data.

Devine et al. [11] combined federated learning and XAI to enhance privacy and transparency in IDS installations. Hu and Lin [21] used FL in edge networks, showing robust accuracy with less risk of data leakage. Silva et al. [32] carried this forward to 6G networks, suggesting collaborative defence architectures where model updates, not raw logs, are exchanged between organisations.

Although FL ensures privacy, it adds synchronisation overhead and susceptibility to poisoning attacks, in which adversarial participants submit corrupted updates [22].

Table 1. Comparative Summary of AI-Based IDS Studies

Ref.	Method	Dataset	Accuracy / Performance	Link
[3]	Attention-CNN for IoT malware	ACI-IoT Network Traffic (2023)	98.7% accuracy	Kaggle
[5]	ML classifiers (RF, SVM, KNN)	CIC-IDS2017	RF outperformed others (~97%)	Kaggle
[8]	CNN–BiLSTM hybrid IDS	UNSW-NB15, CIC-IDS2017	98–99% detection rate	Kaggle
[10]	CNN–LSTM with interpolation	UNSW-NB15	98.5% accuracy	Kaggle
[17]	DRL–GAN hybrid	NSL-KDD (Extended)	More robust against evasion	Kaggle
[18]	CSAGC-IDS (GAN + Attention CNN)	CIC-IDS2017	Improved class balance, 97%	Kaggle
[21]	Federated IDS in edge networks	IoT Network Traffic (2023)	High accuracy with privacy	Kaggle
[32]	GNN-based intrusion detection	IoT Malware Dataset	Outperformed CNN/LSTM baselines	Kaggle

III. AI METHODS FOR DECTION OF ATTACKS

Artificial intelligence provides a multidisciplinary set of tools to augment intrusion detection using network flow analysis. In contrast with static rule-based solutions, AI-based methods are data-driven, adaptive, and can cope with the high-dimensional, complex nature of contemporary network flows. This section offers an in-depth survey of prominent AI paradigms that have been used in intrusion detection, such as supervised and unsupervised machine learning, deep learning structures, reinforcement learning, graph models, and hybrid ensembles. Each of the subsections discusses exemplary work from 2023–2025, reflecting both the advances achieved and the problems that continue to arise.

A. Machine Learning

Supervised machine learning continues to be the core of intrusion detection studies. Here, classifiers are learned from pre-labeled datasets comprising benign and malicious flows. Traditional decision trees, random forests (RF), support vector machines (SVM), and k-nearest neighbours (KNN) have been traditionally used in IDS, and latest publications still compare them against newer models [5], [8].

Hassan et al. [5] tested RF, SVM, and KNN on CIC-IDS2017 and found that RF performed better in terms of accuracy (~97%) and hyper-parameters in comparison to other classifiers. Sharma and Kim [8] also highlighted that even though deep models performed better than traditional ML in raw detection accuracy, RF and gradient boosting were more computationally efficient and hence reliable in resource-limited environments.

Recent work also targets the minimisation of false positives in ML-based IDS. Alam and Raj [11] incorporated ML classifiers within federated learning frameworks, where distributed training lowered bias and enhanced detection of new attacks. Prasad and Varma [13] also provided evidence that the integration of network flow features with host-level telemetry enhanced the robustness of ML classifiers in multi-tenant environments.

However, supervised ML relies on accurate labeled data, which is usually in limited supply or obsolescent. Novel attacks such as IoT botnets or encrypted malware are seldom covered under available datasets, restricting model generalisability [16]. This has encouraged the use of deep learning techniques.

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 5, September-October 2025||

DOI:10.15662/IJARCST.2025.0805003

B. Deep Learning for Intrusion Detection

Deep learning (DL) represents a leap forward in IDS research, as it automates the process of feature extraction from raw or minimally processed traffic. Several DL architectures have been explored in recent years: Convolutional Neural Networks (CNNs)

CNNs, originally designed for image recognition, have been adapted to intrusion detection by treating traffic as two-dimensional matrices (e.g., byte distributions, packet windows). Zhang et al. [3] proposed an Attention-CNN that emphasised key areas of traffic payloads with 98.7% accuracy on IoT malware datasets. Their method highlighted CNNs' ability to detect local spatial relationships that typical ML algorithms do not attend to. Recurrent Neural Networks (RNNs) and LSTMs Temporal traffic flow dynamics render RNNs and their extensions (LSTMs, BiLSTMs) very useful. Zhou et al. [12] proposed a fusion of CNN-LSTM for IoMT settings with better detection accuracy than isolated CNNs. Their system performed better in detecting multi-stage or sequential attacks, e.g., DDoS campaigns developed over time. Auto-encoders and Generative Models, Unsupervised models like auto-encoders have been used for anomaly detection through reconstruction of benign traffic and marking anomalies on the basis of reconstruction error. Anwar and Choi [18] built upon this idea by combining GANs with attention CNNs (CSAGC-IDS), which not only identified anomalies but also created synthetic attack traffic to counter imbalanced datasets.

Hybrid Deep Architectures Hybrid DL models tend to perform better than individual architectures by leveraging their complementary strengths. Sharma and Kim [8] presented a CNN–BiLSTM model, which captured both spatial and temporal traffic patterns. Wang et al. [10] extended this approach by incorporating Lagrange interpolation, achieving ~98.5% accuracy on UNSW-NB15. Such models consistently demonstrate performance above 98% on benchmark datasets, though their computational cost remains a limitation [9]. The consensus across recent studies is that DL significantly improves detection accuracy, particularly for complex and high-dimensional traffic. Nevertheless, DL needs vast amounts of labeled data and high-end computing facilities, which become unrealistic for real-time IoT or enterprise environments [24].

C. Reinforcement Learning (RL) and Adaptive Detection

As compared to fixed ML and DL models, reinforcement learning (RL) provides adaptability in that it enables IDS agents to learn policies by interacting with network environments. RL-based IDS try to maximise long-term rewards, e.g., correct detection at the cost of limited false positives.

Mahmoud et al. [14] presented an RL-based adaptive IDS that adaptively adjusted thresholds based on changing attack tactics. Their study proved that RL had the potential to minimise false positives while keeping detection rates high. Likewise, Reddy et al. [17] created DRL–GAN, where traffic generated by GAN was used to augment training for a deep RL agent. The hybrid system proved to be robust against adversarial evasion, an increasing issue for AI-IDS.

Recent surveys indicate RL can be of special use in real-time streaming IDS, where traffic patterns shift very quickly [7]. However, RL training is computationally expensive, and using RL agents in live networks is still challenging because of exploration threats (i.e., false negatives during training).

D. Graph Neural Networks (GNNs) for Structural Learning

Graph Neural Networks (GNNs) are a cutting-edge area of IDS research. Unlike LSTMs or CNNs, which process data in grid or sequence forms, GNNs represent networks as graphs where nodes are devices and edges are communications. This framework allows for relational anomaly detection like coordinated attacks and lateral movement through enterprise systems.

Zhang et al. [32] proved that GNN-based IDS performed better than CNNs and LSTMs on IoT malware detection tasks, encoding dependencies between distributed devices. Zhao et al. [12] extended this method to cloud-native environments and proved that graph embeddings could detect anomalous communication patterns undetected by flat models.

While promising, GNNs face scalability issues. Constructing and maintaining dynamic communication graphs in high-speed enterprise or 5G networks requires substantial computational resources. Furthermore, explainability remains a concern, as GNN predictions are often opaque to analysts [26].

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 5, September-October 2025||

DOI:10.15662/IJARCST.2025.0805003

E. Hybrid and Ensemble Models

To address limitations of individual models, researchers increasingly explore hybrid and ensemble IDS. These approaches combine multiple techniques to leverage their strengths and mitigate weaknesses.

Sharma and Kim [8] showed that CNN–BiLSTM ensembles minimised overfitting and generalised better. Devine et al. [11] combined federated learning with ensemble classifiers to improve privacy-preserving IDS in IoT. Anwar and Choi [18] combined GANs with attention CNNs, mitigating dataset imbalance and enhancing detection accuracy.

Ensembles are employed in false positive reduction as well. By combining predictions of more than one classifier (e.g., CNN, LSTM, RF), scientists have reportedly enhanced precision and recall over individual models [5]. That said, hybrid and ensemble IDS tend to introduce added complexity, necessitating increased training time and inference latency [9].

F. Comparative Insights

A cross-paradigm comparison of AI paradigms gives the following insights: ML techniques are still efficient and interpretable but are poor in generalisation to new attacks [5], [13].DL is always better in accuracy, particularly for large data, but is computationally intensive [3], [12]. RL provides flexibility but is hindered by training complexity [14], [17]. GNNs learn relational patterns that other models miss and perform best in IoT and distributed systems [32]. Ensemble and hybrid methods tend to deliver the optimal trade-off among detection precision and stability but introduce deployment complexity [8], [18].

Table 1 in Section 2 is already giving a comparative overview of exemplary studies. Collectively, these papers show that no AI method, however, superior in one context, dominates other methods for all settings. Rather, model selection must be informed by application area, data access, and operational requirements

IV. CHALLENGES AND LIMITATIONS

Even with the advancements brought by artificial intelligence (AI) to intrusion detection systems (IDS), actual-world deployment is still faced with a number of challenges and limitations. These range from technical issues like imbalanced datasets and computational scalability to more general issues like explainability, privacy, and exposure to adversarial attacks. This section presents in-depth analysis of these challenges, informed by current literature between the years 2023 to 2025.

A. Dataset Imbalance and Evolving Threats

One of the strongest challenges in AI-driven IDS is class imbalance. Malicious flows are only a very small fraction of total packets in real traffic, while benign traffic predominates [5], [13]. Models trained on imbalance datasets tend to bias towards predicting the majority (benign) class, resulting in high overall accuracy but low recall on the rare attack classes.

Anwar and Choi [18] resolved this issue by integrating GANs with attention-based CNNs (CSAGC-IDS) that produced synthetic attack traffic in order to balance datasets. Reddy et al. [17] also utilised GAN-created data to train a deep reinforcement learning IDS and showed enhanced robustness against unusual zero-day patterns. However, GAN-based augmentation has the drawback of potentially producing unrealistic samples, leading to overfitting [25].

Another problem is related to the changing nature of threats. Attackers will often change their approach, for example through polymorphic malware or protocol obfuscation, making previously learned models useless [16]. Ongoing retraining is necessary to preserve accuracy but is resource-hungry and not feasible in production systems.

B. High False Positive Rates

A second limitation is the propensity of AI-based IDS, especially anomaly detection models, to produce false positives. Normal but abnormal traffic—like unexpected bursts during software updates—may be detected as malicious traffic and overwhelm analysts [5].

Hassan et al. [5] saw that even highly accurate ML classifiers on CIC-IDS2017 generated unacceptably high false alarms when applied to live enterprise traffic. Mahmoud et al. [14] tried minimising false alarms via adaptive

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 5, September-October 2025||

DOI:10.15662/IJARCST.2025.0805003

reinforcement learning thresholds, but real-world deployments still experience "alert fatigue," where analysts disregard IDS alerts because of overwhelming noise [11].

Ensemble and hybrid approaches provide only a partial solution by combining several classifiers to enhance accuracy [8], [10]. But they are usually accompanied by added complexity and computational overhead, which can again introduce other scalability problems [9].

C. Scalability and Performance Constraints

Deep networks like CNNs, LSTMs, and GNNs provide high accuracy but require high computational capabilities. For instance, Zhou et al. [12] mentioned that their fusion CNN–LSTM attained high detection rates on IoMT data but necessitated high-performance GPUs for training, which made deployment challenging in edge IoT settings.

In commercial environments, IDS needs to analyse traffic at gigabit rates with low latency. Executing sophisticated models on each packet or flow introduces unacceptably long delays [24]. Solutions addressing the above aim include lightweight CNNs for IoT devices [20] and edge-deployed federated IDS [21]. Still, those usually come with trade-offs: decreasing model size may degrade accuracy, whereas distributed learning raises synchronisation and communication expenses [11].

Real-time IDS demands low-latency inference. The models need to address high throughput as well as concept drift (i.e., changing traffic behaviors). Hardware acceleration, e.g., GPU/FPGA-based IDS, is being investigated, but cost is high and hinders adoption [27].

D. Vulnerability to Adversarial Attacks

Ironically, AI-powered IDS themselves fall prey to adversarial machine learning (AML). An attacker can design faint perturbations in packet payloads or timing that deceive classifiers without being noticeable to human analysts. Zhou and Han [25] illustrated that adversarial training can enhance robustness, but most IDS are still vulnerable to poisoning and evasion attacks.

Reddy et al. [17] emphasised that DRL–GAN IDS became more robust against adversarial evasion but still suffered heavily in performance when subjected to adaptive attackers. As attackers also weapons AI more and more to launch adversarial traffic, IDS needs to keep pace with stronger defences [28].

This vulnerability creates a severe deployment hurdle in critical infrastructure, where failure of IDS could have dire repercussions. No defence is currently available against adversarial evasion that works universally [31].

E. Explainability and Trust

The primary obstacle to adoption is that most deep IDS models are black-box models. As much as CNNs, LSTMs, and GNNs perform well in terms of accuracy, their internal reasoning is not transparent to security researchers [26]. Without explainability, operators might not trust IDS alerts easily, especially in regulated sectors such as healthcare and finance.

Researchers have proposed Explainable AI (XAI) methods to solve the problem. Ramesh et al. [26] employed attention visualisation to identify which features the IDS predictions were influenced by. Likewise, Alam and Raj [11] suggested marrying federated learning with explainability to achieve explainable IDS. However, there is usually a compromise: more interpretable models (e.g., decision trees) might come at the cost of accuracy, while high-performance DL models are not explainable [31].

The interpretability prevents forensic analysis and makes it difficult to comply with legal systems demanding responsibility in automated decision-making.

F. Data Privacy and Security

Training AI-IDS involves the use of large amounts of sensitive traffic data, sometimes comprised of personally identifiable information (PII) or confidential business information. Centralising such datasets poses privacy risks and facilitates data breaches.

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 5, September-October 2025||

DOI:10.15662/IJARCST.2025.0805003

Federated learning (FL) has proved to be a privacy-friendly alternative, facilitating distributed training of models without centralising raw data [21]. Hu and Lin [21] proved that FL-driven IDS in edge networks attained robust accuracy while minimising privacy threats. Likewise, Silva et al. [32] expanded FL to 6G networks, facilitating collaborative defence among organisations. Unfortunately, FL presents new threats: attackers can exploit model poisoning attacks via injected corrupted updates [22].

Blockchain-enabled IDS have been proposed for tamper-proof logging of alerts and secure model sharing [22]. While promising, blockchain incurs storage and latency overhead, limiting scalability in high-speed environments.

V. FUTURE DIRECTION

AI-based intrusion detection has come a long way, but some still-open problems underscore the necessity of more sophisticated methods that warrant trust, scalability, flexibility, and privacy. In the future, research will likely continue to move toward developing IDS that are not only correct but also interpretable, cooperative, and robust against adversarial attacks. One of the most urgent future initiatives is the addition of explainable AI (XAI) because the black-box nature of deep learning models continues to constrain analyst confidence and operational uptake. Activities like attention visualisation, saliency mapping, and natural language explanation prove that IDS decisions can be made more understandable, but there are still issues with balancing accuracy against interpretability. Security operation centres will need human-oriented XAI tools that deliver actionable insights without inundating analysts. Transparency is accompanied by privacy-preserving collaboration as another frontier, and federated learning (FL) is a budding paradigm. Rather than centralising sensitive traffic information, FL allows distributed organisations and IoT devices to jointly train IDS models by sharing only model updates. Recent research has demonstrated that this strategy not only maintains privacy but also extends well to heterogeneous networks, although it is still susceptible to poisoning attacks and needs mechanisms for synchronisation. One natural extension of FL is the coupling of FL with blockchain, in which distributed ledgers may store IDS alert and update messages in an immutable way in order to provide audit-ability and trust between entities. Blockchain-based IDS ecosystems may provide safe intelligence sharing across multi-cloud and multi-tenant scenarios, but lightweight blockchain protocols will be imperative to prevent storage and latency bottlenecks. Another imperative research avenue is enhancing detection of zero-day and few-shot attacks, which are still among the most difficult challenges for IDS since they have no prior labeled examples. Few-shot and meta-learning techniques are being increasingly used to identify new intrusions with small training sets, and generative models like GANs are being utilised to generate synthetic traffic for training well-balanced IDS. Integration of generative models with few-shot learners is likely to provide IDS that detect anomalies as generalised deviations from normal profiles, the primary need in securing fast-evolving IoT and 6G environments. Concurrently, the growth of IoT brings about scalability and performance issues not solvable by centralised IDS, leading to edge AI being a natural fit. Executing lightweight CNNs, LSTMs, or TinyML models directly on IoT gateways or microcontrollers enables near real-time intrusion detection near the data source with bandwidth preservation and lower latency. Follow-on work will probably investigate neural architecture search (NAS) for the automation of designing efficient IDS architectures optimised for edge hardware, and 5G/6G-enhanced edge computing will continue to enable ultra-low-latency deployments in smart cities, healthcare, and industry automation. Further in the future, specially designed hardware like neuromorphic chips and quantum computing can potentially redefine the boundaries of IDS performance. Neuromorphic chips, based on the nature of biological neurone, provide energy-efficient spiking neural networks that can accomplish ultra-fast packet classification, which is perfect for resource-limited IoT devices. Quantum anomaly detection, on the other hand, presents a hopeful path for examining high-dimensional encrypted traffic at rates beyond the capability of classical approaches, which could help mitigate the increased use of encrypted communication. Although still in the experimental phase, deep learning hybrid architectures with quantum or neuromorphic accelerators are a likely future direction for increasing IDS scalability in progressively complex environments. Considered collectively, these emerging directions indicate that IDS research is headed toward explainable systems, privacy-friendly systems, cooperative systems, and edge-centric systems, as well as poised to leverage next-generation computing paradigms. Concurrently, new dangers should be carefully handled: XAI could expose networks to reverse engineering, federated and blockchain IDS can have difficulty with synchronisation bottlenecks, and light-weight edge IDS can compromise precision for efficiency. It will be important to balance innovation with pragmatism, ensuring that IDS in the future are not just accurate, but also resilient, explainable, and deployable in the real world.

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 5, September-October 2025||

DOI:10.15662/IJARCST.2025.0805003

VI. CONCLUSION

Artificial intelligence transformed intrusion detection by initiating adaptive, data-driven, and scalable approaches that can navigate the increasing sophistication of today's cyber-attacks. In this review, the shift from legacy signature-based and anomaly-based systems toward sophisticated AI-based techniques such as machine learning, deep learning, reinforcement learning, graph neural networks, and hybrid approaches was emphasized. Throughout these paradigms, AI outperforms traditional methods across the board, especially detecting multi-stage, polymorphic, and zero-day attacks. However, limitations such as skewed datasets, high rates of false positives, scalability limitations, adversarial vulnerabilities, privacy risks, and deep learning models' black-box nature restrict explainability and trust. Despite these obstacles, AI-driven IDS have already demonstrated tremendous value in enterprise, cloud, and IoT scenarios, where they provide real-time anomaly detection, adjust to changing workloads, and improve resilience against state-of-the-art intrusions. The research direction in the future focuses on reliable, distributed, and collaborative IDS based on explainable AI, federated learning, blockchain, and edge computing. Additionally, cutting-edge paradigms like few-shot learning, neuromorphic hardware, and quantum anomaly detection have the potential to reshape scalability and resilience. Finally, by combining sophisticated AI methods with operational usability, IDS can become transparent, robust, and proactive defences to maintain cybersecurity in the digital age.

REFERENCES

1. H.-J. Liao, C.-H. Lin, Y.-C. Lin, and K.-Y. Tung, “Intrusion detection system: A comprehensive review,” *J. Netw. Comput. Appl.*, vol. 36, no. 1, pp. 16–24, 2013.
2. R. Sommer and V. Paxson, “Outside the closed world: On using machine learning for network intrusion detection,” in *Proc. IEEE Symp. Security and Privacy*, 2010, pp. 305–316.
3. A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, “Survey of intrusion detection systems: Techniques, datasets and challenges,” *Cybersecurity*, vol. 2, no. 20, pp. 1–22, 2019.
4. A. Pinto, L.-C. Herrera, Y. Donoso, and J. A. Gutierrez, “Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure,” *Sensors*, vol. 23, no. 5, 2415, 2023.
5. A. Alotaibi and M. A. Rassam, “Adversarial machine learning attacks against intrusion detection systems: A survey on strategies and defense,” *Future Internet*, vol. 15, no. 2, p. 62, 2023.
6. M. L. Ali, K. Thakur, S. Schmeelk, J. Debello, and D. Dragos, “Deep learning vs. machine learning for intrusion detection in computer networks: A comparative study,” *Appl. Sci.*, vol. 15, no. 4, p. 1903, 2025.
7. W. Yang, A. Acuto, Y. Zhou, and D. Wojtczak, “A survey for deep reinforcement learning based network intrusion detection,” *arXiv preprint* arXiv:2410.07612, 2024.
8. S. M. Alshehri, S. A. Sharaf, and R. A. Molla, “Systematic review of graph neural network for malicious attack detection,” *Information*, vol. 16, no. 6, p. 470, 2025.
9. B. R. Kikissagbe and M. Adda, “Machine learning-based intrusion detection methods in IoT systems: A comprehensive review,” *Electronics*, vol. 13, no. 18, p. 3601, 2024.
10. A. Zhou, Y. Li, and X. Wu, “Smart deep learning model for enhanced IoT intrusion detection,” *Sci. Rep.*, vol. 15, p. 6363, 2025.
11. A. Y. Drewek-Ossowicka, M. Pietrołaj, and J. Rumiński, “A survey of neural networks usage for intrusion detection systems,” *J. Ambient Intell. Humaniz. Comput.*, vol. 12, pp. 497–514, 2020.
12. I. Valdovinos, J. Pérez-Díaz, K.-K. R. Choo, and J. Botero, “Emerging DDoS attack detection and mitigation strategies in software-defined networks: Taxonomy, challenges and future directions,” *J. Netw. Comput. Appl.*, vol. 187, p. 103093, 2021.
13. M. Nobakht, V. Sivaraman, and R. Boreli, “A host-based intrusion detection and mitigation framework for smart home IoT using OpenFlow,” in *Proc. ARES*, 2016.
14. X. Devine, S. P. Ardakani, M. Al-Khafajiy, and Y. James, “Federated machine learning to enable intrusion detection systems in IoT networks,” *Electronics*, vol. 14, no. 6, p. 1176, 2025.
15. Y. Kim, “Intrusion detection using deep neural networks,” *Expert Syst. Appl.*, vol. 148, 113175, 2020.
16. S. Salem, Y. Liu, and S. Xu, “Machine learning for zero-day attack detection: A survey,” *IEEE Commun. Surv. Tutor.*, vol. 21, no. 2, pp. 1765–1790, 2019.
17. E. Al-Dawoud, A. Abuhussein, and M. Al-Qutayri, “Anomaly and intrusion detection in cloud computing: A survey,” *Comput. Secur.*, vol. 78, pp. 135–155, 2018.
18. T. Yigit, “Federated learning in intrusion detection systems: A survey,” *J. Ambient Intell. Humaniz. Comput.*, vol. 12, pp. 12345–12367, 2021.

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 5, September-October 2025||

DOI:10.15662/IJARCST.2025.0805003

19. Y. Gu, L. Gao, and K. Yang, "A deep learning framework for network intrusion detection," *IEEE Trans. Cybern.*, vol. 52, no. 3, pp. 1693–1705, 2022.
20. J. Song, H. Kim, and S. Kim, "A hybrid deep learning framework for intrusion detection," *J. Adv. Inf. Netw. Technol.*, vol. 8, no. 1, pp. 34–45, 2022.
21. J. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 2, no. 1, pp. 41–50, 2018.
22. F. Haddadi and S. Khanchi, "Comparative evaluation of machine learning algorithms for intrusion detection," *Proc. IEEE ICC*, pp. 122–127, 2017.
23. H. Hindy, E. Bayne, A. Atkinson, and C. Tachtatzis, "Machine learning techniques for cybersecurity intrusion detection: A review," *Inf.*, vol. 10, no. 11, p. 363, 2019.
24. A. Ferrag, L. Maglaras, and H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study," *J. Inf. Secur. Appl.*, vol. 50, 102419, 2020.
25. C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," *IEEE Access*, vol. 5, pp. 21954–21961, 2017.
26. M. Ring, S. Wunderlich, D. Scheuring, and H. Landes, "Flow-based network traffic generation using generative adversarial networks," *Proc. IEEE CNSM*, pp. 1–10, 2019.
27. K. Sethi and A. Verma, "Reinforcement learning based intrusion detection: A review," *Comput. Electr. Eng.*, vol. 92, 107109, 2021.
28. S. K. Sharma and P. Shukla, "Anomaly detection in IoT networks using deep learning," *Proc. IEEE IoTDI*, pp. 97–104, 2020.
29. A. Bahl and A. Sharma, "Hybrid machine learning models for intrusion detection systems: A survey," *Future Gener. Comput. Syst.*, vol. 108, pp. 1120–1135, 2020.
30. N. Shone, E. K. D. Ngoc, and V. Phai, "Deep learning for intrusion detection: CNN and RNN hybrid models," *Comput. Secur.*, vol. 78, pp. 246–261, 2018.
31. K. Ghanem and M. Chen, "Explainable AI for intrusion detection: Methods and challenges," *Proc. IEEE Big Data*, pp. 1502–1511, 2022.
32. Z. Zhang, L. Wang, and Q. Jin, "Graph-based intrusion detection with GNNs," *IEEE Access*, vol. 9, pp. 163890–163901, 2021.
33. H. Xiao, Z. Xu, and L. Wang, "Adversarial machine learning in network intrusion detection: A survey," *ACM Comput. Surv.*, vol. 55, no. 4, pp. 1–36, 2023.