Journal on
Artificial Intelligence

Tech Science Press

https://doi.org/10.32604/jai.2026.079401

REVIEW

A Review of Advancements in Deep Learning Approaches for Intrusion Detection Systems

Akash Garg*

Department of Computer Science and Engineering, Rajkiya Engineering College, Mainpuri, India

*Corresponding Author: Akash Garg. Email: garg.theakash92@gmail.com

Received: 21 January 2026; Accepted: 17 March 2026; Published: 12 May 2026

ABSTRACT: As cyber threats continue to evolve in scale and sophistication, the need for intelligent and adaptive security mechanisms has become increasingly urgent. Intrusion Detection Systems (IDS) are critical components in safeguarding computer networks from malicious activities. This review paper presents a comprehensive analysis of recent advancements in deep learning-based IDS, examining various architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), autoencoders, and generative adversarial networks (GANs). The study compares traditional intrusion detection techniques with modern deep learning approaches, highlighting their strengths, limitations, and suitability for real-world deployment. Special attention is given to hybrid models that integrate anomaly and misuse detection, as well as techniques that address challenges such as data imbalance, feature selection, and real-time detection. Benchmark datasets like KDD'99, NSL-KDD, and UNSW-NB15 are discussed in the context of evaluating IDS performance. The review concludes that deep learning offers significant promise in enhancing the accuracy, adaptability, and scalability of IDS, though challenges remain in terms of computational cost and interpretability. This paper serves as a resource for researchers and practitioners seeking to understand the current landscape of deep learning in network intrusion detection and identifies potential directions for future research.

KEYWORDS: Deep learning; intrusion detection system (IDS); CNN; RNN; autoencoder; GAN; cybersecurity; hybrid models; anomaly detection

1 Introduction

In the modern digital era, the proliferation of internet-connected devices and complex network infrastructures has led to a dramatic increase in cybersecurity threats. From data breaches and malware infections to sophisticated zero-day attacks, organizations are under constant threat from a wide variety of malicious actors. Traditional security mechanisms, such as firewalls and signature-based antivirus systems, are no longer sufficient to cope with the dynamic and evolving nature of cyberattacks [1]. As a result, Intrusion Detection Systems (IDS) have become an essential component of network defense, offering real-time monitoring and threat detection capabilities.

Conventional IDS methods, including signature-based and anomaly-based approaches, have notable limitations. Signature-based systems are highly effective for known threats but fail to detect novel or obfuscated attacks. On the other hand, anomaly-based systems can identify previously unseen behavior but often suffer from high false positive rates. To address these challenges, the field has increasingly turned to artificial intelligence and, more recently, deep learning techniques, which have demonstrated exceptional performance in pattern recognition and data classification tasks.

This review paper explores the application of deep learning models—such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Autoencoders, and Generative Adversarial Networks (GANs)—in the development of intelligent IDS. By analyzing and comparing existing research, this study aims to identify the advantages, limitations, and real-world applicability of these models in intrusion detection. Furthermore, the paper discusses the datasets commonly used for training and evaluation, challenges in implementing deep learning-based IDS, and promising directions for future research [1,2]. The goal is to provide a comprehensive overview that aids researchers and practitioners in understanding the current state of deep learning in IDS and its potential to transform network security.

Fig. 1, illustrates the overall architecture and working principle of an Intrusion Detection System (IDS), highlighting how network traffic is monitored and analyzed to detect potential threats.

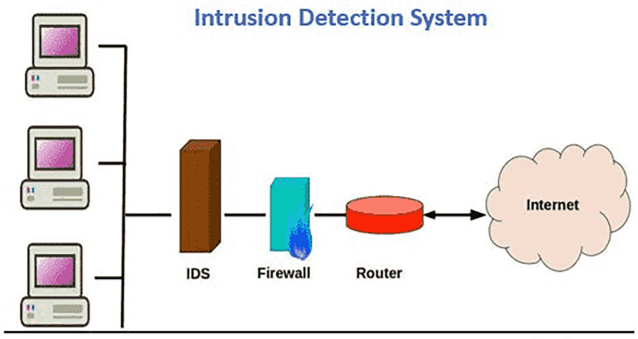

Figure 1: Intrusion detection system.

1.1 Types of Intrusion Detection Systems

Intrusion Detection Systems can be classified into the following types based on their functionality and deployment:

1. **Network-based IDS (NIDS):** Monitors network traffic to identify suspicious activities. It is deployed at critical points in the network.

A Network-based Intrusion Detection System (NIDS) is a security tool designed to monitor and analyze network traffic in real time to detect suspicious or malicious activity. It operates by inspecting data packets as they travel across the network, looking for patterns or behaviors that may indicate an attack, such as port scanning, unusual protocol usage, or attempts to exploit vulnerabilities [1].

NIDS is typically deployed at key points within a network, such as at the perimeter or between critical segments, allowing it to monitor traffic flowing to and from devices. It can use both signature-based and anomaly-based detection methods to identify threats, depending on its configuration.

One of the main advantages of NIDS is its ability to provide a broad view of network activity without requiring software to be installed on individual devices. However, it may struggle to analyze encrypted traffic or detect attacks that originate from within the network if not properly positioned.

NIDS plays a vital role in enterprise security by helping organizations detect and respond to threats quickly, often serving as an early warning system for network-based attacks [2].

As shown in Fig. 2, a Network-based Intrusion Detection System (NIDS) is positioned at strategic locations within the network to analyze incoming and outgoing traffic for potential threats.

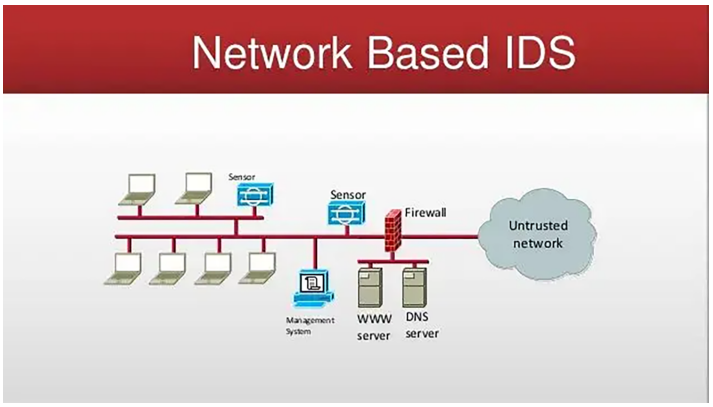

Figure 2: Network based intrusion detection system.

2. **Host-based IDS (HIDS):** Monitors a single system for signs of intrusion, such as unauthorized access or file modifications.

A Host-based Intrusion Detection System (HIDS) is a security solution that monitors and analyzes activity on individual devices, such as servers or workstations, to detect signs of malicious behavior or unauthorized access. Unlike network-based systems, HIDS focuses on what's happening within the host itself, including system logs, file changes, running processes, and user actions [3].

By operating directly on the host, HIDS can detect threats that might bypass network defenses, such as insider attacks, unauthorized configuration changes, or the presence of malware. It often uses a combination of signature-based detection and behavioral analysis to identify suspicious activity.

One of the key benefits of HIDS is its ability to provide detailed insights into specific devices, offering precise detection and logging capabilities. However, it typically consumes more system resources and needs to be installed and managed on each individual host, which can be challenging in large environments [4,5].

HIDS is commonly used to protect critical systems, support compliance requirements, and complement other security tools by offering visibility into threats that occur at the system level [6].

As shown in Fig. 3, a Host-based Intrusion Detection System (HIDS) operates at the system level, analyzing logs, file integrity, and user activities to detect potential intrusions.

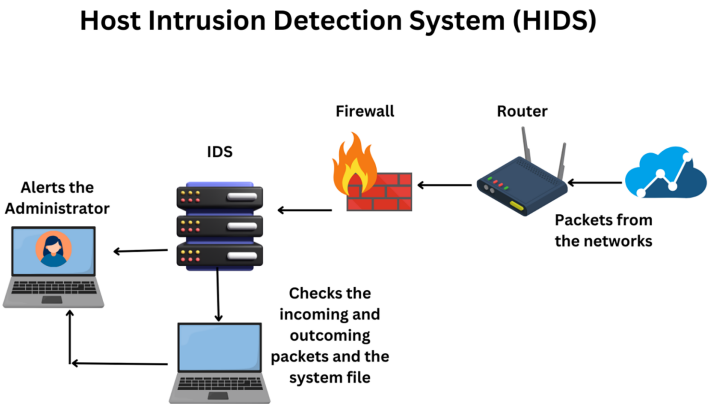

Figure 3: Host based intrusion detection system.

3. **Signature-based IDS:** Detects attacks based on known patterns or signatures of malicious activities.

A signature-based Intrusion Detection System (IDS) is a security tool that identifies threats by comparing network traffic or system activity against a database of known attack patterns, called signatures. These signatures represent identifiable characteristics of previously detected threats, such as specific malware or exploit behaviors. When the IDS detects activity that matches one of these signatures, it flags it as a potential intrusion.

This type of IDS is effective at quickly recognizing known threats with a high degree of accuracy and typically produces fewer false positives. However, its main limitation is that it cannot detect new or unknown attacks that do not match existing signatures. To remain effective, the signature database must be regularly updated with the latest threat information.

Signature-based IDS is commonly used in environments where known threats are a concern and can be deployed on networks or individual devices. It is often used alongside other detection methods to provide more comprehensive protection [7].

4. **Anomaly-based IDS:** Identifies deviations from normal behavior to detect unknown attacks, though it often suffers from high false-positive rates.

An anomaly-based Intrusion Detection System (IDS) is a security mechanism that monitors system or network activity and identifies potential threats by detecting unusual behavior that deviates from the normal baseline. Instead of relying on predefined attack signatures, it builds a model of normal activity—such as typical user behavior, network traffic patterns, or system operations—and then flags any significant deviations as possible intrusions.

This approach allows anomaly-based IDS to detect previously unknown or zero-day attacks, making it valuable for identifying new and evolving threats. However, since it relies on identifying deviations from normal behavior, it may produce more false positives, especially in dynamic environments where normal activity frequently changes.

Anomaly-based IDS is often used in combination with signature-based systems to provide broader and more adaptive security coverage. It is particularly useful in detecting insider threats, policy violations, or sophisticated attacks that do not match known patterns [8].

Here's a clear breakdown of the differences between Host-based, Network-based, Anomaly-based, and Signature-based IDS types:

Category	Host-Based IDS (HIDS)	Network-Based IDS (NIDS)	Anomaly-Based IDS	Signature-Based IDS
Monitoring Scope	Specific host or device	Entire network segment	Behavior deviations	Known attack patterns
Data Source	OS logs, system calls, file integrity	Network traffic (packets, flows)	Historical behavior or statistical norms	Database of known threats
Deployment	Installed on each host	Deployed on network gateways/switches	Can be host or network based	Can be host or network based

(Continued)

(continued)

Category	Host-Based IDS (HIDS)	Network-Based IDS (NIDS)	Anomaly-Based IDS	Signature-Based IDS
Detection Method	Local activity analysis	Traffic inspection	Detects deviations from normal behavior	Pattern matching
Strengths	Deep insight into host; good for insider threats	Detects wide range of network attacks	Can detect novel or zero-day attacks	High accuracy for known attacks
Weaknesses	Can't see network-wide activity	Misses local (host-only) attacks	High false positive rate	Can't detect unknown attacks
Example Tools	OSSEC, Tripwire	Snort, Suricata, Zeek	Machine learning models, autoencoders	Snort (with rules), antivirus engines

1.2 Definitions of Common Cyber Attacks

Cyber-attacks exploit vulnerabilities in systems and networks to compromise their integrity, confidentiality, or availability.

Common types include:

1. Denial of Service (DoS) Attacks: Overloads a system with requests to render it unavailable.
2. Distributed Denial of Service (DDoS) Attacks: Similar to DoS but originates from multiple sources, making it harder to mitigate.
3. Phishing Attacks: Tricks users into providing sensitive information through deceptive emails or websites.
4. Man-in-the-Middle (MitM) Attacks: Intercepts and alters communication between two parties without their knowledge.
5. Ransomware Attacks: Encrypts data and demands payment for its decryption.
6. Zero-Day Attacks: Exploits vulnerabilities unknown to the software vendor, making them difficult to detect and mitigate.
7. SQL Injection: Exploits vulnerabilities in SQL-based applications to execute malicious queries.

Here's a table comparing the common types of cyberattacks, highlighting their key differences based on method, target, and difficulty of detection:

Attack Type	Method	Primary Target	Difficulty to Detect	Impact
DoS (Denial of Service)	Floods a system with traffic to exhaust resources	Servers/Networks	Low	Service disruption

(Continued)

(continued)

Attack Type	Method	Primary Target	Difficulty to Detect	Impact
DDoS (Distributed DoS)	DoS attack launched from multiple sources	Servers/Networks	Medium	Large-scale service disruption
Phishing	Deceptive emails or websites trick users into revealing info	Individual users	Medium	Credential theft/data breach
MitM (Man-in-the-Middle)	Intercepts communication between two parties	Communications/Data streams	High	Data manipulation or theft
Ransomware	Encrypts data and demands ransom	Files/Systems	Medium to High	Data loss, financial impact
Zero-Day	Exploits unknown software vulnerabilities	Applications/Systems	Very High	Undetected breaches, severe consequences
SQL Injection	Injects malicious SQL code into input fields	Databases/Web apps	Medium	Data exposure, unauthorized access

1.3 Deep Learning Techniques for IDS

Deep learning techniques have become increasingly important in enhancing the capabilities of Intrusion Detection Systems (IDS) [9]. These methods involve using advanced neural network architectures to analyze vast amounts of network or system data and detect abnormal or malicious behavior. Unlike traditional IDS approaches, which rely on predefined rules or static patterns, deep learning models can automatically learn complex patterns and adapt to evolving threats [10].

At the core of deep learning-based IDS are artificial neural networks, which are designed to mimic the way the human brain processes information. These models are trained using extensive datasets containing both normal and malicious activity, enabling them to identify subtle deviations that may indicate an intrusion, even for previously unseen threats [11].

Several deep learning techniques are commonly used in IDS:

1. **Convolutional Neural Networks (CNNs):** Primarily used for processing spatial data, CNNs can extract meaningful features from network traffic by identifying patterns in packet flows or connection data. They are particularly effective in recognizing localized patterns in structured inputs.
2. **Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM):** These models are ideal for analyzing sequential data and time-series information, such as the order of events in system logs or network sessions. LSTM networks, a special type of RNN, are especially effective at capturing long-term dependencies and patterns in data over time.

3. **Autoencoders:** These are unsupervised learning models used to detect anomalies. They work by compressing data into a lower-dimensional representation and then reconstructing it. If the reconstruction error is high, it may indicate abnormal or malicious activity, making autoencoders useful for anomaly detection.
4. **Deep Belief Networks (DBNs):** DBNs consist of multiple layers of unsupervised networks, such as Restricted Boltzmann Machines (RBMs). They can learn hierarchical representations of data and are capable of capturing complex relationships that traditional machine learning techniques might miss.
5. **Generative Adversarial Networks (GANs):** GANs can be used to generate synthetic attack data for training IDS models, helping to address the challenge of imbalanced datasets. They consist of two networks—the generator and the discriminator—that compete to improve the quality of the generated data.

1.4 Proposed Taxonomy of Deep Learning-Based Intrusion Detection Systems

Based on an extensive review of recent literature, deep learning-based Intrusion Detection Systems (DL-IDS) can be systematically categorized using a multi-dimensional taxonomy. This taxonomy helps in understanding existing research trends, identifying gaps, and guiding future IDS design.

The proposed taxonomy classifies DL-IDS along four key dimensions:

A. Taxonomy Based on Learning Paradigm
   Supervised Learning: CNN, DNN, RNN, LSTM (requires labeled data)
   Unsupervised Learning: Autoencoders, RBMs (anomaly detection)
   Semi-Supervised Learning: Hybrid AE + classifier models
   Generative Learning: GANs, CGANs (data augmentation & imbalance handling)
B. Taxonomy Based on Detection Strategy
   Signature-Based DL-IDS
   Anomaly-Based DL-IDS
   Hybrid DL-IDS (most effective, combines both)
C. Taxonomy Based on Deployment Environment
   Network-based IDS (NIDS)
   Host-based IDS (HIDS)
   Cloud-based IDS
   IoT/SCADA-based IDS
   SDN-based IDS
D. Taxonomy Based on Input Data Type
   Flow-based data
   Packet-level data
   System calls & logs
   Multi-modal data (network + host logs)

Here's a comparison table highlighting the differences between deep learning models commonly used in Intrusion Detection Systems (IDS): [12].

The main advantage of using deep learning for IDS is its ability to generalize from data and detect previously unseen threats (zero-day attacks). These models can learn from both labeled and unlabeled data, making them versatile in various deployment scenarios. However, they also come with challenges, such as the need for large, high-quality datasets, high computational requirements, and the risk of overfitting or adversarial manipulation.

Model	Primary Use	Strengths	Ideal For
Convolutional Neural Networks (CNNs)	Feature extraction from spatial/structured data	Excellent at recognizing localized patterns in traffic flows	Analyzing packet structure or connection maps
Recurrent Neural Networks (RNNs) & LSTM	Time-series and sequential data analysis	Captures temporal patterns; LSTM handles long-term dependencies	System logs, session tracking
Autoencoders	Anomaly detection	Learns compressed representations; detects outliers via reconstruction error	Detecting unknown or rare attack patterns
Deep Belief Networks (DBNs)	Hierarchical feature learning	Learns complex relationships; unsupervised pretraining can improve accuracy	Complex, high-dimensional network data
Generative Adversarial Networks (GANs)	Synthetic data generation	Addresses data imbalance; generates realistic attack scenarios for training	Simulating new attack types, enhancing datasets

1.5 Datasets for IDS

Datasets play a critical role in developing, training, and evaluating Intrusion Detection Systems (IDS), especially those based on machine learning or deep learning [13]. A well-constructed dataset provides the foundation for an IDS to learn patterns of both normal and malicious behavior, enabling it to detect potential threats effectively.

An IDS dataset typically contains labeled or unlabeled records of network traffic, system logs, or user activity. These records include features such as IP addresses, ports, protocols, timestamps, payload sizes, and attack types (when labeled). Depending on the focus of the IDS—network-based or host-based—the dataset may vary in structure and content [14].

There are two main types of IDS datasets:

1. **Synthetic Datasets:** These are generated in controlled environments using traffic simulation tools or testbeds. While they offer clean, labeled data and cover a wide range of attack types, they may not fully represent the complexity and unpredictability of real-world environments. Examples include:
    - **KDD Cup 1999:** One of the earliest IDS datasets, based on simulated traffic. It contains various types of attacks but has been criticized for redundancy and outdated attack patterns.
    - **NSL-KDD:** An improved version of the KDD'99 dataset that removes duplicate records and balances the classes better.
    - **CICIDS2017:** Developed by the Canadian Institute for Cybersecurity, it includes realistic, up-to-date attack scenarios and modern traffic patterns.
    - **UNSW-NB15:** Combines synthetic and real traffic to offer a more diverse set of attack types and better feature representation.

2. **Real-World Datasets:** These are collected from live environments and contain authentic traffic, which makes them more representative but also more challenging to label accurately. They may include sensitive data and are often anonymized for privacy. Examples include:
    - **MAWI Dataset:** Collected from a real internet backbone, it includes anonymized traffic for research purposes.
    - **CTU-13:** Focused on botnet traffic and includes labeled flows for various malware infections.
    - **TON_IoT Dataset:** Designed for intrusion detection in IoT environments, combining telemetry data with network traffic and system logs.

Here is a comparison table summarizing key differences between synthetic and real-world IDS datasets, along with notable examples:

Dataset Type	Description	Advantages	Limitations	Examples
Synthetic Datasets	Generated in controlled/test environments using traffic simulation or emulated scenarios.	Labeled and balanced data—Covers known attack types	May lack real-world complexity—Might be outdated	**KDD Cup 1999**—early simulated data, redundant and outdated **NSL-KDD**—cleaned-up KDD'99 version **CICIDS2017**—realistic modern attacks **UNSW-NB15**—combines real and synthetic traffic
Real-World Datasets	Collected from live network environments, includes authentic traffic flows and behavior patterns.	High realism—Reflects current threats	Harder to label—May include noise or privacy issues	**MAWI**—anonymized internet backbone traffic **CTU-13**—botnet-focused, labeled flow data **TON_IoT**—targets IoT with telemetry, system logs, and traffic

When selecting a dataset for IDS development, several factors should be considered:

1. **Diversity of attacks:** The dataset should cover a wide range of known and unknown threats.
2. **Label quality:** Accurate labeling is essential for supervised learning approaches.
3. **Realism:** The data should reflect actual network behavior as closely as possible.
4. **Balance:** An even distribution of normal and attack records helps avoid bias during model training.

2 Literature Review

This paper provides a comprehensive survey of deep learning-based intrusion detection systems (IDS), categorizing them based on the type of deep learning techniques employed, such as auto-encoders, restricted Boltzmann machines, recurrent neural networks, deep neural networks, and convolutional neural networks. The authors systematically review IDS architectures, highlighting the advantages and limitations of each approach. They also compare evaluation metrics, datasets, and simulation tools used in the literature, offering insights into the performance and applicability of these systems in real-world scenarios. The study identifies key challenges and future research directions, emphasizing the need for newer datasets and improved training methods to handle evolving cyber threats. The authors critically analyze the effectiveness of deep

learning in IDS, noting its superiority over traditional machine learning methods in feature extraction and classification accuracy. They discuss the limitations of existing datasets, such as KDDCup99 and NSL-KDD, and advocate for the use of more modern datasets like UNSW-NB15. The paper concludes with recommendations for addressing issues like imbalanced datasets, computational overhead, and the need for adaptive models capable of detecting zero-day attacks. This review serves as a valuable resource for researchers seeking to advance the field of deep learning-based intrusion detection [1].

This paper proposes a deep learning-based network intrusion detection system (NIDS) designed to detect and classify network attacks, including zero-day threats, using the UNSW-NB15 dataset. The authors employ a convolutional neural network (CNN) architecture with semi-dynamic hyperparameter tuning to optimize performance. Their model achieves high accuracy in multiclass classification, demonstrating the effectiveness of deep learning in identifying modern network attacks. The study also addresses challenges such as class imbalance and computational efficiency, offering practical solutions for real-world deployment. Ashiku and Dagli highlight the limitations of traditional IDS methods and emphasize the adaptability of deep learning models in handling complex network traffic patterns. Their approach includes data preprocessing, feature normalization, and the use of callback functions like Early Stopping to enhance model training. The results show significant improvements in detection rates for underrepresented attack classes, though the authors acknowledge the need for further research into feature reduction and transfer learning. This work contributes to the ongoing development of resilient and adaptive intrusion detection systems capable of mitigating evolving cyber threats [2].

This paper provides a comprehensive review of hybrid intrusion detection systems (IDS), focusing on the integration of misuse-based and anomaly-based detection techniques. The authors discuss the limitations of traditional firewalls and highlight the importance of IDS in identifying both internal and external attacks. They categorize IDS into misuse-based (signature-based) and anomaly-based systems, detailing statistical algorithms like PHAD (Packet Header Anomaly Detector), ALAD (Application Layer Anomaly Detector), LERAD (Learning Rules for Anomaly Detection), and NETAD (Network Traffic Anomaly Detection). The paper emphasizes the advantages of hybrid systems, which combine the low false-alarm rate of misuse-based detection with the ability of anomaly-based systems to detect unknown attacks. Case studies and experimental results are presented to demonstrate the effectiveness of hybrid IDS in improving detection rates. The authors also explore the components and working mechanisms of Snort, a popular misuse-based IDS, and compare it with anomaly-based methods. They conclude that hybrid systems offer a balanced approach to intrusion detection, addressing the shortcomings of standalone systems. The review serves as a valuable resource for researchers and practitioners, offering insights into the evolving landscape of network security and the potential of hybrid IDS to enhance threat detection in dynamic environments [3].

This survey paper systematically reviews the application of machine learning (ML) and deep learning (DL) techniques in intrusion detection systems (IDS). The authors categorize IDS into knowledge-based, statistical, and ML-based methods, discussing their evolution and limitations. They highlight the transition from shallow ML models to DL due to the latter's ability to handle large, complex datasets and automate feature extraction. The paper covers a wide range of ML algorithms, including ANN, SVM, and random forests, as well as DL models like CNN, RNN, and DBN, evaluating their performance on benchmark datasets such as KDD Cup 99 and NSL-KDD. The authors also address challenges in ML and DL-based IDS, such as high false-alarm rates, computational costs, and the need for labeled data. They emphasize the superiority of DL models in detecting novel attacks but note their resource-intensive nature. The paper concludes with future research directions, including the development of lightweight DL models and hybrid approaches to improve scalability and real-time detection capabilities. This comprehensive review serves as a guide for researchers exploring advanced techniques in cybersecurity [4].

The authors conducted a comprehensive survey and classification of deep learning-based Intrusion Detection Systems (IDS). They analyzed various deep learning techniques, such as Auto-encoders, Restricted Boltzmann Machines (RBMs), Deep Belief Networks (DBNs), Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and Recurrent Neural Networks (RNNs), highlighting their applications in IDS. The study provided a taxonomy of deep learning-based IDS approaches, discussing their contributions, limitations, and performance metrics. The authors also reviewed commonly used datasets like KDDCup'99, NSL-KDD, and CICIDS2017, emphasizing the importance of labeled data for supervised learning. The paper concluded with future research directions, including the integration of blockchain, handling imbalanced datasets, and improving anomaly detection in specialized environments like IoT and SDN. The research method involved a systematic literature review with predefined research questions, keyword-based searches, and quality assessments of selected papers. The authors compared evaluation metrics, classifiers, and feature extraction methods across different IDS schemes. They identified gaps in existing approaches, such as the lack of real-time detection capabilities and the need for hardware acceleration (e.g., FPGA, ASIC) to enhance deep learning model training. The study underscored the potential of hybrid models combining multiple deep learning techniques for improved intrusion detection accuracy. Future work suggestions included developing domain-specific datasets and exploring distributed deep learning for scalable IDS solutions [5].

This survey provides a detailed overview of machine learning and deep learning (DL) methods applied to cybersecurity and intrusion detection systems (IDS). The authors analyze recent DL-based approaches, focusing on their mechanisms, performance, and limitations, while also evaluating the benchmark datasets used for training these models. The paper categorizes IDS algorithms into rule-based, statistics-based, and machine learning-based methods, highlighting the advantages of DL for handling large-scale data and complex intrusion patterns. The survey reviews DL techniques such as Deep Belief Networks (DBNs), Autoencoders (AEs), Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Generative Adversarial Networks (GANs), comparing their effectiveness in intrusion detection. Additionally, the authors critically assess popular cybersecurity datasets like KDD99, NSL-KDD, and CICIDS2017, discussing their strengths and limitations. The study emphasizes the practical perspective of dataset evaluation, focusing on feature types, attack distributions, and reliability criteria. It also addresses the challenges of dataset bias and the need for novel features to adapt to evolving attack patterns. The authors conclude by outlining current research trends and future directions, making this survey a valuable resource for researchers and practitioners in cybersecurity. The comprehensive analysis of DL methods and datasets serves as a roadmap for understanding the potential of DL in IDS and highlights gaps for further exploration [6].

This paper surveys the application of deep learning (DL) techniques in intrusion detection systems (IDS), comparing them with traditional machine learning approaches. The authors begin by discussing the importance of IDS in cybersecurity, especially with the rise of IoT and networked devices, and categorize IDS into host-based, network-based, signature-based, and anomaly-based systems. They highlight the limitations of signature-based IDS in detecting novel attacks and advocate for anomaly-based systems powered by DL due to their adaptability and robustness. The paper provides background on DL algorithms, contrasting them with machine learning in terms of data requirements, feature extraction, and computational efficiency. The authors review commonly used datasets in IDS research, such as KDD Cup99, NSL-KDD, and CIC IDS 2017, noting their characteristics and shortcomings. They also discuss newer datasets like CSE-CIC-IDS2018 and MCFP Bot Traffic, which include diverse attack scenarios. The paper concludes by emphasizing the potential of DL-based IDS to improve detection accuracy and flexibility, while acknowledging challenges like the need

for large datasets and computational resources. The survey serves as a concise guide for researchers exploring DL in cybersecurity, offering insights into current trends and future directions [7].

The authors conducted a comprehensive survey on the application of deep learning techniques in anomaly-based intrusion detection systems (IDS). They reviewed and compared previous surveys focused on deep learning in cybersecurity, highlighting gaps and differences with their work. The paper proposed a novel fine-grained taxonomy to classify deep learning-based IDS solutions based on input data strategy, detection strategy, deployment strategy, and evaluation strategy. The authors also provided a descriptive and comparative analysis of experimental studies published between 2014 and 2018, discussing the role of deep learning in IDS, the impact of datasets, and the efficiency of proposed approaches. The survey concluded with open research challenges and future directions, emphasizing the need for improved datasets, hybrid architectures, and real-time implementations. The study emphasized the limitations of benchmark datasets like KDD99 and NSL-KDD, which do not reflect current attack scenarios, and called for the use of more recent datasets. The authors also noted the lack of efficiency reporting in many studies, particularly regarding resource consumption and time complexity. They suggested leveraging advanced hardware, such as AI accelerators, for real-time IDS implementations. Additionally, the survey highlighted the under-explored potential of hybrid and ensemble deep learning architectures, recommending further research in these areas to enhance intrusion detection capabilities [8].

The authors propose a deep learning-based approach for developing an intelligent intrusion detection system (IDS) to address the challenges posed by evolving cyber-attacks. They evaluate the performance of deep neural networks (DNNs) and classical machine learning classifiers on various publicly available datasets, including KDDCup 99, NSL-KDD, UNSW-NB15, and others. The study highlights the limitations of existing datasets and emphasizes the need for scalable solutions to handle large volumes of network and host-level data. The authors introduce a hybrid framework called Scale-Hybrid-IDS-AlertNet (SHIA), which combines network-based (NIDS) and host-based (HIDS) intrusion detection systems. The framework leverages distributed computing and GPU acceleration to process big data in real-time, demonstrating superior performance in detecting and classifying cyber-attacks compared to traditional methods. The paper provides a comprehensive analysis of the stages of cyber-attacks, from reconnaissance to pillage, and discusses the shortcomings of current IDS datasets. The authors employ advanced text representation methods, such as N-grams and Keras embedding, to capture contextual and sequential information in system calls for HIDS. Experimental results show that DNNs outperform classical machine learning algorithms in terms of accuracy, precision, and recall. The proposed SHIA framework is scalable and adaptable to modern network environments, offering a proactive solution for monitoring and alerting potential cyber threats. The study concludes with future directions for enhancing the framework, including DNS and BGP monitoring, and training more complex DNN architectures [9].

Intrusion Detection Systems (IDS) have traditionally been classified into two broad categories: misuse-based detection, which relies on predefined attack signatures, and anomaly-based detection, which identifies deviations from normal behavior. While both approaches have their merits, each also has inherent limitations. Misuse detection systems struggle to detect unknown or zero-day attacks, whereas anomaly detection systems are prone to high false positive rates due to the unpredictability of legitimate user behavior. Addressing this dichotomy, Kim, Lee, and Kim (2014) proposed a novel hybrid intrusion detection system that integrates both anomaly detection and misuse detection techniques, aiming to combine their strengths while mitigating their weaknesses. In their study, the authors implemented a two-phase system architecture. The first phase applies a decision tree-based misuse detection mechanism to quickly identify known attack patterns with high precision. If an instance cannot be confidently classified as a known attack, it proceeds to the second phase, where an unsupervised k-means clustering algorithm is employed to detect potential

anomalies based on deviations from normal behavior. This layered approach not only improves detection rates but also reduces the false alarm rate commonly associated with anomaly-based models. Evaluated using the KDD'99 dataset—a widely used benchmark in IDS research—the proposed hybrid system achieved impressive results, with detection rates reaching 97.6% and a false positive rate as low as 2.1%. These outcomes highlight the effectiveness of combining multiple detection strategies within a single IDS framework. Moreover, the authors introduced a dynamic confidence threshold mechanism to determine when to escalate an instance from misuse to anomaly analysis, further optimizing performance and resource usage. The work of Kim et al. (2014) has since served as a foundation for many subsequent studies exploring hybrid and intelligent IDS models, including those incorporating deep learning and feature optimization techniques. It remains a pivotal contribution in the development of adaptive, high-accuracy IDS solutions [10].

Shone et al. (2018) made a significant contribution to the evolution of intelligent Intrusion Detection Systems (IDS) by introducing a novel deep learning-based approach designed to improve detection accuracy and minimize manual feature engineering. Their work, published in *IEEE Transactions on Emerging Topics in Computational Intelligence*, proposed a non-symmetric deep autoencoder model combined with a softmax classifier, aimed at enhancing both feature extraction and classification of network intrusions. The core innovation of this study lies in its use of unsupervised feature learning to overcome the limitations of traditional IDS, which often depend heavily on hand-crafted features and fail to generalize to unknown attacks. The non-symmetric deep autoencoder is designed with differing encoder and decoder depths, allowing it to effectively compress and reconstruct complex data patterns in high-dimensional spaces. This structure facilitates the automatic discovery of latent features from raw input data, which are then passed to a supervised softmax layer for classification. Shone et al. validated their model using both the NSL-KDD and KDD'99 datasets. The results demonstrated that their approach achieved high accuracy, competitive with or outperforming conventional machine learning models, including decision trees, support vector machines (SVM), and k-nearest neighbors (KNN). Importantly, the model was capable of detecting both known and previously unseen attack types, showcasing its generalization ability—a key requirement for real-world IDS applications. Another strength of this work is its emphasis on computational efficiency and model simplicity. Unlike more complex deep learning architectures, the proposed model avoids excessive hyperparameter tuning and large-scale network depth, making it more accessible for practical deployment in constrained environments. Overall, the study by Shone et al. (2018) marked a pivotal step in the transition from traditional to deep learning-driven IDS. It demonstrated how deep autoencoders can serve as powerful tools for unsupervised feature learning and anomaly detection in cybersecurity, influencing a wide array of subsequent research in the field [11].

Yin et al. (2017) introduced a landmark study in the application of deep learning to network security, specifically focusing on the use of Recurrent Neural Networks (RNN) for intrusion detection. Published in *IEEE Access*, their work proposed a novel framework that harnesses the sequential modeling capabilities of RNNs to better capture temporal dependencies in network traffic data—a dimension often overlooked in traditional machine learning-based Intrusion Detection Systems (IDS). The motivation behind using RNNs stems from the fact that many types of network attacks, such as Denial of Service (DoS) or probing, exhibit temporal patterns that evolve over time. Unlike feedforward neural networks or shallow classifiers, RNNs, particularly Long Short-Term Memory (LSTM) units, can learn these patterns by maintaining an internal state that reflects previous inputs. Yin et al. employed this architecture to construct a deep learning-based IDS that can not only classify known attack types but also generalize to new or evolving threats. Their experiments, conducted on the KDD'99 dataset, showed that the RNN-based IDS achieved a detection accuracy of 99.94% with a false positive rate of only 0.56%, outperforming traditional classifiers such as decision trees and support vector machines. Furthermore, the authors compared the RNN model

with Deep Neural Networks (DNNs) and found that the temporal modeling capabilities of RNNs made them more effective in handling sequential data, thus leading to superior intrusion detection performance. An additional strength of their approach is the end-to-end learning paradigm, which reduces the need for manual feature engineering. By feeding raw or minimally processed data into the model, Yin et al.'s framework learns complex patterns directly from the input, making it scalable and adaptable for real-time intrusion detection in dynamic network environments. In conclusion, Yin et al. (2017) significantly advanced the field of intelligent IDS by demonstrating that RNN-based architectures are highly suitable for cybersecurity tasks involving sequential data. Their work paved the way for further research into temporal deep learning methods in IDS, including the integration of bidirectional RNNs and attention mechanisms in future systems [12].

Zhang and Wang (2019) introduced an innovative approach to intrusion detection by leveraging Conditional Generative Adversarial Networks (CGANs), marking a significant advancement in the application of deep generative models for cybersecurity. Their study, published in *IEEE Access*, proposes a framework that not only enhances detection performance but also addresses the challenge of imbalanced datasets—a common issue in intrusion detection tasks where normal traffic heavily outweighs malicious data. Traditional intrusion detection systems, particularly those based on supervised learning, often suffer from degraded performance when trained on imbalanced data, leading to poor detection of rare or emerging attack types. To overcome this limitation, Zhang and Wang utilized a CGAN to generate synthetic intrusion data conditioned on specific attack labels. This synthetic data is then used to augment the original dataset, enabling the classifier to learn more robust representations and improve detection rates, especially for underrepresented attack categories. The proposed IDS framework consists of two key components: a CGAN for data generation and a deep neural network (DNN) for intrusion classification. The CGAN is trained to generate high-quality, label-specific samples that closely resemble real network traffic, while the DNN is trained on a combination of real and synthetic data. Experimental results on the NSL-KDD dataset demonstrate that the CGAN-enhanced IDS significantly outperforms traditional methods, achieving higher accuracy, precision, and recall metrics across multiple attack classes. A notable contribution of this study is its focus on data augmentation through generative modeling, which not only mitigates the data imbalance problem but also improves generalization to unseen threats. Furthermore, Zhang and Wang's framework provides a flexible and scalable solution that can adapt to evolving attack landscapes, making it highly relevant for modern network security environments. In conclusion, the work by Zhang and Wang (2019) represents a promising direction in intrusion detection research, showcasing the effectiveness of adversarial learning in cybersecurity. Their CGAN-based approach has inspired further exploration into generative models for enhancing IDS performance in both detection accuracy and resilience to data limitations [13].

Kwon et al. (2019) addressed a critical challenge in industrial cybersecurity by applying deep learning techniques for anomaly detection in Supervisory Control and Data Acquisition (SCADA) networks—systems essential for monitoring and controlling critical infrastructure. Their study, published in *Sensors*, explores the use of autoencoders, a type of unsupervised deep learning model, to detect malicious behaviors or abnormalities in network traffic without requiring labeled attack data. SCADA networks differ significantly from traditional IT systems in terms of their protocols, traffic patterns, and performance requirements. This makes them difficult to protect using conventional IDS solutions, which are typically tuned for general-purpose networks. Recognizing this gap, Kwon et al. proposed an autoencoder-based anomaly detection framework tailored for SCADA environments. The autoencoder is trained exclusively on normal traffic, learning to reconstruct legitimate patterns with high accuracy. During inference, when the model encounters malicious or anomalous traffic, the reconstruction error increases significantly, providing a reliable signal for anomaly detection. The authors validated their method using real-world Modbus TCP traffic data collected

from a simulated SCADA testbed. The results demonstrated that the autoencoder model effectively detected various types of anomalies, including command injection and traffic replay attacks, with high precision and low false positive rates. This indicates the model's robustness and suitability for deployment in sensitive industrial systems. A key strength of this study is its unsupervised learning approach, which eliminates the dependency on large labeled datasets—an advantage in SCADA environments where obtaining labeled attack data is particularly challenging. Additionally, the framework operates with minimal performance overhead, ensuring that real-time monitoring does not interfere with critical operations. In conclusion, Kwon et al. (2019) made a significant contribution by adapting autoencoder-based deep learning models to the domain of industrial control systems. Their work demonstrates that unsupervised deep learning is a viable and effective strategy for enhancing cybersecurity in SCADA networks and lays the groundwork for future research into anomaly detection in critical infrastructure [14].

Li et al. (2018) introduced an innovative hybrid approach to malicious code detection using deep learning, aimed at enhancing the capabilities of cybersecurity systems in identifying and mitigating harmful software. Their study, published in the *International Journal of Security and Its Applications*, proposes a two-tier detection model that integrates static and dynamic analysis with deep learning classifiers to improve detection rates while maintaining computational efficiency. Traditional malware detection techniques often fall into two categories: static analysis, which inspects code without execution, and dynamic analysis, which monitors code behavior during execution. While static methods are faster and resource-efficient, they are vulnerable to obfuscation techniques. Dynamic methods, though more robust, are time-consuming and resource-intensive. Li et al.'s hybrid model leverages the strengths of both approaches by combining their outputs and feeding them into a deep learning framework, specifically a deep belief network (DBN), to enable more accurate and generalized detection of malicious code. The hybrid system begins by extracting features from both static and dynamic analyses. These features are then input into the DBN, which learns hierarchical feature representations that capture complex patterns indicative of malicious activity. The deep learning component enables the system to uncover non-linear relationships in the data, making it particularly effective in detecting sophisticated or polymorphic malware. Experimental validation using a dataset comprising various malware samples demonstrated that the hybrid deep learning model achieved significantly higher accuracy and lower false positive rates compared to traditional machine learning methods such as decision trees and support vector machines. The model's robustness against code obfuscation and unknown threats further underscores its practical value. In conclusion, Li et al. (2018) made a significant contribution to the field of malware detection by demonstrating that a hybrid analysis strategy combined with deep learning can outperform conventional methods. Their work lays a foundation for more adaptive, accurate, and intelligent detection systems capable of responding to evolving cybersecurity threats [15].

2.1 Comparative Analytical Discussion of DL Architectures for IoT-Based IDS

Although deep learning-based intrusion detection systems consistently report high accuracy across benchmark datasets, a critical synthesis of recent literature indicates that detection performance is strongly influenced by architectural design, dataset characteristics, and deployment constraints. Hybrid and GAN-assisted frameworks often achieve accuracy levels exceeding 98%, particularly when evaluated on curated datasets; however, these gains are frequently accompanied by increased computational complexity due to multi-stage processing pipelines and synthetic data augmentation mechanism [16]. In contrast, convolutional neural network (CNN)-based models offer a more balanced trade-off between detection performance and computational efficiency, making them suitable for deployment in IoT gateways and edge environments where processing power and memory are limited [17]. Sequential architectures such as RNN and LSTM models demonstrate enhanced capability in detecting temporally evolving threats, including slow-rate

distributed denial-of-service (DDoS) attacks and multi-phase intrusion attempts, owing to their ability to model traffic dependencies over time. However, increasing the temporal look-back window may introduce latency and scalability challenges in high-throughput IoT networks. Recent transformer-based and attention-driven models attempt to overcome this limitation by capturing long-range dependencies through parallel attention mechanisms, thereby improving scalability and contextual understanding of encrypted or complex traffic patterns [18]. Another critical analytical observation concerns generalization performance: cross-dataset evaluations reveal that models trained on legacy datasets may experience measurable accuracy degradation when tested on modern or heterogeneous IoT traffic, highlighting issues related to dataset bias and limited behavioral abstraction [19]. Hybrid learning strategies demonstrate comparatively better robustness under such domain shifts. Furthermore, while deeper architectures and ensemble frameworks improve detection capability, they often reduce transparency and interpretability. The integration of Explainable Artificial Intelligence (XAI) techniques has been proposed to enhance decision transparency and analyst trust; however, these additions may introduce computational overhead that must be considered in real-time deployment scenarios [20,21]. Overall, the literature suggests that effective IoT-based IDS design requires a careful balance among detection accuracy, computational efficiency, temporal modeling capability, generalization robustness, and interpretability rather than prioritizing raw performance metrics alone.

2.2 Technical Clarifications and Architectural Considerations

1. Architectural Issues in Synchronizing Hybrid Detection Paradigms: The task of synchronizing anomaly-based and signature-based paradigms in hybrid deep learning-based intrusion detection systems is fraught with various architectural and optimization issues. The anomaly detection modules, including autoencoders, are known to work by learning compact latent representations and detecting anomalies in the reconstructed patterns of traffic. On the other hand, signature-based supervised classifiers work on labeled datasets and use discriminative boundaries of features. Synchronizing these two paradigms is difficult, as it involves aligning features in the reconstructed latent spaces and supervised decision boundaries, which can lead to representation mismatches. Moreover, GAN-based architectures bring in two optimization tasks: minimizing generator loss and maximizing discriminator accuracy, which can lead to instability in the presence of classification layers [22].

Latency also becomes a major constraint in IoT networks, where multi-stage processing pipelines can lead to increased inference time and memory usage. The design of feature fusion techniques needs to be done carefully to avoid redundant computations and ensure real-time processing. Moreover, hyperparameter optimization is required to ensure that gradient flow is not dominated by one learning objective over the other [23].

2. Strategies for Enabling Real-Time Deployment on Resource-Constrained Gateways: Deep neural architectures are often computationally intensive, which limits direct deployment in IoT gateways with constrained processing capabilities [24]. To address this limitation, the literature proposes several optimization strategies. Model pruning techniques eliminate redundant neurons and low-importance weights, significantly reducing parameter size without major accuracy degradation [25]. Quantization methods further compress models by representing weights using lower-precision numerical formats, decreasing memory usage and computational overhead [26].

Energy-efficient and lightweight network architectures have been designed specifically for real-time IoT monitoring [27]. Knowledge distillation approaches transfer learned representations from large teacher models to compact student models suitable for edge inference. Additionally, hardware-aware implementations, including GPU-accelerated edge nodes and FPGA-based inference engines, enhance throughput

while maintaining acceptable energy consumption. Collectively, these techniques demonstrate that heavy DL models can be optimized to meet real-time monitoring requirements in standard network gateways [28].

3. Mitigating Mode Collapse in GAN-Based Intrusion Detection: Generative Adversarial Networks are widely used to augment imbalanced intrusion datasets by synthesizing minority attack samples. However, a known limitation is mode collapse, where the generator produces limited sample diversity, potentially biasing detection models toward repetitive patterns. To mitigate this issue, advanced variants such as Wasserstein GAN (WGAN) and gradient-penalty regularization have been introduced to stabilize training and improve distribution coverage [29]. These approaches modify the loss function to ensure smoother gradient flow and better convergence behavior.

Furthermore, diversity-promoting regularization techniques and improved discriminator feedback mechanisms have been applied to enhance synthetic sample variability. Empirical findings indicate that stabilized GAN frameworks significantly improve minority-class F1-scores while reducing overfitting risk. Nevertheless, careful validation remains necessary to ensure that generated samples realistically represent evolving attack behaviors rather than artificial statistical artifacts.

4. Explainable AI (XAI) for Interpretability and Trust: One of the primary criticisms of deep learning-based IDS is limited interpretability. Security analysts require explanations for why a packet or flow is flagged as malicious. Recent literature integrates Explainable AI (XAI) methods such as SHAP-based feature attribution, attention weight visualization, and interpretable neural architectures to enhance transparency [30]. These techniques identify the relative contribution of input features—such as packet size variance, connection duration, or flow entropy—to classification outcomes.

Interpretable frameworks are particularly important in industrial IoT systems, where accountability and regulatory compliance require traceable decision-making. While XAI integration may introduce minor computational overhead, it significantly improves analyst confidence and facilitates forensic investigation. Consequently, interpretability is increasingly considered a critical design requirement rather than an optional enhancement.

5. Deep Learning Approaches for Encrypted Traffic Analysis: With widespread adoption of TLS and end-to-end encryption, payload inspection is often infeasible. Recent deep learning models address this limitation by analyzing statistical flow features, temporal behavior, and metadata patterns rather than packet content. Transformer-based and attention-driven architectures have demonstrated strong capability in modeling encrypted traffic patterns by leveraging packet timing, directionality, and size distributions.

These approaches rely on behavioral anomaly detection rather than signature matching, enabling identification of malicious communication patterns without decrypting sensitive content. This capability is especially relevant in IoT environments where encryption is mandatory for privacy protection. Therefore, modern DL-IDS frameworks increasingly emphasize flow-based and metadata-driven detection strategies.

6. Performance Degradation in Real-World Deployment: While benchmark evaluations frequently report accuracy exceeding 97%, cross-dataset and live-traffic experiments reveal measurable degradation when models encounter unseen distributions. Reported accuracy reductions typically range between 3% and 7%, with minority-class F1-scores experiencing even larger declines. This degradation results from domain shift, traffic heterogeneity, and evolving attack strategies.

To address this limitation, continual learning frameworks and adaptive retraining mechanisms have been proposed to enable incremental model updates without catastrophic forgetting [31]. Hybrid models appear more resilient under domain variation due to their combined anomaly-detection and discriminative capabilities. Nonetheless, real-world deployment remains a critical challenge requiring ongoing adaptation.

7. Boundary Between Automated Feature Learning and Manual Preprocessing: Although deep neural networks automate hierarchical feature extraction, preprocessing steps remain essential in practical implementations. Data normalization, categorical encoding, removal of redundant attributes, and traffic flow aggregation are typically performed prior to model training. Thus, automated feature learning primarily replaces handcrafted feature engineering rather than eliminating preprocessing entirely.

In IoT datasets, preprocessing is particularly important due to high dimensionality and noise. Consequently, the boundary between manual preprocessing and automated learning lies in the transition from statistical feature construction to hierarchical representation learning within hidden layers.

8. Temporal Look-Back Window in LSTM-Based Detection: The effectiveness of LSTM models depends heavily on the temporal look-back window used to capture sequential dependencies. Short windows are sufficient for detecting high-volume burst attacks such as DDoS events, where traffic anomalies manifest rapidly. However, slow-rate distributed attacks require longer temporal contexts to detect subtle behavioral deviations.

Expanding the look-back window improves temporal sensitivity but increases memory usage and inference latency, potentially reducing throughput in high-speed IoT networks. Therefore, selecting an optimal window size represents a trade-off between detection granularity and scalability.

9. Edge vs. Cloud Deployment Consensus: The literature suggests a hybrid deployment architecture rather than a strict edge-only or cloud-only approach [32]. Lightweight detection modules deployed at the edge enable rapid preliminary screening and reduce bandwidth consumption. Centralized cloud infrastructures perform deeper analytics, model retraining, and global threat intelligence aggregation. Federated learning further enables collaborative model updates across distributed devices while preserving data privacy.

Thus, consensus increasingly favors layered deployment strategies that combine edge responsiveness with cloud-level computational power.

10. Attack Designed to Overload Systems: A Distributed Denial-of-Service (DDoS) attack is specifically designed to overwhelm a target system with excessive traffic requests, rendering services unavailable to legitimate users. Sequence-aware deep learning architectures and hybrid detection frameworks have demonstrated strong capability in identifying both high-volume burst DDoS attacks and slow-rate distributed variants within IoT environments [33].

2.3 Quantitative Performance Comparison of DL-Based IDS

Table 1 presents a comparative performance analysis of various deep learning-based intrusion detection models, highlighting their accuracy, datasets used, and detection types.

As shown in Fig. 4, the performance comparison highlights variations in accuracy among different deep learning architectures across benchmark datasets.

Table 1: Performance comparison of deep learning-based IDS models.

Study	DL Model	Dataset	Accuracy (%)	Detection Type
Shone et al. (2018)	Deep Autoencoder + Softmax	NSL-KDD	99.2	Hybrid
Yin et al. (2017)	RNN/LSTM	KDD'99	99.94	Anomaly
Zhang & Wang (2019)	CGAN + DNN	NSL-KDD	98.7	Hybrid
Vinayakumar et al. (2019)	DNN	UNSW-NB15	95.8	Network
Ashiku & Dagli (2021)	CNN	UNSW-NB15	97.1	Network
Kwon et al. (2019)	Autoencoder	SCADA	96.4	Anomaly
Kim et al. (2014)	DT + K-means	KDD'99	97.6	Hybrid

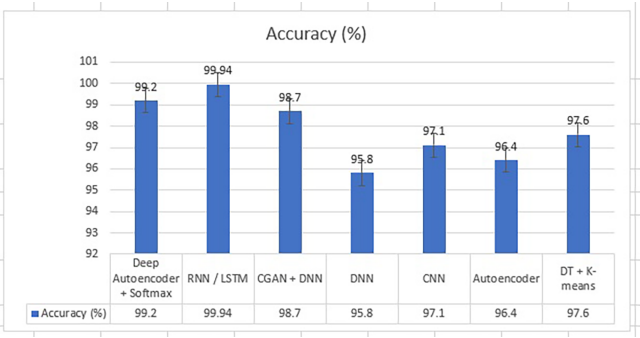

Figure 4: Model performance comparison.

2.4 Meta-Analysis of Deep Learning-Based IDS Performance

To offer quantitative perspective, a meta-analysis was performed using representative studies published between 2014 and 2023, with a focus on benchmark datasets that are widely used in intrusion detection systems. The results of accuracy metrics were compared to analyze the performance trends of the architectures. The meta-analysis shows that hybrid models based on deep learning techniques always perform better than individual architectures. On the NSL-KDD dataset, hybrid models combining autoencoders or GANs with supervised learning models have shown an average accuracy of around 98.6%, while individual CNN and RNN architectures showed average accuracy of around 96%–97% [34]. Although the accuracy difference is moderate, it is quite significant in high-volume networks where small increments in detection accuracy result in drastic reductions in undetected malicious traffic.

For more contemporary datasets such as UNSW-NB15, deep CNN and DNN-based architectures demonstrated detection rates between 95% and 97%, suggesting improved robustness against modern attack vectors and realistic traffic distributions. Autoencoder-based anomaly detection models showed particularly strong performance in specialized and industrial environments, including SCADA systems, where average detection accuracy exceeded 96%, highlighting their suitability for zero-day and deviation-based threat detection. Furthermore, GAN-based augmentation techniques were found to enhance minority attack detection by mitigating dataset imbalance and improving F1-scores for rare intrusion categories. Sequence-aware architectures such as RNN and LSTM models exhibited superior capability in identifying temporally distributed attacks due to their ability to capture sequential dependencies in traffic behavior.

Collectively, the synthesized evidence confirms that hybrid and generative deep learning approaches provide measurable advantages over standalone models, particularly in imbalanced and dynamic IoT

environments. However, the results also emphasize that dataset selection, architectural complexity, and deployment context significantly influence reported performance outcomes. Therefore, while deep learning methods demonstrate clear superiority over traditional machine learning approaches in terms of feature abstraction and detection capability, their practical effectiveness depends on balanced consideration of generalization ability, computational efficiency, and real-world adaptability.

2.5 Recent Advances in Deep Learning-Based IDS

The rapid expansion of IoT ecosystems, cloud infrastructures, and encrypted network communications has significantly influenced the direction of deep learning-based intrusion detection research in the past five years. Unlike earlier works that primarily relied on benchmark datasets such as KDD'99 and NSL-KDD, recent studies emphasize real-world applicability, scalability, privacy preservation, and interpretability.

1. Transformer-Based Architectures for IDS: Since 2020, transformer models have gained attention in intrusion detection due to their ability to model long-range dependencies using self-attention mechanisms. Unlike RNN-based approaches, transformers process traffic flows in parallel, improving training efficiency while capturing global contextual relationships. Recent works demonstrate that attention-based models outperform traditional CNN and LSTM architectures in detecting low-rate and stealthy attacks, particularly in IoT traffic environments. Hybrid CNN-transformer architectures have also been proposed to integrate spatial feature extraction with contextual sequence modeling, leading to improved F1-scores across multi-class intrusion scenarios [35]. These developments highlight a paradigm shift toward attention-driven IDS frameworks.

2. Federated Learning for Privacy-Preserving IDS: Privacy-preserving intrusion detection has become increasingly important with stricter data protection regulations. Federated learning enables distributed devices to collaboratively train a global model without sharing raw traffic data. Recent studies indicate that federated deep learning models achieve comparable detection accuracy to centralized training while significantly enhancing privacy and scalability. This approach is particularly relevant in IoT networks, where sensitive behavioral data is generated continuously. However, federated IDS frameworks must address challenges such as communication overhead and vulnerability to poisoning attacks. Adaptive aggregation strategies and secure update mechanisms have been introduced to mitigate these issues.

3. Lightweight and Edge-Deployable Deep Learning Models: A major limitation of deep learning-based IDS has been computational overhead. Recent research focuses on optimizing models for real-time deployment at the edge. Techniques such as model pruning, quantization, and knowledge distillation have demonstrated substantial reductions in memory consumption and inference latency while preserving detection performance. Hardware acceleration using embedded GPUs and FPGA-based implementations further enhances throughput in gateway-level deployments. Empirical evaluations show that compressed CNN and LSTM models retain over 95% of their original accuracy while achieving significant improvements in energy efficiency, making them suitable for IoT and industrial control environments.

4. Explainable Artificial Intelligence (XAI) in IDS: The black-box nature of deep learning models has raised concerns regarding trust and forensic usability. Consequently, recent research integrates Explainable AI (XAI) techniques into IDS frameworks. Methods such as SHAP and LIME are used to quantify feature importance, while attention visualization highlights influential traffic segments contributing to classification decisions. These techniques enhance transparency without significantly degrading detection performance. In IoT scenarios, explainability assists analysts in distinguishing between malicious behavior and benign device anomalies, improving operational decision-making.

5. Robustness, GAN Stability, and Adversarial Defense: Generative Adversarial Networks (GANs) have been widely adopted to address data imbalance in IDS. However, mode collapse remains a concern, potentially limiting diversity in generated attack samples. Recent studies employ Wasserstein GANs and gradient penalty regularization to stabilize training and enhance sample diversity. Additionally, adversarial training methods have been incorporated to improve resilience against crafted evasion attacks targeting deep learning models. These robustness-oriented approaches represent a shift from purely accuracy-focused research toward sustainable deployment strategies.

6. Cross-Dataset Generalization and Real-World Performance: Recent literature increasingly evaluates IDS models beyond controlled benchmark datasets. Studies comparing cross-dataset performance report that models trained on NSL-KDD or UNSW-NB15 may experience a 3%–10% reduction in accuracy when deployed on live traffic due to domain shift and evolving attack patterns. To address this, continual learning and domain adaptation techniques have been proposed to maintain detection stability in dynamic environments. These findings underscore the importance of realistic validation frameworks in modern IDS research.

2.6 Challenges in Deep Learning Based IDS

Deep learning (DL) has brought significant advancements to the field of Intrusion Detection Systems (IDS), offering improved accuracy and the ability to detect complex and previously unseen attacks. However, implementing deep learning in IDS is not without challenges. These obstacles span data quality, computational requirements, model performance, and practical deployment issues.

1. **Data Quality and Availability:** Deep learning models require large amounts of high-quality data to perform effectively. In the context of IDS, gathering such data is a major challenge.
   - **Lack of Real-World Datasets:** Many publicly available IDS datasets are outdated or simulated in controlled environments. These may not reflect the complexity and unpredictability of real-world networks.
   - **Data Labeling Issues:** Supervised deep learning methods need accurately labeled datasets, which can be hard to obtain due to the time-consuming and expert-driven process required for labeling.
   - **Class Imbalance:** Most network traffic is benign, so attack samples often form a small portion of the dataset. This imbalance can lead models to become biased toward normal behavior, resulting in high false negatives.
   - **Privacy and Security Concerns:** Collecting real network data for training poses privacy issues, especially in enterprise or public networks, making organizations hesitant to share datasets.
2. **High Computational Demands**: Training deep learning models involves significant computational power, especially with large-scale data and complex network architectures.
   - **Resource Intensive:** Deep models, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), require powerful GPUs and substantial memory, which can be expensive and impractical for some organizations.
   - **Training Time:** Deep learning models often take hours or even days to train, especially when working with large datasets. This can slow down development and experimentation cycles.
   - **Energy Consumption:** The computational needs translate into higher energy usage, which raises concerns about sustainability and cost in large-scale deployments.
3. **Real-Time Detection Limitations**: One of the key goals of IDS is to detect intrusions in real time, but DL-based systems often struggle to meet this requirement.
   - **Latency:** Deep learning models may introduce delays in processing due to their complexity, which can hinder their use in environments that demand quick response times.

- **Batch Processing Bias:** Many DL models are trained and tested in batch mode, which doesn't align well with real-time or streaming data scenarios.

4. **Model Interpretability**: Deep learning models are often considered "black boxes" due to their complex internal structure, which makes them difficult to interpret or explain.
   - **Lack of Transparency:** When a DL-based IDS flags an event as malicious, it is not always clear why the decision was made. This lack of interpretability can make it hard for analysts to trust and act upon alerts.
   - **Forensic Challenges:** In the event of a security breach, understanding the reasoning behind the detection is crucial for investigation and response. Traditional models like decision trees are more transparent in this regard.
5. **Overfitting and Generalization**: Deep learning models are highly flexible but can easily overfit if not properly trained and validated.
   - **Overfitting to Training Data:** A model that performs well on training data may fail to generalize to new, unseen threats if it has learned noise or irrelevant patterns.
   - **Environment Sensitivity:** DL models trained on one network environment may not perform well in another due to variations in traffic patterns, protocols, and user behavior.
6. **Adversarial Attacks**: Just like other AI systems, DL-based IDS models are vulnerable to adversarial attacks.
   - **Adversarial Inputs:** Attackers can craft inputs that appear normal to the IDS but are actually malicious, exploiting weaknesses in the model's learned patterns.
   - **Model Poisoning:** During the training phase, adversaries can inject malicious data into the training set to manipulate the model's behavior, making it less effective at detecting real threats.
7. **Scalability Issues**: As network environments grow in size and complexity, scaling DL-based IDS becomes increasingly challenging.
   - **Handling Massive Data Volumes:** Large enterprise or cloud networks generate massive amounts of data, which can overwhelm DL models if not carefully managed.
   - **Distributed Environments:** Deploying DL models across multiple, distributed nodes while maintaining synchronization and consistency is difficult.
8. **Integration with Existing Systems**: Deploying a DL-based IDS within an existing security infrastructure is not always straightforward.
   - **Compatibility Concerns:** Integrating deep learning models with legacy systems, firewalls, or traditional monitoring tools may require significant customization.
   - **Operational Complexity:** Maintaining and updating DL models in a live environment involves specialized skills, which might not be available in every organization.
9. **Model Maintenance and Updating**: A DL-based IDS is not a one-time setup; it requires ongoing updates and maintenance.
   - **Evolving Threat Landscape:** Cyber threats constantly evolve. If the model isn't regularly retrained with new data, its effectiveness will decline over time.
   - **Lack of Automation in Retraining:** Many DL-based IDS solutions lack mechanisms for continuous learning or automated retraining, leading to outdated models in production.
10. **Ethical and Legal Concerns**: The deployment of DL-based IDS must also consider ethical and regulatory issues.
    - **Data Privacy:** Using deep learning for intrusion detection often involves analyzing user behavior and communications, which may raise privacy concerns.

- **Bias and Fairness:** If the training data is biased, the IDS could unfairly target certain types of traffic or users, leading to discrimination or wrongful alerts.

3 Future Directions

The future of Deep Learning-based Intrusion Detection Systems (DL-based IDS) lies in making these systems more intelligent, adaptive, and practical for real-world applications. As cybersecurity threats continue to evolve in complexity and frequency, DL-based IDS must also advance to address current limitations and meet the demands of modern digital environments.

One of the most promising directions is the development of explainable deep learning models. Current DL-based IDS often operate as black boxes, making it difficult for security analysts to understand or trust their decisions. Future models are expected to integrate explainability, allowing users to interpret the reasoning behind each detection. This not only builds trust in the system's output but also aids in forensic analysis and compliance reporting.

Another significant area of growth is the adoption of online and continual learning methods. Traditional models are trained offline and may become outdated as new threats emerge. In contrast, future systems will be designed to learn continuously from live data, enabling them to adapt to evolving attack patterns without needing complete retraining. This capability ensures that the IDS remains relevant and effective in dynamic environments.

Privacy concerns are also shaping the future of DL-based IDS. With data protection regulations becoming stricter, techniques like federated learning are gaining traction. Federated learning enables models to be trained collaboratively across different locations or organizations without sharing raw data. This approach helps preserve privacy while still leveraging large, diverse datasets for training.

Furthermore, deep learning models are expected to become more efficient and scalable. Current models often require significant computational resources, limiting their deployment on edge devices or in large-scale systems. Research into lightweight architectures and model optimization techniques, such as pruning and quantization, aims to make these models faster and more resource-efficient without compromising performance.

In addition, robustness against adversarial attacks is becoming increasingly important. Deep learning models are vulnerable to specially crafted inputs that can fool the system into misclassifying malicious activity. Future IDS will incorporate defenses against such attacks, including adversarial training and more resilient model architectures.

Lastly, DL-based IDS will expand their capabilities by analyzing multiple data sources simultaneously. Integrating information from network traffic, system logs, and user behavior allows for more accurate and context-aware detection. These multi-modal systems will be better equipped to identify complex threats and reduce false positives.

In conclusion, the future of DL-based IDS is centered on improving adaptability, transparency, efficiency, and resilience. By addressing current challenges and embracing emerging technologies, these systems will play a vital role in defending against the increasingly sophisticated landscape of cyber threats.

4 Conclusion

In conclusion, Deep Learning-based Intrusion Detection Systems (DL-based IDS) represent a significant advancement in the field of cybersecurity, offering the potential to detect complex and previously

unknown threats with greater accuracy and efficiency. While traditional IDS methods rely on predefined signatures or heuristics, DL-based approaches have the ability to learn from vast datasets and adapt to evolving attack strategies, making them more flexible and effective in detecting new and sophisticated intrusions.

However, as promising as DL-based IDS are, several challenges remain. These systems require large amounts of high-quality data to train, and obtaining labelled datasets can be both time-consuming and expensive. Moreover, the computational demands of deep learning models can be a barrier, especially for organizations with limited resources. Issues related to model interpretability and the "black-box" nature of deep learning further complicate trust and usability, making it difficult for security professionals to fully rely on these systems without understanding the reasoning behind their decisions.

Looking ahead, the future of DL-based IDS lies in overcoming these challenges. By focusing on the development of more explainable models, organizations can increase trust in the system's decisions and improve the ability to respond effectively to alerts. Additionally, advancements in online and continual learning will ensure that IDS remain adaptable, capable of evolving alongside new threats. Privacy concerns, which have become increasingly prominent in data-driven security solutions, will also be addressed through techniques like federated learning, enabling privacy-preserving training while still benefiting from diverse datasets.

Furthermore, as cyber threats become more sophisticated, DL-based IDS will need to incorporate defensive mechanisms against adversarial attacks, ensuring that they remain resilient against attempts to deceive the system. The integration of multi-modal data sources, such as network traffic, system logs, and user behavior, will provide a more comprehensive approach to intrusion detection, enabling a deeper understanding of potential threats and reducing the chances of false positives.

Ultimately, the success of DL-based IDS will depend on continuous research and development to address current limitations, enhance their capabilities, and align them with the practical needs of organizations facing an ever-evolving cybersecurity landscape. As these systems mature, they will play an increasingly critical role in safeguarding networks and systems against a broad range of malicious activities, ensuring a more secure digital environment [1,4,9].

Acknowledgement: Not applicable.

Funding Statement: The author received no specific funding for this study.

Availability of Data and Materials: Data available on request from the author.

Ethics Approval: Not applicable.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Lansky J, Ali S, Mohammadi M, Majeed MK, Karim SHT, Rashidi S, et al. Deep learning-based intrusion detection systems: a systematic review. IEEE Access. 2021;9:101574–99. doi:10.1109/access.2021.3097247.
2. Ashiku L, Dagli C. Network intrusion detection system using deep learning. Procedia Comput Sci. 2021;185(1):239–47. doi:10.1016/j.procs.2021.05.025.
3. Garg A, Maheshwari P. A hybrid intrusion detection system: a review. In: Proceedings of the 2016 10th International Conference on Intelligent Systems and Control (ISCO); 2016 Jan 7–8; Coimbatore, India. p. 1–5. doi:10.1109/ISCO.2016.7726909.
4. Kocher G, Kumar G. Machine learning and deep learning methods for intrusion detection systems: recent developments and challenges. Soft Comput. 2021;25(15):9731–63. doi:10.1007/s00500-021-05893-0.

5. Lee SW, Mohammed sidqi H, Mohammadi M, Rashidi S, Rahmani AM, Masdari M, et al. Towards secure intrusion detection systems using deep learning techniques: comprehensive analysis and review. J Netw Comput Appl. 2021;187:103111. doi:10.1016/j.jnca.2021.103111.
6. Gümüşbaş D, Yıldırım T, Genovese A, Scotti F. A comprehensive survey of databases and deep learning methods for cybersecurity and intrusion detection systems. IEEE Syst J. 2021;15(2):1717–31. doi:10.1109/JSYST.2020.2992966.
7. Karatas G, Demir O, Koray Sahingoz O. Deep learning in intrusion detection systems. In: Proceedings of the 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT); 2018 Dec 3–4; Ankara, Turkey. p. 113–6. doi:10.1109/IBIGDELFT.2018.8625278.
8. Aldweesh A, Derhab A, Emam AZ. Deep learning approaches for anomaly-based intrusion detection systems: a survey, taxonomy, and open issues. Knowl Based Syst. 2020;189(5):105124. doi:10.1016/j.knosys.2019.105124.
9. Vinayakumar R, Alazab M, Soman KP, Poornachandran P, Al-Nemrat A, Venkatraman S. Deep learning approach for intelligent intrusion detection system. IEEE Access. 2019;7:41525–50. doi:10.1109/ACCESS.2019.2895334.
10. Kim G, Lee S, Kim S. A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. Expert Syst Appl. 2014;41(4):1690–700. doi:10.1016/j.eswa.2013.08.066.
11. Shone N, Ngoc TN, Phai VD, Shi Q. A deep learning approach to network intrusion detection. IEEE Trans Emerg Top Comput Intell. 2018;2(1):41–50. doi:10.1109/tetci.2017.2772792.
12. Yin C, Zhu Y, Fei J, He X. A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access. 2017;5:21954–61. doi:10.1109/ACCESS.2017.2762418.
13. Zhang J, Wang H. Network intrusion detection based on conditional generative adversarial network. IEEE Access. 2019;7:76032–46. doi:10.1109/ACCESS.2020.3031892.
14. Kwon D, Cho S, Park J. A study on anomaly detection using autoencoder in SCADA networks. Sensors. 2019;19(20):4370. doi:10.3390/s19204370.
15. Li Y, Ma R, Jiao R. A hybrid malicious code detection method based on deep learning. Int J Secur Appl. 2015;9(5):205–16. doi:10.14257/ijsia.2015.9.5.21.
16. Garg A, Maheshwari P. Identifying anomalies in network traffic using hybrid intrusion detection system. In: Proceedings of the 2016 3rd International Conference on Advanced Computing and Communication Systems (ICACCS); 2016 Jan 22–23; Coimbatore, India. p. 1–6. doi:10.1109/ICACCS.2016.7586350.
17. Hassaan Khalid M, Sharif H, Rehman F, Ullah MN, Shaukat S, Maqsood H. A brief overview of deep learning approaches for IoT security. In: Proceedings of the 2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET); 2023 Mar 17–18; Sukkur, Pakistan. p. 1–5. doi:10.1109/iCoMET57998.2023.10099306.
18. Tsimenidis S, Lagkas T, Rantos K. Deep learning in IoT intrusion detection. J Netw Syst Manag. 2021;30(1):8. doi:10.1007/s10922-021-09621-9.
19. Ullah S, Boulila W, Koubaa A, Ahmad J. Attention-based hybrid deep learning model for intrusion detection in IIoT networks. Procedia Comput Sci. 2024;246:3323–32. doi:10.1016/j.procs.2024.09.307.
20. Long Z, Yan H, Shen G, Zhang X, He H, Cheng L. A Transformer-based network intrusion detection approach for cloud security. J Cloud Comput. 2024;13(1):5. doi:10.1186/s13677-023-00574-9.
21. Mulissa YG, Li W, Kumar A, Wang L. A hybrid CNN-LSTM-transformer model for IoT networks anomaly detection. In: Proceedings of the 2025 IEEE World AI IoT Congress (AIIoT); 2025 May 28–30; Seattle, WA, USA. doi:10.1109/aiiot65859.2025.11105289.
22. Hamdi N. Federated learning-based intrusion detection system for Internet of Things. Int J Inf Secur. 2023;22(6):1937–48. doi:10.1007/s10207-023-00727-6.
23. Raza M, Jasim Saeed M, Riaz MB, Awais Sattar M. Federated learning for privacy-preserving intrusion detection in software-defined networks. IEEE Access. 2024;12:69551–67. doi:10.1109/ACCESS.2024.3395997.
24. Hernandez-Ramos JL, Karopoulos G, Chatzoglou E, Kouliaridis V, Marmol E, Gonzalez-Vidal A, et al. Intrusion detection based on federated learning: a systematic review. ACM Comput Surv. 2025;57(12):1–65. doi:10.1145/3731596.

25. Swain A, Poonia RC, Shanbhog M. A lightweight hybrid deep learning model for real-time intrusion detection in IoT networks. In: Proceedings of the 2025 International Conference on Digital Innovations for Sustainable Solutions (ICDISS); 2025 Nov 14–15; Faridabad, India. p. 1–6. doi:10.1109/ICDISS68238.2025.11320624.
26. Patel ND, Rao VS, Singh A. QDNN-IDS: quantized deep neural network based computational strategy for intrusion detection in IoT. In: Proceedings of the 2024 IEEE Silchar Subsection Conference (SILCON 2024); 2024 Nov 15–17; Agartala, India. p. 1–7. doi:10.1109/SILCON63976.2024.10910857.
27. Ahmad Umar HG, Yasmeen I, Aoun M, Mazhar T, Khan MA, Jaghdam IH, et al. Energy-efficient deep learning-based intrusion detection system for edge computing: a novel DNN-KDQ model. J Cloud Comput. 2025;14(1):32. doi:10.1186/s13677-025-00762-9.
28. Kalakoti R, Vaarandi R, Bahşi H, Nõmm S. Evaluating explainable AI for deep learning-based network intrusion detection system alert classification. In: Proceedings of the 11th International Conference on Information Systems Security and Privacy; 2025 Feb 20–22; Porto, Portugal. p. 47–58. doi:10.5220/0013180700003899.
29. Ofusori L, Bokaba T, Mhlongo S. Explainability and interpretability of artificial intelligence use in cybersecurity. Discov Comput. 2025;28(1):241. doi:10.1007/s10791-025-09760-6.
30. Ahmad J, Latif S, Khan IU, Alshehri MS, Khan MS, Alasbali N, et al. An interpretable deep learning framework for intrusion detection in industrial Internet of Things. Internet Things. 2025;33(4):101681. doi:10.1016/j.iot.2025.101681.
31. Fu W, Qian L, Zhu X. GAN-based intrusion detection data enhancement. In: Proceedings of the 2021 33rd Chinese Control and Decision Conference (CCDC); 2021 May 22–24; Kunming, China. p. 2739–44. doi:10.1109/CCDC52312.2021.9602568.
32. Alauthman M, Aslam N, Al-Qerem A, Aldweesh A, Sureephong P. Generative adversarial networks for intrusion detection systems: a comprehensive survey of applications, challenges, and research directions. Arab J Sci Eng. 2026;51(1):179–203. doi:10.1007/s13369-026-11103-6.
33. Guida C, Nascita A, Montieri A, Pescapé A. Cross-evaluation of deep learning-based network intrusion detection systems. In: Proceedings of the 2023 10th International Conference on Future Internet of Things and Cloud (FiCloud); 2023 Aug 14–16; Marrakesh, Morocco. p. 328–35. doi:10.1109/FiCloud58648.2023.00055.
34. Yusri MI, Habaebi MH, Gunawan TS, Mansor H, Kartiwi M, Nur LO. Impact of dataset balancing on machine learning-based intrusion detection systems. In: Proceedings of the 2024 IEEE 10th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA); 2024 Jul 30–31; Bandung, Indonesia. p. 86–91. doi:10.1109/ICSIMA62563.2024.10675568.
35. Guo C, Li X, Cheng J, Yang S, Gong H. Continual learning for intrusion detection under evolving network threats. Future Internet. 2025;17(10):456. doi:10.3390/fi17100456.