frontiers | Frontiers in Computer Science

TYPE Review
PUBLISHED 10 June 2024
DOI 10.3389/fcomp.2024.1387354

OPEN ACCESS

EDITED BY
Muhammad Adnan Khan,
Gachon University, Republic of Korea

REVIEWED BY
Hariprasath Manoharan,
Panimalar Institute of Technology, India
Adian Fatchur Rochim,
Diponegoro University, Indonesia

*CORRESPONDENCE
Boudour Ammar
✉ boudour.ammar@enis.usf.tn
Faisal Albalwy
✉ faisal.albalwy@manchester.ac.uk

RECEIVED 17 February 2024
ACCEPTED 28 May 2024
PUBLISHED 10 June 2024

CITATION
Ali AH, Charfeddine M, Ammar B, Hamed BB, Albalwy F, Alqarafi A and Hussain A (2024) Unveiling machine learning strategies and considerations in intrusion detection systems: a comprehensive survey.
Front. Comput. Sci. 6:1387354.
doi: 10.3389/fcomp.2024.1387354

Unveiling machine learning strategies and considerations in intrusion detection systems: a comprehensive survey

Ali Hussein Ali[1], Maha Charfeddine[2], Boudour Ammar[2]*, Bassem Ben Hamed[3], Faisal Albalwy[4,5]*, Abdulrahman Alqarafi[4] and Amir Hussain[6]

[1]REGIM-Lab: REsearch Groups in Intelligent Machines (REGIM), National School of Electronics and Telecommunications of Sfax, University of Sfax, Sfax, Tunisia, [2]REsearch Groups in Intelligent Machines (REGIM), National Engineering School of Sfax (ENIS), University of Sfax, Sfax, Tunisia, [3]Laboratory of Signals, systeMs, aRtificial Intelligence and neTworkS (SM@RTS), National School of Electronics and Telecommunications of Sfax, University of Sfax, Sfax, Tunisia, [4]Department of Computer Science, College of Computer Science and Engineering, Taibah University, Madinah, Saudi Arabia, [5]Division of Informatics, Imaging and Data Sciences, Stopford Building, University of Manchester, Manchester, United Kingdom, [6]Centre of AI and Robotics, Edinburgh Napier University, Edinburgh, United Kingdom

The advancement of communication and internet technology has brought risks to network security. Thus, Intrusion Detection Systems (IDS) was developed to combat malicious network attacks. However, IDSs still struggle with accuracy, false alarms, and detecting new intrusions. Therefore, organizations are using Machine Learning (ML) and Deep Learning (DL) algorithms in IDS for more accurate attack detection. This paper provides an overview of IDS, including its classes and methods, the detected attacks as well as the dataset, metrics, and performance indicators used. A thorough examination of recent publications on IDS-based solutions is conducted, evaluating their strengths and weaknesses, as well as a discussion of their potential implications, research challenges, and new trends. We believe that this comprehensive review paper covers the most recent advances and developments in ML and DL-based IDS, and also facilitates future research into the potential of emerging Artificial Intelligence (AI) to address the growing complexity of cybersecurity challenges.

KEYWORDS

intrusion detection system, network security, machine learning, deep learning, benchmark datasets

1 Introduction

The concern for network security has developed and is now an unavoidable issue. Many security reports and research papers show an annual increase in hostile actions (Mohammadi et al., 2021; Establishment, 2023). It has been observed that many attacks attempt to exploit system vulnerabilities to harm the confidentiality, integrity, and availability of data. Typical harmful behaviors include stealing users' accounts, gaining illegal access, capturing critical information, and blocking or rejecting services (Sumaiya Thaseen et al., 2021). Access control, encryption, authentication, and a sophisticated firewall are security procedures and techniques developed to detect and mitigate these threats. An intrusion detection system (IDS) is designed to address the inadequacies of other security solutions. There is an urgent need for a sophisticated IDS that can automatically detect known and unknown threats. The fundamental role

of an IDS is to monitor the exchange of data for suspicious behavior (Jatti and Sontif, 2019). Different approaches to designing IDS systems based on misuse detection, anomaly detection, or combining the two concepts have been presented in recent years. Because it looks to have additional major implications, anomaly detection is becoming more of a focus of investigation in network intrusion detection. Anomaly detection relies on statistics, expertise, and machine learning (ML) (Haji and Ameen, 2021; Prasath et al., 2022). The number of IDS deployments involving ML approaches has lately surged. The mode of learning deployed by various ML approaches allows for classifying these strategies into two major categories: supervised and unsupervised techniques (Hindy et al., 2020). The training and testing stages make up the total supervised learning process. The model is built with a labeled training set during the training process. The created model is tested for its capacity to produce accurate predictions, which results in a classification of the testing set instances. To learn from data, unsupervised learning does not require a training stage. It uses metrics to classify similar models into clusters. ML and DL aim to extract valuable information from massive data repositories. Predicting normal and aberrant behavior from learned patterns and monitoring network traffic are two of the most significant applications of ML (Alzahrani and Alenazi, 2021) and DL (Kim et al., 2021; Agrawal et al., 2022). Researchers have proposed several ML and DL-based IDS detection algorithms during the last decade. Much more study may be done on IDS to increase its ability to quickly and accurately identify network intrusions.

This paper comprehensively reviews recent advancements and trends in ML and DL-based IDS systems. It is an inventory of the most up-to-date research publications on intrusion detection, with a focus on the latest methodologies. The discussion focused on the prominent machine learning and deep learning algorithms, as well as the essential factors utilized for evaluating the outcomes. While previous survey articles have been published on machine learning-based intrusion detection systems (IDS), our research makes several novel contributions. We have carefully selected the recent research papers in intrusion detection to highlight the most advanced methods. Our analysis goes beyond simply listing papers; we investigated modern and widely used datasets, commonly used metrics and indicators, and studied and categorized major IDS-detected attacks. Furthermore, we covered important recent ML and DL-based IDS algorithms, as well as key parameters for evaluating their results. Moreover, we thoroughly addressed the challenges and potential advancements in ML and DL-based IDS systems. In addition, to evaluate our research, we compared it to other studies, identifying both similarities and differences between our methodology and previous surveys. This level of analysis provides an in-depth overview of the current landscape in ML and DL-based IDS research. By compiling the obtained research findings, our survey offers insight and direction for future studies in the field, ultimately contributing to the advancement of intrusion detection technology.

The following is the outline for the rest of the paper. The study's methodology is explained in Section 2. Section 3 introduces the fundamental IDS concepts and the various categorization algorithms. Section 4 reviews the DL and ML techniques in greater depth. Section five details the evaluation metrics and the performance indicators. Section 6 outlines the public datasets used as benchmarks. Section 7 discusses the most significant findings in ML and DL-based IDS, addresses the research challenges related to this subject and highlights novel trends and future directions. Section 8 compares our proposed study to other surveys on ML and DL-based IDS. The ninth section concludes this review article.

2 Methodology

This study covers the most important ML and DL-based IDS studies published recently in peer-reviewed papers since 2020. We find relevant articles, assess them, and gather important information. This analysis, for the most part, attempts to answer the following questions:

- How have AI-based intrusion detection systems evolved recently?
- What are the most popular and modern ML and DL approaches employed for IDS?
- What are the strengths, weaknesses and implications of ML and DL-based IDS systems?
- What datasets are commonly and recently utilized in AI-based IDS testing?
- Which metrics and performance indicators are most frequently used to evaluate performance?
- What are the research challenges, emerging trends, and expected future developments in ML and DL-based IDS systems?
- Are there any previous state-of-the-art surveys that have tackled this important research topic?
- What are the differences and similarities between our systematic approach and existing surveys, as well as the specific types of concerns we addressed in our survey?

The paper provides a comprehensive survey of the effectiveness of modern algorithms in intrusion detection, describing the most recent solutions, datasets, metrics, indicators, and approaches used. It is a useful resource for researchers working in these domains, addressing the challenge and emerging trends in ML and DL-based IDS systems, and thereby advancing intrusion detection technology.

3 IDS: concept and classification

This section provides a first exposition of the fundamental principles underlying IDSs, followed by information on how IDS are classified based on deployment and threat identification. Table 1 lists the abbreviations used in this article.

3.1 IDS concept

Dorothy E. Denning invented intrusion detection systems (IDS) in 1987 to detect network and computer attacks. IDS is a collection of approaches designed to detect suspicious, malicious, or unusual behavior that threatens the security of networks and computers (Oprea et al., 2021). Computer or network systems

TABLE 1 Meaning of acronyms.

Acronyms	Meanings	Acronyms	Meanings
AE	AutoEncoder	IDS	Intrusion Detection System
AI	Artificial Intelligence	ID3	Iterative Dichotomiser 3
AIDS	Anomaly Intrusion Detection System	IoT	Internet of Things
ANN	Artificial Neural Network	KNN	K-Nearest Neighbor
ACO	Ant Colony Optimization	LR	Linear Regression
CART	Classification and Regression Trees	LSTM	Long Short-Term Memory
CIC	Canadian Institute of Cybersecurity	ML	Machine Learning
CM	Confusion Matrix	NB	Naive Bayes
CNN	Convolutional Neural Network	NIDS	Network-based Intrusion Detection System
CSE	Communication Security Establishment	OCSVM	One-Class Support Vector Machine
DBN	Deep Belief Network	1D CAE	One-Dimensional Convolutional AutoEncoder
DL	Deep Learning	PSO	Particle Swarm Optimization
DNN	Deep Neural Network	RandNN	Random Neural Networks
DNN- SAVER	DNN-Supervised Adversarial Variational	R2L	Remote to Local
DoS	Denial of Service	ReLU	Rectified Linear Unit
DT	Decision Tree	RBM	Restricted Boltzmann Machine
ELM	Extreme Learning Machine	RF	Random Forest
FAR	False Alarm Rate	RNN	Recurrent Neural Network
FFNN	Feed Foeward Neural Netwoks	SMOTE	Synthetic Minority Over-Sampling Technique
FLN	Fast Learning Network	SIDS	Signature-based Intrusion Detection System
FN	False Negative	SVM	Support Vector Machine
FP	False Positive	SOM	Self-Organizing Maps
FSL	Few-shot Learning	TN	True Negative
GA	Genetic Algorithm	TNR	True Negative Rate
GAN	Generative Adversarial Networks	TP	True Positive
GRU	Gated Recurrent Unit	U2R	User to Root
GPU	Graphics Processing Unit	XGBoost	eXtreme Gradient-Boosting
HIDS	Host-based Intrusion Detection System		

intruders can jeopardize data security by modifying, destroying, or making information unavailable. Conversely, a detecting system is a preventive mechanism for identifying this illegal activity. IDS is software or hardware used to monitor computerized systems to detect intruders. Today, many commercial and open-source IDSs have varying capabilities depending on their components, such as the type of attack they can detect, their categories or classes, and their strategy (Wester, 2021).

Figure 1 depicts an IDS classification based on the detection approach used and its environments.

3.2 Categorization of IDS based on environment

Depending on what is being tracked, the IDS can be classified principally into two types: those operating on hosts and across networks. Other types of IDS include graph, application, distribution, and hypervisor-based IDS (Borkar et al., 2017).

3.2.1 Host-based IDS

HIDS is deployed on individual hosts or endpoints, such as servers, workstations, or network devices (Gu and Lu, 2021). HIDS monitors activities and events occurring on the host where it is deployed and requires installation of agents or sensors on each host to gather data. It is placed directly on the host or endpoint it is intended to protect. HIDS offers detailed insight into activities and events specific to the host where it is deployed. It is ideal for protecting critical servers or endpoints where detailed monitoring and analysis are required. Nonetheless, HIDS has some drawbacks. It uses many computer system resources, can interfere with operating systems and firewalls, and is difficult to maintain in large quantities networks (Panagiotou et al., 2021).

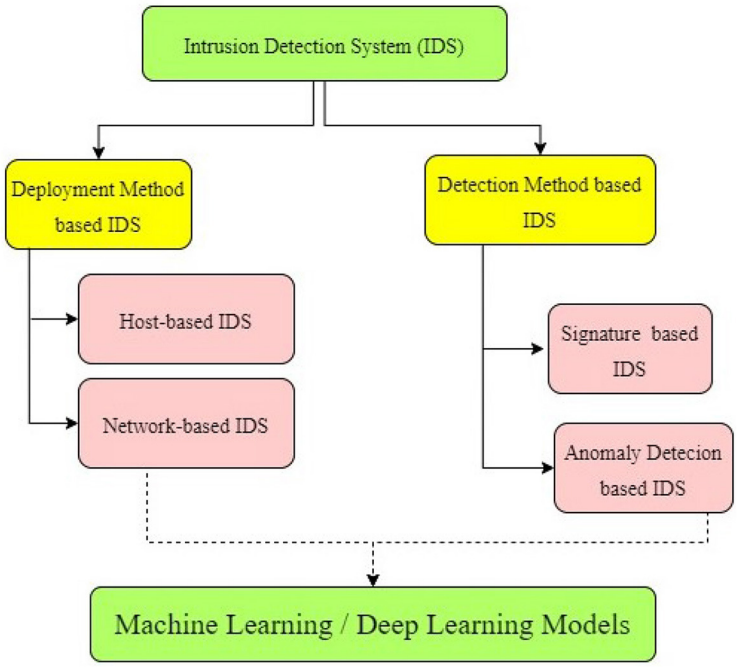

FIGURE 1
Classification of IDSs (Hindy et al., 2020).

3.2.2 Network-based IDS

NIDS is deployed on strategic points within the network infrastructure. It is positioned at network perimeter, internal segments, or critical chokepoints. It does not require agents on individual hosts, making deployment simpler and less intrusive. It analyzes and monitors the network traffic passing through designated points to detect suspicious or malicious activity (Borkar et al., 2017). Because it provides an up-to-date view of the entire network, deploying NIDS on the network influences its effectiveness (Sultana et al., 2019). NIDS combines two detection approaches for intrusions: misuse and anomaly. NIDS has some downsides: it cannot analyze encrypted packets, is susceptible to DoS attacks, and has limited visibility into the host machine (Hindy et al., 2020).

In the following, we will explain the different detection method-based IDS.

3.3 Detection method-based IDS

IDS can be classified into three categories based on their ability to detect misuse, abnormal, or hybrid behavior (Riyaz and Ganapathy, 2020). Detection methods are described below, along with their benefits and weaknesses.

1. Misuse detection
   The signature-based or misuse detection approach is based on known signatures stored in the system as patterns or rules. Each received packet is compared to the signatures that are provided (Lansky et al., 2021). When there is a match, the plan sends an alert. Misuse detection effectively detects frequent cyberattacks but fails to detect new ones. Furthermore, if an error is made in the definition of signatures, the False Alarm Rate is increased. State-based strategies, rule-based methodologies, pattern matching, and data analysis methods implement the misuse detection techniques (Hindy et al., 2020).
2. Anomaly detection
   The anomaly detection approach is based on creating a profile to distinguish between normal and attack behavior. Each incoming packet is examined using several extracted or generated features to determine whether it is normal or malicious (Kunhare and Tiwari, 2018). When an attack activity is detected, an alarm is issued. In contrast to the misuse detection approach, the anomaly detection method can effectively detect a new attack, but at the expense of a high FAR. In the literature, many methods, such as rule-based models, biological models, models based on signal processing techniques, statistical models, and learning models, are used to implement an anomaly detection strategy (Hindy et al., 2020).
3. Hybrid detection
   The hybrid IDS, which combines anomaly and misuse detection approaches, is more effective than either. As previously stated, the anomaly and abuse techniques have advantages and disadvantages (Einy et al., 2021). The disadvantages of the two strategies can be mitigated by combining them. IDS's capacity to detect most network threats has improved (Maseno et al., 2022).

3.4 Significant cyberattacks detected by intrusion detection systems

Intrusion Detection Systems play an essential role in protecting networks and systems from a wide range of cyberthreats. As a

result, it is important to understand how IDS may efficiently detect different types of cyberattacks. Organizations may develop an exhaustive defense plan tailored to their specific security requirements by categorizing attacks into major classes and assigning them to the appropriate IDS type. Table 2 summarizes several classes of cyberattacks, providing a thorough overview of each category as well as particular examples. It also specifies which type of Intrusion Detection System, Host-based (HIDS) or Network-based (NIDS), is best suited to detecting each example inside the appropriate class of attack.

Choosing between HIDS and NIDS depends on the type of attack and its position inside the network. NIDS are ideal for monitoring network-wide traffic patterns for detecting attacks on multiple hosts or network services. On the other hand, HIDS are effective at monitoring particular hosts or systems for signs of unauthorized access or malicious activity.

As illustrated in Table 2, we emphasize the broad nature of cyberattacks and the importance of deploying both NIDS and HIDS to effectively detect and mitigate a wide range of threats to network infrastructure and individual systems.

We propose in Figure 2, an alternative classification for the major attacks detected by IDSs.

- Disruptive attacks. These attacks aim to disrupt or impair the normal functioning of network services or systems. Examples include DoS attacks, DDoS attacks, ICMP floods, Heartbleed, etc.
- Exploratory attacks. This class of attack intended to probe and gather information about network infrastructure and vulnerabilities. Examples include port scanning, OS fingerprinting, network reconnaissance, etc.
- Privilege escalation attacks. This category attempts to elevate user privileges or gain unauthorized access to privileged accounts. Examples include User-to-Root (U2R) attacks, exploiting Sudo vulnerabilities, buffer overflows, Trojan Horse, Spyware, Ransomware, MITM, etc.
- Unauthorized access attacks. This class involves unauthorized access attempts or the exploitation of system vulnerabilities to gain entry. Examples include remote-to-local (R2L) attacks, brute-force attacks, exploiting vulnerable services, Web attacks, infiltration, Botnet, spoofing, Mirai, etc.

This alternative classification depicted in Figure 2 underscores the importance of deploying appropriate IDSs to detect and mitigate threats in all categories, strengthening defenses against evolving cyber threats.

4 Artificial intelligence methods for IDS

The utilization of ML and DL techniques often entails the execution of three primary steps, as seen in Figure 3: (i) Data preprocessing step, (ii) Training step, and (iii) Testing step. Before utilizing the technique, the dataset undergoes preprocessing to convert it into a usable format. During this phase, the process usually includes encoding and normalization. During the step 1, it is necessary to clean the dataset by deleting entries that have missing data and duplicate records. So, the first step includes transformations to numeric, data visualization and analysis, scalling and normalization. The preprocessed data is subsequently partitioned into two random subsets: the training and testing datasets. Usually, the training dataset consists of approximately 80% of the original dataset, while the remaining 20% is used for testing purposes. The training dataset is used to train the ML or DL algorithm : step 2. After the model has been trained, it is sent to testing using a separate dataset and assessed by its predictions. For IDS models, the network traffic instance will be classified as benign or an attack : step 3.

This section overviews the ML and DL methodologies frequently employed in developing an efficient IDS.

4.1 Machine learning algorithms

ML is an AI discipline that allows machines to learn from enormous datasets by automatically building mathematical models (Xin et al., 2018). This subsection describes the most often used ML approaches for IDS.

1. K-Nearest Neighbors (KNN)
   KNN is a nonparametric classification approach known as instance-based learning. A lazy learner favors classification over training (Singhal et al., 2021). The KNN algorithm begins by computing distances among points in an n-dimensional space. Second, it finds the k locations closest to the unlabeled moment (Belgrana et al., 2021). Finally, by majority vote, it assigns the unlabeled point to the class of its KNN. The k value impacts classification accuracy (Kunhare and Tiwari, 2018).
2. Support vector machine (SVM)
   SVM is a binary data-supervised classifier. It can, however, be applied to unsupervised machine learning (Binbusayyis and Vaiyapuri, 2021). The main aim of SVM is to determine the optimal hyperplane that effectively separates a collection of training vectors within a high-dimensional space into two distinct classes (Mohammadi et al., 2021). An SVM raises the dimensionality of the input vector to make its components separate to reach the high dimensional space. Maximizing the distance between it and the support samples is essential to find the best hyperplane rather than the complete set of outlier-resistant training vectors (Alsarhan et al., 2021). An SVM is applied in intrusion detection, producing good results regarding FAR compared to other approaches (Wisanwanichthan and Thammawichai, 2021). This article (Zou et al., 2023) proposes a network intrusion detection approach called HC-DTTWSVM, based on DT twin SVM and hierarchical clustering. HC-DTTWSVM is designed to effectively detect various forms of network intrusion. The hierarchical clustering algorithm is initially utilized to create the DT for network traffic data. The bottom-up merging strategy is employed to optimize the separation of the higher nodes in the DT, minimizing error accumulation throughout the construction process. Subsequently, twin SVMs are integrated into the created DT to execute the network intrusion detection model. This model is capable of accurately identifying the network

TABLE 2 Major cyberattacks detected by IDS.

Class of attack (Saranya et al., 2020)	Description	Examples	IDS type	IDS type justification
Denial-of-Service (DoS)	Attacks designed to prevent or restrict the use of a network or computer system's services.	1. Distributed Denial of Service (DDoS) occurs when attackers flood a target system with a large volume of traffic, overwhelming its resources.	NIDS	NIDS is chosen because DoS attacks often involve flooding network resources with traffic.
		2. ICMP Flood: Attackers send a huge number of ICMP echo request (ping) packets to a target system, consuming its network bandwidth and resources.	NIDS	NIDS is appropriate for detecting ICMP Flood attacks, which include flooding a target system with network packets, affecting network bandwidth.
Probing Attacks	Attempts to collect information about a network or computer system, usually in order to detect possible vulnerabilities for exploitation.	1. Port Scanning: Attackers use tools such as Nmap to scan a target system for open ports and services.	NIDS	NIDS is often used for detecting probing attacks since they entail network scanning.
		2. OS Fingerprinting: Hackers analyze the responses of a target system to identify its operating system and possible vulnerabilities.	NIDS	NIDS is useful for detecting OS fingerprinting since it is strategically placed within the network, monitoring network traffic and analyzing its patterns to identify target system characteristics.
User-to-Root (U2R)	A non-privileged user attempts to get root or admin-level access to a system where they originally only had user access.	1. Buffer Overflow: An attacker exploits a buffer overflow vulnerability in a system service to run arbitrary code with elevated privileges.	HIDS	HIDS is elected because U2R attacks target specific hosts or systems, aiming to exploit vulnerabilities locally.
		2. Exploiting Sudo Vulnerabilities: An attacker exploits flaws in the sudo configuration to elevate their privileges to root.	HIDS	HIDS is beneficial for monitoring system logs, file integrity, and user activity on particular hosts for signs of unauthorized access or privilege escalation.
Remote-to-Local (R2L)	The attacker sends packets to the victim's station in an attempt to gain unauthorized access or escalate privileges.	1. Brute Force Attacks: Attackers attempt to guess login credentials by repeatedly trying different username and password combinations.	NIDS	NIDS is ideal for detecting R2L attacks because it involves unauthorized access attempts from external sources that target network services.
		2. Exploiting Weak Attackers exploit vulnerabilities in services like FTP, SSH or RDP for gaining unauthorized access to a target system.	NIDS	NIDS is effective in detecting the exploitation of vulnerable services throughout the network.

intrusion type in a hierarchical approach. The performance of the HC-DTTWSVM approach is assessed using the NSL-KDD and UNSW-NB15 intrusion detection benchmark datasets. The experimental findings demonstrate that HC-DTTWSVM is capable of efficiently detecting various types of network intrusion and achieve similar detection performance to recently suggested approaches for network intrusion detection.

3. Artificial neural networks (ANN)

ANN is a parallel processing model inspired by the brain's neural networks. An ANN's processing unit comprises multiple nodes or neurons connected by a network of synapses, each with a weight and a learning process (supervised or unsupervised). An ANN comprises many unique layers (Sumaiya Thaseen et al., 2021; Hassija et al., 2024). The input layer receives data from the outside world. The hidden layer consists of nodes whose input and output signals remain within the network, and the output layer processes the data and sends it to the outside world (Kavitha and Manikandan, 2022; Javed et al., 2023). Different ANNs, such as Multilayer Perceptron (MLP) and Self-Organizing Map (SOM) can be used in intrusion detection, depending on how many hidden layers they have and the network design (Choraś and Pawlicki, 2021).

This research (Das et al., 2021) presents a comprehensive security solution for network intrusion detection utilizing a machine learning approach. The authors utilize an ensemble-supervised machine learning framework and the ensemble feature selection algorithms: NN, LR, DT, NB, and SVM Ali Hussein Ali (2024b). In addition, they offer a comparative examination of multiple machine-learning models and strategies for selecting features. The objective of this research is to develop a universal detection mechanism that attains superior precision while minimizing the occurrence of false positive rates (FPR). The experiment utilized the NSL-KDD, UNSW-NB15, and CICIDS2017 datasets. The results indicate that the detection model can accurately identify 99.3% of intrusions while maintaining a low false alarm rate of 0.5%. This

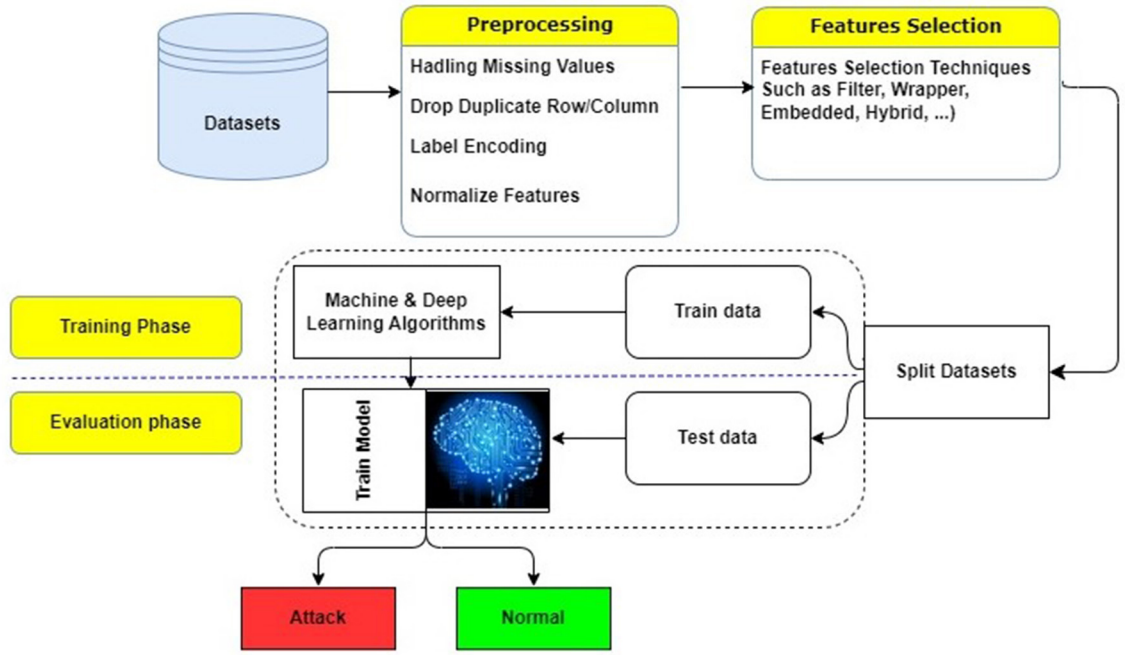

FIGURE 2
Significant cyberattacks detected by IDS.

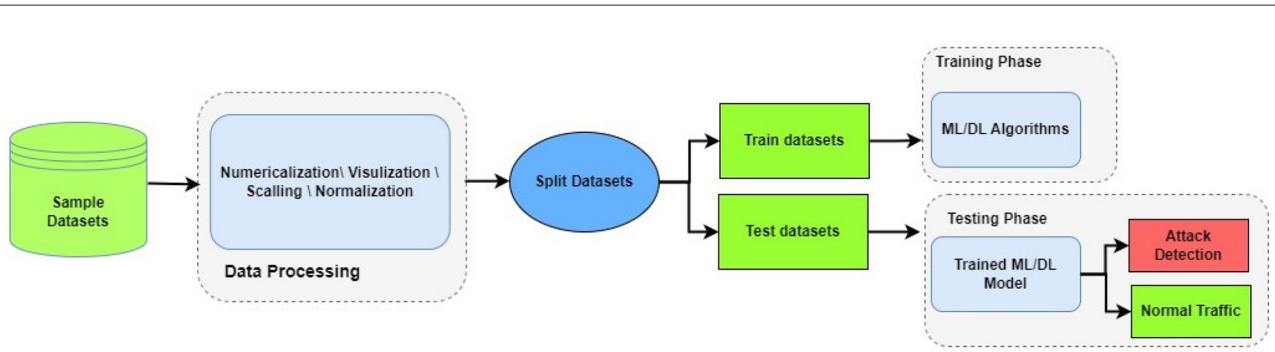

FIGURE 3
Generalized ML/DL-based intrusion detection system methodology.

demonstrates superior performance metrics in comparison to current solutions.

4. K-Means clustering

K-means clustering has become one of the most commonly used unsupervised learning methods due to its ease of use and rapid convergence. It is a method for categorizing a dataset into k separate and non-overlapping clusters (Liu et al., 2021). Before beginning the algorithm, the cluster number k must be determined. The K-Means algorithm starts by randomly selecting the kth object and assigning it to a cluster mean. The remaining objects are then transferred to the kth similar collection based on their distance from the cluster mean. This procedure is performed until the cluster assignments no longer change. The K-means algorithm results generally depend on the initial cluster assignment of the algorithm's first phase (Maseer et al., 2021). As a result, the method must be run numerous times to find the solution with the smallest objective. The author (Chandra et al., 2019) proposes a hybrid model that uses Filter-based Attribute Selection to reduce the dimensionality of the dataset's features. The KDDCUP99 dataset was used for training and testing. This model is evaluated using a variety of performance criteria. The proposed model significantly improves detection accuracy.

The authors of this study (Alenezi and Aljuhani, 2023) suggest a smart intrusion detection strategy that employs principal components analysis (PCA) as a method for feature engineering. This technique aims to identify the most important characteristics, decrease data complexity, and enhance the accuracy of intrusion detection. During the classification phase, the authors utilize clustering methods like K-means to ascertain whether a specific flow of IIoT communication is normal or under attack for binary classification. To assess the suggested

model's efficiency and resilience, it was tested using a novel dataset known as X-IIoTID. The detection method attained a superior accuracy rate of 99.79% and a decreased error rate of 0.21% in the performance results, outperforming current techniques.

5. Tree-based machine learning techniques

Decision Trees (DT) is a classifier that uses a set of known cases to predict the class of an unknown example by applying a series of decisions that can be easily translated into classification rules (Ogundokun et al., 2021). DT is classified into two types based on the task to solve: regression and classification problems. A regression tree is utilized for quantitative classification (numerical class labels), whereas a classification tree is used for qualitative classification. A DT is a flowchart with a node hierarchy (Guezzaz et al., 2021). Each branch, beginning with a root node, indicates the result of a test performed on a non-leaf node that represents an attribute, while a leaf node represents a class label. The attribute value of the unknown classified instance is checked using a DT to trace the path from the root node to the leaf node, reflecting the class prediction for that instance (Al-Omari et al., 2021). A DT is built by increasing the information obtained at each attribute split, leading to a natural feature ranking or selection. In general, DTs offer higher accuracy and simpler implementation than more sophisticated algorithms such as SVMs, and it does not require parameter setup or domain knowledge. The ease of extracting rules from DTs is proportional to the size of the tree (Bhosale et al., 2020). DT-based intrusion detection methods are now in use. Iterative Dichotomiser 3 (ID3), RF and Classification and Regression Trees (CART) are the three most well-known algorithms for implementing DTs. Each ID3, C4.5, and CART uses a greedy, top-down approach in constructing the tree (Riyaz and Ganapathy, 2020). Another known advanced tree-based ML technique, eXtreme Gradient-Boosting XGBoost is selected to enhance attack detection (Alzahrani and Alenazi, 2021). The suggested approach is trained and tested on the NSL-KDD dataset. Compared with basic tree-based ML systems, the dataset is subjected to several sophisticated preprocessing approaches to extract the best form of the data, yielding exceptional results. A multiclass classification challenge identifies attacks and classifies their types with 95.95% accuracy, utilizing only five of NSL-KDD's 41 features. This research enhances NIDS's accuracy and monitoring.

6. Naive Bayes (NB)

The NB Network is a classifier based on the Bayes theorem. It represents probabilistic correlations between relevant variables to simulate an uncertain domain (Kurniawan et al., 2021). NB is represented by a directed cyclic graph, with each node representing a variable, its conditional probability table, and each link encoding how one node affects the others. Because it is a classifier, NB can be used for intrusion detection. Although the usefulness of NB has only been proved in one situation, its outcomes are equivalent to threshold-based systems while needing less computer work (Wester, 2021). The researchers of this paper (Gu and Lu, 2021) propose an IDS based on SVM and NB feature embedding. Fisr, the NB feature transformation is implemented on the original features to generate novel data with high quality; then, an SVM classifier is trained using the new data to create the intrusion detection model. Experiments on multiple intrusion detection datasets reveal the proposed detection method's good and robust performance, with accuracy rates of 93.75% on the UN-SWNB15 dataset and 98.92% on the CICIDS2017, 99.35% on the NSL-KDD dataset, and 98.58% on the Kyoto 2006+ dataset.

4.2 Deep learning algorithms

This section illustrates the DL techniques used by the reviewed studies to deliver DL-based IDS solutions. DL is a subclass of ML that uses deep neural network features provided by several hidden layers. These approaches are characterized by their complex architecture and intrinsic ability to understand the main aspects of a dataset and give an output with minimum human assistance.

1. Recurrent neural networks (RNN)

RNNs are feed-forward neural networks that can be used to represent data consecutively. RNNs are input, concealment, output units, and the model's "memory components." Each RNN unit decides depending on the input and output of earlier inputs (Al-Emadi et al., 2020). RNNs have many more applications than those described above. Within an IDS, supervised classification and feature extraction can be performed using an RNN. If the sequences are excessively long, RNNs have short-term memory problems. Several RNN variants, including the LSTM and GRU types of RNNs, have been created to overcome these concerns (Tang et al., 2019; Mittal et al., 2021). Authors Naseer et al. (2018) conducted a comparative analysis of IDS on a GPU-based testbed using multiple DL and ML approaches. Experiments utilizing LSTM and Deep CNN outperformed those using other models on the NSL-KDD benchmarking dataset. An RNN-based IDS with GRU as the primary memory, a multilayer perceptron, and a softmax classifier has been published in Xu et al. (2018). Tests were done on the KDD Cup'99 and NSL-KDD datasets. The experimental results indicated that the detection performance outperformed other approaches. There is a serious issue with the system's inability to recognize less prevalent forms of attack, such as U2R and R2L. Authors Kasongo (2023) conducted a comparison study on IDS on the NSL-KDD and UNSW-NB15 bench-marking datasets, with XGboost-LSTM achieving higher accuracy than alternative models.

The authors in this study (Bakhsh et al., 2023) propose a DL-based IDS, employing Feed Forward Neural Networks (FFNN), Long Short-Term Memory (LSTM), and Random Neural Networks (RandNN) as defense mechanisms against cyberattacks in IoT networks. The suggested technique performs better than the present state-of-the-art DL-IDS utilizing the CIC-IoT22 dataset. The FFNN model achieves an accuracy of 99.93%, the LSTM model achieves an accuracy of 99.85%, and the RandNN model achieves an accuracy of 96.42% in detecting incursion.

2. AutoEncoder (AE)

AE is a popular DL method that uses unsupervised neural networks. The best features are learned so that the output closely

resembles the input. It includes similar information and output layers. However, the size of the hidden levels is frequently less than those of the input layer (Khan and Kim, 2020; Saheed et al., 2023). AE is symmetric and operates with an encoder-decoder arrangement. AE variants include Stacking AE, Sparse AE, and Variational AE (Rahman et al., 2021; Hameed et al., 2024). Authors Al-Qatf et al. (2018) suggested using a comparable concept of self-learning based on sparse AE and SVM. They validated their performance by running tests with the suggested model with the NSL-KDD dataset. The overall performance improved when the results were compared to different DL and ML methods. In Binbusayyis and Vaiyapuri (2021), the autoencoder (1D CAE) and a one-class support vector machine (OCSVM) are suggested. To test the model, the authors use the NSL-KDD and UNSWNB15 datasets.

Authors Yang et al. (2020) proposed a DNN-Supervised Adversarial Variational (DNN-SAVER) system based on an AE with regularization. It was tested using the datasets UNSW-NB15 and NSL-KDD. According to experimental results, the model effectively recognizes occasional and previously unknown risks. A multistage model with a 1D convolution layer and two fully stacking linked layers was reported (Andresini et al., 2020). To recreate the data, two AEs were trained independently utilizing benign and attack flows. New models from the recovered dataset are supplied into the network as input for creating the 1D-CNN. Finally, a softmax classifier classifies the dataset using the convolution layer results. The proposed technique outperforms other DL models on the KDD Cup'99, CICIDS2017, and UNSW-NB15 datasets.

This work (Catillo and Villano, 2023) introduces CPS-GUARD, an innovative intrusion detection method that utilizes a single semi-supervised autoencoder and a strategy for determining the threshold that separates regular activities from attacks. The method is designed to be sensitive to outliers, using outlier identification to address intrinsic flaws in the training data. CPS-GUARD undergoes evaluation by direct experimentation, utilizing both regular and intrusive data points from individual sensing devices, an HTTP server, and four comprehensive systems, which include Cyber-Physical Systems. The tests encompass a diverse array of attacks present in six cutting-edge datasets. The intrusion detection findings of CPS-GUARD exhibit recall values ranging from 0.949 to 1.000, precision values ranging from 0.961 to 0.999, and false positive rates ranging from 0.006 to 0.027, depending on the particular system under evaluation. The examination also encompasses a comparative analysis of alternative methodologies for selecting thresholds and identifying outliers.

The researchers Hnamte et al. (2023) propose a novel method that combines AE and LSTM algorithms and trained and tested the model using two datasets: CICIDS2017 and CSE-CICIDS2018. The AE encrypts the original data, creating a bottleneck, while the decoding network restores all of the data. The proposed model's key problems are the merging of two types of architectures and training under smoothing limitations. When trained for up to 30 epochs, the suggested hybrid model demonstrated an impressive multiclass detection accuracy of 99.99% on the CICIDS2017 dataset, surpassing the 99.10% achieved on the CSE-CICIDS2018 dataset. The experimental results surpassed the accuracy performance measures of other state-of-the-art intrusion detection methods.

3. Deep neural network (DNN)

   A DNN is a fundamental DL structure that allows multilayer models to be trained. Authors in RM et al. (2020) described the system as having an input layer, an output layer, and several other components. The model's abstraction level increases as the number of hidden layers increases, boosting its effectiveness. In Jia et al. (2019), the KDD cup'99 and NSL-KDD datasets were subjected to categorization using a DNN-based IDS network, including four hidden layers. The activation function employed by authors for the buried layer was the Rectified Linear Unit (ReLU). In Kavitha and Manikandan (2022), the authors apply the bottleneck layer method to the CICIDS-2017 dataset to show how well it can identify cyberattack features. According to the findings, the bottleneck model architecture, which combines ANN and DNN models, is superior to conventional ANN, DNN, and SVM variants. Multiple datasets, such as KDDCup 99, NSL-KDD, Kyoto, UNSW-NB15, and CICIDS 2017, were used to measure the performance of the proposed IDS model. The experimental findings showed that the suggested model performed better than other ML methods.

4. Deep belief network (DBN)

   DBN is a deep learning model that utilizes Restricted Boltzmann Machines (RBMs) followed by a softmax classification layer. In an RBM, two layers of data flow in both directions (Tan et al., 2019). All nodes in the preceding and subsequent layers of the layer are linked, while nodes in the current layer are not. Unsupervised layer-wise learning is used to pre-train DBN before using supervised fine-tuning to discover useful features. The IDS system uses DBN to extract and classify characteristics (Süzen, 2021).

   The paper of He et al. (2023) examines the characteristics of adversarial challenges in Network Intrusion Detection Systems (NIDS). Authors focused on the offensive approach, which involves developing methods to create adversarial examples to bypass various machine-learning models. They specifically investigated the utilization of evolutionary computation techniques such as Particle Swarm Optimization (PSO), Genetic Algorithms (GA), DBN, and deep learning methods like Generative Adversarial Networks (GANs) to generate these examples. To evaluate their ability, the researchers utilized these algorithms on two datasets that are accessible to the public: NSL-KDD and UNSW-NB15. The findings indicated that their methodologies elevated misclassification rates across eleven ML models, including a voting classifier.

5. Convolutional neural network (CNN)

   Regarding data structures, CNN is better suited for data stored in arrays. A sequence of convolutional and pooling layers, followed by a fully connected layer and a softmax classifier, comprise the framework for feature extraction (Riyaz and Ganapathy, 2020). Regarding computer vision, CNN has a long history of success (Fki et al., 2023). IDS uses them for feature extraction and classification, so they are supervised (Khan et al., 2019; Azizjon et al., 2020). Proposals for an IDS model using

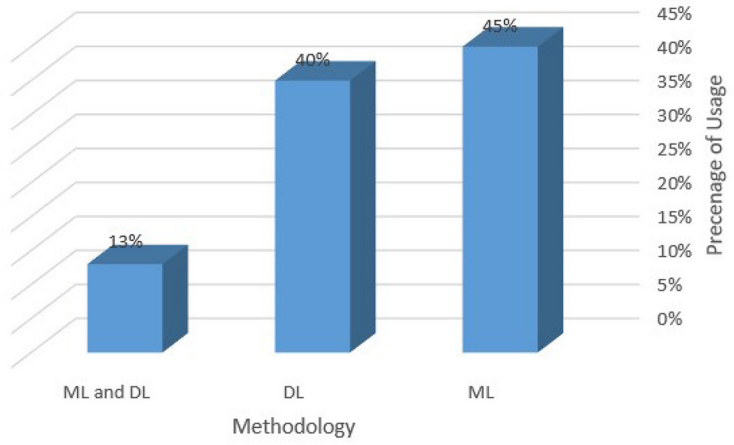

FIGURE 4
Methodology distribution.

CNN and gcForest were made, and a new P-Zigzag approach for generating two-dimensional grayscale images from raw data was also available (Zhang et al., 2019). An advanced CNN model (GoogLeNetNP) was applied in a coarse grit layer. The anomalous classes are further sub-classed using gcForest in a fine-grained layer. It was chosen to combine the UNSW-NB15 and CIC-IDS2017 datasets to create a new dataset. According to the trials, their methodology significantly decreases FAR when compared to single methods. For an effective IDS system, Jiang et al. (2020) proposed a deep hierarchical CNN-BiLSTM system. This approach uses both CNN and BiLSTM to handle the class imbalance issue; the SMOTE is employed to increase minority samples, aiding the model learning process (Ali Hussein Ali, 2024a). To extract geographical and temporal properties, CNN and belts were used. The datasets NSL-KDD and UNSW-NB15 were used in the studies. As a result, the proposed approach has higher accuracy. Because of the structure's complexity, the training duration is longer. The few-shot learning (FSL) model was described as an IDS model (Yu and Bian, 2020). Small amounts of uniformly dispersed labeled data from the dataset are employed for training. Two datasets are used, the NSL-KDD and UNSW-NB15, to illustrate the model's usefulness. The article (Wang et al., 2023) employs the CSE-CIC-IDS2018 dataset and evaluates its performance using standard evaluation metrics. Six models, namely DNN, CNN, RNN, LSTM, CNN + RNN, and CNN + LSTM, were developed to ascertain the presence of a malicious attack in network traffic. The proposed model greatly enhances the performance of detection. In addition, the processing time for combinations of CNN with RNN and CNN with LSTM is greater than that of individual DNNs, RNNs, and CNNs. Thus, when implemented in an IDS device, it can be inferred that DNNs, RNNS, and CNNs are superior to utilizing combinations like CNN+RNN and CNN+LSTM.

This study (Saba et al., 2022) introduces a CNN technique for anomaly-based intrusion detection systems (IDS) that leverages the capabilities of the IoT to effectively analyze all network traffic in the IoT environment. The suggested approach can identify all potential intrusions and atypical traffic patterns. The model was trained and evaluated using the NID Dataset and BoT-IoT datasets, attaining accuracy rates of 99.51% and 92.85%, respectively.

This research (Madwanna et al., 2023) presents two deep learning-based intrusion detection systems (IDSs). The first IDS is a fusion of LuNet and Bidirectional Long Short-Term Memory (Bi-LSTM), while the second IDS combines Temporal Convolutional Network (TCN), Convolutional Neural Network (CNN), and Bi-LSTM. In order to maintain the IDS (Intrusion Detection System) up-to-date and precise, it is necessary to provide it with a sufficient quantity of samples. The first model has undergone training and evaluation using two established benchmark datasets, namely NSL-KDD and UNSW-NB15. The second model has undergone training and testing using the NSL-KDD dataset. In order to address the issue of limited sample size, the models have employed a method known as Synthetic Minority Oversampling Technique (SMOTE). These models yielded superior experimental results compared to conventional machine learning-based methods and numerous deep learning approaches. Their classification accuracy and detection rate are superior. The first model achieved a classification accuracy of 82.19% for UNSW-NB15 and 98.87% for NSL-KDD. The second model achieved a classification accuracy of 98.8% for NSL-KDD.

Referring to Figure 4, observations indicate that 45% of the suggested methods exclusively utilize machine learning (ML) approaches and 40% of the solutions apply DL methods. In contrast, only 13% of the recommended solutions are based on a hybrid strategy that mixes ML and DL algorithms.

Table 3 displays a compilation of recent ML and DL algorithms that researchers have developed to detect network attackers.

Various metrics can be used to evaluate ML and DL algorithms for IDS.

5 Evaluation metrics and performance indicators

This section discusses commonly used metrics and performance indicators. The Confusion Matrix (CM) is a

TABLE 3 An overview of recent research on network intrusions with ML/DL techniques.

Study	Algorithms		Methodology
	ML	DL	
Zou et al. (2023)	X		SVM and DT
Alenezi and Aljuhani (2023)	X		K-means
Bakhsh et al. (2023)	X	X	FFNN, LSTM, and RandNN
Hnamte et al. (2023)		X	AE and LSTM
He et al. (2023)		X	PSO, GA and GANs
Catillo and Villano (2023)		X	Single semi- supervised autoencoder
Saba et al. (2022)		X	CNN
(Das et al., 2021)	X		NN, SVM, DT, NB and LR
Mbona and Eloff (2022)	X		OCSVM
Choraś and Pawlicki (2021)		X	DNN
Maseer et al. (2021)	X	X	ANN, DT, k-NN, NB, RF, SVM, CNN, EM and k-means
Mohammadi et al. (2021)	X		SVM
Wisanwanichthan and Thammawichai (2021)	X	X	1D CAE and OCSVM
Alsarhan et al. (2021)	X		GA, PSO, ACO, and SVM
Gu and Lu (2021)	X		SVM and NB
(Wang et al., 2023)		X	DNN, CNN, RNN and CNN+RNN

two-dimensional matrix that defines the actual and expected categories (Deng et al., 2016; Zhu and Liu, 2024).

- True Positive (TP): The classifier successfully recognizes data objects as Attacks.
- False Negative (FN): Incorrectly identified as Normal.
- False Positives (FP): Instances in the data that were wrongly identified as Attacks.
- True Negative (TN): The instances are correctly classified as Normal.

The following are the various metrics used in the most recent evaluation research:

- Accuracy: the ratio of accurately identified cases to total cases. Detection accuracy only matters if the dataset is uniformly distributed.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- Precision: refers to the percentage of accurately predicted attacks relative to the total number of samples labeled as attacks.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

- A recall (Sensitivity): is the fraction of samples correctly labeled as attacks relative to the total number of attack samples.

$$Recall = TPR = Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

- False alarm rate (Specificity): determines the proportion of wrongly predicted positive labels among all actual negative labeled attacks.

$$FAR = FPR = Specificity = \frac{FP}{FP + TN} \tag{4}$$

- True negative rate (TNR): is the fraction of samples correctly labeled as normal as a percentage of all normal models.

$$TNR = \frac{TN}{TN + FP} \tag{5}$$

- F-Measure: is the mathematical middle ground between Precision and Recall. It is a statistical method for assessing a system's reliability by looking at its performance in terms of precision and recall.

$$F - Measure = 2 * (\frac{Precision * Recall}{Precision + Recall}) \tag{6}$$

- False discovery rate: this metric measures the proportion of incorrectly predicted positive labels among all positive predictions. It is also known as false alarm rate or type I error rate, calculated as:

$$FDR = \frac{FP}{FP + TP} \tag{7}$$

- Matthews correlation coefficient: this metric measures the correlation between the predicted and actual values of a classification model. It takes into account the true positives, false positives, true negatives, and false negatives generated by the model.

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN}} \tag{8}$$

Besides, statistics-based Intrusion Detection System constructs a model that can represent the distribution of normal behavior profiles. It, then, identifies low-probability events and flags them as potential intrusions. Statistical AIDS essentially considers the statistical metrics like the median, mean, mode, and standard deviation of packets. Instead of inspecting entire data traffic, each packet is individually monitored, serving as a fingerprint of the flow. Statistical IDSs effectively identify deviations in present behavior from the established normal behavior (Li et al., 2019).

In addition to the common used metrics, some performance indicators can be useful for further comparative analysis. In fact, ANNs require careful selection and fitting of activation functions to

TABLE 4 Overview of the datasets used in ML and DL-based IDS.

Dataset	Attack types	Total Features
KDD-Cup 99	Normal, DoS, Probe, U2R, R2L	41
NSL-KDD	Normal, DoS, Probe, U2R, R2L	41
Kyoto 2006+	Normal (no attack), attack (known attack), and unknown attack.	24
UNSW-NB15	Fuzzers Analysis, Backdoors, DoS, Exploits, Generic	49
CICIDS2017	Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet and DDoS	80
CSE-CICIDS2018	HeartBleed, DoS, Botnet, DDoS, Brute Force, Infiltration, Web.	81
CIC IoT 2023	DoS, DDoS, Brute Force, Spoofing, Recon, Web, Mirai.	46
CIC-MalMem- 2022	Trojan Horse, Spyware, Ransomware	56
IOT Intrusion 2020	DoS, Mirai, MITM, Scan	79

effectively model complex relationships in data. However, they can be prone to convergence issues, where the training process struggles to reach an optimal solution (Dini et al., 2023). Convergence in neural networks refers to the stage during training where further adjustments to the model's parameters result in diminishing improvements in performance. At this point, the changes in the learning rate become minimal, and the errors produced by the model on the training data approach a minimum. Another way to identify convergence in a deep learning model is when the loss, which quantifies the disparity between predicted and actual values, reaches its lowest achievable value. These variations aim to improve the convergence speed which can be an added metric and overcome issues related to large datasets.

The complexity of the ANN algorithm is related to the time computational time. It's noteworthy that binary classification generally outperformed multi-class classification, achieving precise results with low false negatives and false positives. Conversely, multi-class classification is more computationally intensive and intricate, resulting in less effective outcomes, as evidenced in this study. For example, generally, DT demonstrated a quick computation time, likely due to their lower complexity compared to other methods. However, all the methods exhibited relatively high computation times, highlighting the necessity for efficient algorithms in intrusion detection systems (Dini et al., 2023).

6 Benchmark datasets

This section outlines the datasets used by the researchers to test their methodologies. Table 4 includes a detailed overview of the dataset and the corresponding attacks.

- KDD Cup'99: It receives much attention and is a popular IDS dataset. There are around 5 million training recordings and 2 million test records available. Based on 41 criteria, each entry is classified as normal or an attack (Al Tobi and Duncan, 2018).
- NSL-KDD: Kyoto University gathered network traffic records for this dataset using honeypots, darknet sensors, email servers, web crawlers, and other network security protocols. Each record has 24 statistical attributes, 14 of which are extracted from the KDD Cup'99 dataset and ten new ones (Salo et al., 2019).
- Kyoto 2006+: Kyoto University used honeypots, darknet sensors, email servers, web crawlers, and other network security protocols to gather network traffic records for this dataset. Each record has 24 statistical attributes, 14 drawn from the KDD Cup'99 dataset, and 10 are additional features (Gu and Lu, 2021).
- UNSW-NB15: The Center for Cyber Security in Australia developed this dataset. Bro-IDS, Argus, and other novel approaches were used to recover around two million records with 49 features (Michelena et al., 2024).
- CIC-IDS2017: This dataset was created in 2017 by the Canadian Institute of Cyber Security (CIC) and is now publicly available. This version will find recent iterations of real-world attacks and the typical flow patterns (Kumar and Pathak, 2022).
- CSE-CIC-IDS2018: The Communications Security Establishment (CSE) and the CIC partnered to create this dataset in 2018. Abstract representations of the numerous occurrences in user profiles are made. These profiles are combined into a single dataset by employing a distinct collection of features (Karatas et al., 2020).
- CIC IoT dataset 2023: This dataset contains a real-time benchmark for large-scale attacks in IoT environments. These attacks are classified into seven categories: DDoS, DoS, Recon, Web-based, Brute Force, Spoofing, and Mirai. These attacks are executed by malicious IoT devices targeting other IoT devices. The dataset includes 46 features (Jony and Arnob, 2024).
- CIC-MalMem-2022 : The data was designed to closely replicate a real-life scenario by employing malware that is commonly utilized in the real world. The dataset consists of Spyware, Ransomware, and Trojan Horse malware. It can be used to assess obfuscated malware monitoring systems. The dataset is evenly distributed, with 50% consisting of malignant memory dumps and the other 50% consisting of benign memory dumps (Talukder et al., 2023).
- IOTINTRUSION-2020 : The IoT network intrusion dataset was developed in 2020. This dataset contains eight IoT cyberattacks, including flooding, brute force, spoofing, and scanning, as well as 79 network traffic features that describe benign and malicious network traffic. Network features were extracted using a CIC flowmeter (Ullah and Mahmoud, 2020).

Benchmark datasets play a crucial role in evaluating the effectiveness of the suggested methodology. Figure 5 examines the utilization of the public datasets. It is demonstrated that 41% of the time, NSL-KDD and KDD Cup'99 were utilized for the objectives of testing and validation. Both datasets are considered old-fashioned, although they remain highly favored among academics due to

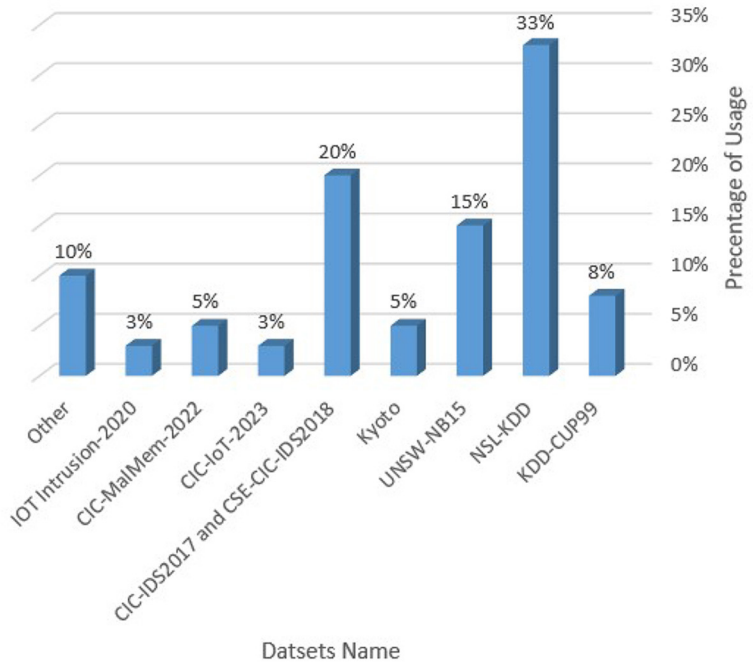

FIGURE 5
Datasets distribution.

the abundance of comprehensive findings documented in the literature. It is clear that IoT Intrusion 2020, CIC-MalMem-2022 and CIC-IoT-2023 datasets represent only 11% of utilization in research papers because they are new comparing to others.

7 Discussion, challenges and future trends

Following the previous sections, we discuss the most significant findings in ML and DL-based IDS. Next, we address the research challenges related to this subject. Then, we highlight novel trends and future directions to effectively detecting intrusions. Finally, we outline the practical and managerial implications and impacts of using ML, DL, and advanced related technologies within IDS.

7.1 Discussion of the findings

The AI-powered IDS's success relies mainly on training with a sufficient dataset. Training an ML model can give satisfactory results even with a small dataset. However, ML is only suitable for large datasets if the data is naturally tagged. Because labeling is time-consuming and expensive, DL techniques are recommended for large datasets. These algorithms can learn and discover interesting patterns from raw datasets.

Computational power and time to process increase the time and resource requirements of the learning process. Training the IDS model improves its ability to identify attacks. Table 5 shows the strengths and weaknesses of approaches used in recent articles regarding complexity, amount of execution time, available dataset, and evaluation metrics used. We also found that the model's performance in some proposed solutions datasets could be stronger in more current datasets. Another key problem that most approaches share is their failure to detect attacks when given a sparse training dataset successfully. Because of the class imbalance issue, the accuracy for these underrepresented attacks should be given additional thought.

Conversely, we discovered that certain approaches are more complex, demanding longer model training durations. DL techniques show a tradeoff between model complexity and a more detailed organizational structure. The more complex the model, the longer it takes to execute and the more resources it requires. As a result, by carefully picking important features for the model's training, this disadvantage can be overcome.

Furthermore, Figure 6 and Table 3 illustrate how researchers utilize ML- or DL-based algorithms while developing an effective IDS solution. The three most commonly utilized algorithms are SVM, RNN, and CNN. Next, approaches such as DT, K-Means and ANN are included in the list and are mostly employed in hybrid designs to support and enhance DL algorithms.

Figure 7 and Table 6 display the performance metrics utilized by the researchers to evaluate the methodology. The two most commonly utilized performance indicators are accuracy and precision. To demonstrate the efficacy of an IDS developed using ML or DL techniques, it is essential to consider Accuracy, Recall, Precision, and F-measure as the primary performance metrics, among others, to showcase its capability in detecting intrusions. MCC, FNR, TPR, and FAR can also, help to evaluate the performance of the IA algorithm in IDS.

TABLE 5 Strengths and weaknesses of the ML and DL-based IDS approaches.

Study	Strengths	Weaknesses
Zou et al. (2023)	The experimental findings demonstrate that HC-DTTWSVM is capable of efficiently detecting various types of network intrusion and achieve similar detection performance to recently suggested approaches for network intrusion detection.	The primary issue with employing an SVM classifier in dynamic IDS systems is that conventional hyperplanes are unable to discern between benign and malicious behavior over an extended period. Additionally, the model is based on old datasets.
Alenezi and Aljuhani (2023)	The proposed model was tested using a novel dataset known as X-IIoTID. The results showed an accuracy rate of 99.79% and decreased error rate of 0.21%, outperforming current techniques.	The detection performance for multi-class classification needs improvement as some attacks have achieved lower results. Additionally, the work needs to be expanded to include more clustering techniques and additional IIoT datasets.
Bakhsh et al. (2023)	The suggested technique performs better than the present state-of-the-art DL-IDS utilizing the CIC-IoT23 dataset. The FFNN model achieves an accuracy of 99.93%, the LSTM model achieves an accuracy of 99.85%, and the RandNN model achieves an accuracy of 96.42% in detecting incursion.	This work requires a wide and diverse dataset to ensure optimal training of models, as well as the need for computational resources capable of effectively managing the complexity of IoT data.
Hnamte et al. (2023)	The suggested hybrid model demonstrated an impressive multiclass detection accuracy of 99.99% on the CICIDS2017 dataset, surpassing the 99.10% achieved on the CSE-CICIDS2018 dataset.	The suggested model has limits in terms of both training time duration and complexity when applied to a large dataset. It still takes longer than usual DNN and CNN models.
He et al. (2023)	Algorithms are tested on NSL-KDD and UNSW-NB15, two public datasets. Their methods considerably raised misclassification rates across eleven ML models, including a voting classifier.	The amount of time spent training is significant. In addition, the proposed system uses old datasets.
Catillo and Villano (2023)	The testing included six modern datasets of various attacks. CPS-GUARD intrusion detection yielded recall values of 0.949 to 1.000, accuracy values of 0.961 to 0.999, and FPR values of 0.006 to 0.027, depending on the system.	The system needs to be expanded in its analysis to include other attacks, in order to gain better insights into the potential limitations of CPS-GUARD. Therefore, this analysis requires more time and calculations.
Saba et al. (2022)	The model was trained and assessed on the NID and BoT-IoT datasets, attaining 99.51% and 92.85% accuracy rates.	More epochs, batch sizes, and parameters are required to improve the proposed model's performance in a real-world IoT environment.
Das et al. (2021)	Comparative analysis of several ML models and feature selection methods based ensemble learning. The experiment used three datasets. Results reveal that the detection model can detect 99.3% of intrusions with a 0.5% FAR.	Reducing amount of data for training purposes compromises the probability of a near-perfect performance.
Mbona and Eloff (2022)	Benford's Law can identify benign and zero-day network traffic. The authors tested this analysis on several new datasets.	The proposed system is based on a single feature selection method. Additionally, the system relies on outdated datasets which cannot detect zero-day attacks.
Choraś and Pawlicki (2021)	It is proposed to merge ANN and DNN to build a new method for detecting network attacks.	It is important to train the system on larger datasets to ensure that the suggested algorithm's selection is unbiased in every circumstance.
Maseer et al. (2021)	Several ML or DL algorithms have been suggested to implement anomaly-based IDS for web attack detection, with benchmarking results showing low false positive and false negative detection rates, as well as short training and prediction times.	A large amount of processing and time takes to get the best hyperparameters.

(Continued)

TABLE 5 (Continued)

Study	Strengths	Weaknesses
Mohammadi et al. (2021)	The proposed model uses the SVM algorithm with other algorithms to improve the diagnosis of the threats and reduce time where the highest accuracy appeared with the GA algorithm.	The fundamental problem with using an SVM classifier in dynamic IDS systems is that the traditional hyper-planes cannot differentiate between benign and malicious behaviuor over the long term.
Wisanwanichthan and Thammawichai (2021)	The suggested method integrates 1D CAE and OCSVM utilizing a joint optimization framework. The outcomes indicate the viability of the proposed method for creating effective IDS.	Training and data preprocessing take a considerable amount of time. Also, it uses only the old dataset NSL-KDD.
Alsarhan et al. (2021)	The suggested ANN model outperforms state-of-the-art accuracy, sensitivity, and specificity approaches.	Local minimum convergence makes ANN unstable. It is resolved by merging CFS and ANN for intrusion identification which takes more processing time.
Gu and Lu (2021)	Experiments on various intrusion detection datasets have shown the proposed detection method's excellent and robust performance, with better accuracy rates on the NSL-KDD dataset.	Lots of time to preprocess the three datasets and train the suggested model.
Wang et al. (2023)	The proposed study uses six models to judge whether network traffic comprised a malevolent attack. Compared to other papers' IDS, this approach significantly improves detection performance.	The proposed study used only one dataset to evaluate the performance of detection.

7.2 Research challenges for ML and DL-based intrusion detection systems

This subparagraph emphasizes the research obstacles in the field of IDS.

1. This research has revealed the lack of a current dataset that accounts for new attacks on modern networks. Most proposed algorithms failed to detect zero-day attacks because these models were not appropriately trained with multiple attack kinds and patterns (Vangipuram et al., 2020). Testing and validating an IDS model using a dataset that includes historical and recent incidents is critical. If the dataset has a definition for the maximum number of attacks, the ML/DL model can better understand patterns and ultimately provide security against a wider range of invasion scenarios. On the other hand, building a dataset requires a large investment of time, energy, and the specialized knowledge of many experts. As a result, one of the challenges of IDS research is systematically developing a contemporary dataset with enough examples of practically all attack types.
2. Unbalanced data reduces the detection rate: Most proposed IDS systems have detection accuracies for specific attack types lower than the model's overall accuracies. It is related to the uneven distribution of data. On average, less common attacks are more difficult to detect than more prevalent ones. There are two approaches to this problem. The first is to assemble a broad and accurate data set. Finding effective strategies to increase the number of minority attack occurrences to build a more representative dataset is one option. SMOTE, RandomOverSampler, adaptive synthetic sampling technique, and other researchers have recently devised strategies for lowering the dataset imbalance ratio to increase performance (Wu et al., 2020; Ali Hussein Ali, 2024a). However, there is still an opportunity for advancement in this subject, necessitating further research.
3. Complex models use many resources: Most IDS strategies reported in the literature are based on complex models that require a large amount of time and computer resources. This may cause the CPU to perform unnecessary work, lowering the IDS's efficacy. Using a powerful computer with advanced capabilities may reduce processing time and effort, but at a large cost. As a result, a viable technique for selecting the most significant attributes while minimizing computational and processing overhead is required. Despite academic efforts to examine alternative optimization algorithms for feature selection, there is undeniably a need for improvement. More research will be necessary to develop a suitable feature selection optimization technique.

7.3 New trends in ML and DL-based intrusion detection systems

We present novel trends in ML and DL applied to intrusion detection systems, focusing on some key elements to broaden their scope. Hence, we propose future directions to effectively detect intrusions in real-world environments.

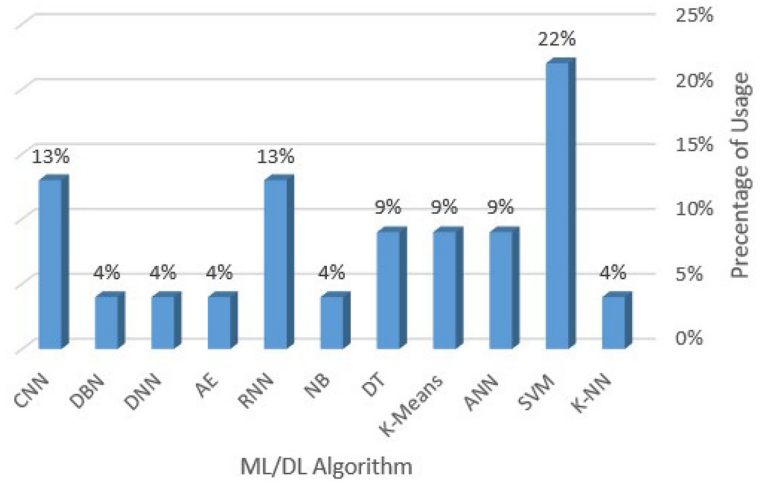

FIGURE 6
Total frequency of usage of ML and DL algorithms.

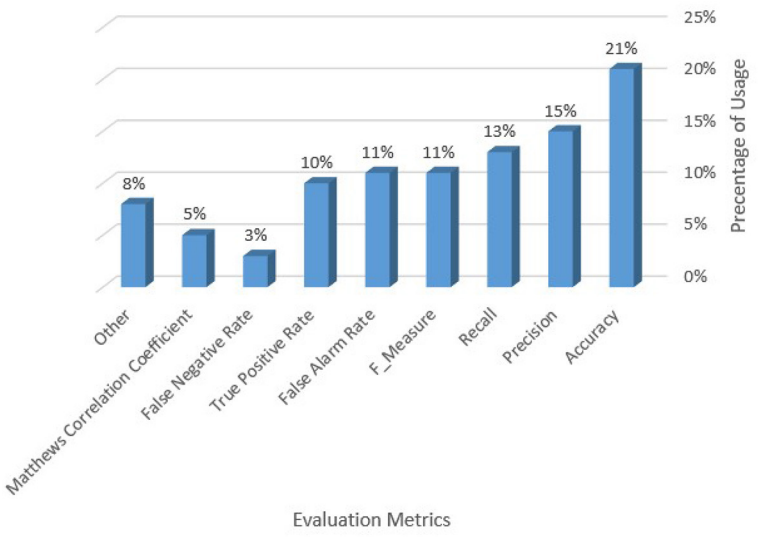

FIGURE 7
Evaluation metrics.

1. The IDS is critical to the security of any network. A recent study has shown that the approach cannot consistently detect zero-day attacks due to a substantial FAR (Wu et al., 2020). A current, thorough, well-balanced dataset can help with this goal. Such a field of inquiry can assist researchers in developing a comprehensive IDS framework capable of protecting networks from any potential attack. The IDS model's capacity to identify zero-day threats and reduce false positives will be enhanced.
2. The solution to complex models: The success of DL-based IDSs is primarily due to the effectiveness of deep feature learning in identifying malicious intrusions (Vangipuram et al., 2020). Processing power, storage space, and time are all required to execute DL algorithm-based models. Because of the intricacy of these systems, implementing IDS in such a dynamic environment is difficult. One option to address these challenges is to use high-performance GPUs to analyze massive datasets quickly and efficiently. Graphics processing units, on the other hand, are relatively expensive. As a result, efficiency and affordability are opposed. One approach for training models at a cheaper cost is to look into GPU platforms or cloud-based services (Salvakkam et al., 2023). This problem can be handled by implementing effective and intelligent feature engineering to reduce the complexity of DL algorithms. A smaller number of features can achieve the same detection accuracy as all available data. As a result, less processing power is required in real-time, and the model's complexity is reduced.

TABLE 6 Datasets and performance evaluation metrics.

Study	Dataset										Evaluation metrics								
	KD	NS	KT	U15	CI	CS	MM	CT	IO	OT	AC	PR	RE	F-M	FAR	TPR	FNR	MCC	OT
Zou et al. (2023)	X	X	X							X	X	X	X		X	X		X	X
Alenezi and Aljuhani (2023)		X						X				X	X	X	X	X	X		
Bakhsh et al. (2023)		X	X	X			X				X				X				X
Hnamte et al. (2023)		X			X						X	X				X			X
He et al. (2023)		X			X						X	X	X						
Catillo and Villano (2023)		X		X					X		X			X	X				
Saba et al. (2022)		X				X					X					X			
Das et al. (2021)		X		X	X						X	X	X	X	X	X	X		
Mbona and Eloff (2022)		X			X	X	X			X	X	X	X					X	X
Choraś and Pawlicki (2021)		X			X					X	X					X			
Maseer et al. (2021)		X			X										X				X
Mohammadi et al. (2021)		X								X		X	X					X	
Wisanwanichthan and Thammawichai (2021)		X		X							X			X					
Alsarhan et al. (2021)	X			X							X	X	X	X	X				
Gu and Lu (2021)	X			X							X			X					
Wang et al. (2023)											X	X	X	X					

KD: KDDCup-99, NS: NSL-KDD, KT: Kyto 2006+, U15: UNSW-NB15, CI: CICIDS2017, CS: CSE-CIC-IDS2018, MM: CIC-MalMem2022, CT: CIC-IoT-2023, IO: IoT Intrusion2020, OT: Other.
AC, Accuracy; PR, Precision; RE, Recall; F-M, F-Measure; FAR, False Alarm Rate; TPR, True Positive Rate; FDR, False Discovery Rate; MCC, Matthews Correlation Coefficient; OT, Other.

10.3389/fcomp.2024.1387354

3. Use of DL algorithms: Recently, it has been suggested that DL-based algorithms be employed in IDS architecture (Yi et al., 2023). The investigation into DL's potential use of IDS is in its early stages. Certain DL algorithms have been researched, with many of them being put to good use in the formulation of acceptable solutions. Some DL approaches, such as Deep Reinforcement Learning (Alavizadeh et al., 2022), require more research before being utilized to propose an adequate IDS solution. Recently, Nguyen et al. investigate DRL approaches developed for cyber security Nguyen and Reddi (2023). They touch on different aspects, including protecting cyber-physical systems by DRL-based security methods, defense strategies against cyberattacks using multiagent DRL-based game theory simulations and autonomous intrusion detection techniques.

   Alternatively, researchers can combine DL feature extraction and ML classification. As a result, the proposed model will become more straightforward.
4. In the context of intrusion detection systems, adversarial attacks may significantly affect the performance of such systems (Martins et al., 2020; Alhajjar et al., 2021; Khalid Albulayhi, 2023).

   An adversarial attack is an attack meant to fool the target ML model regardless of the type of ML model being used. This means that the adversarial attacks try to bypass a system so that the affected system behaves in an unwanted manner (Miller et al., 2020). Essentially, adversarial attacks and the subsequent degradation in the performance of intrusion detection system could lead to enormous risks associated with cybersecurity. This comes in the form of incorrect classifications of network traffic, which results in many genuinely malicious activities going undetected. Researchers have shown that most intrusion detection systems are quite susceptible to adversarial attacks (Frank and Nancy, 2019). However, it has been explained that there exists an inherent resiliency in deep learning systems, that is, proper tuning may prevent against some adversarial attacks which are based on poisoning the training data (Abou Khamis et al., 2020). The most prominent method in which adversaries disrupt intrusion detection systems is through adversarial sampling. This is where the adversary specifically crafts disturbances, ensuring that the intrusion detection system fails to detect malicious instances (Ángel Luis Perales Gómez et al., 2021). In conjunction with this method, the dynamic environment in which intrusion detection systems operate, along with the vast number of possibilities for adversarial attacks, will mean that research will constantly be required to ensure the security of these systems. Indeed, generative artificial intelligence models such as ChatGPT can be used to disrupt the functionality of security tools such as IDS via automated hacking and various attack scenarios (Charfeddine et al., 2024). To counter the threat of these adversarial attacks, some defense strategies have been proposed, including adversarial training, preprocessing techniques, the addition of extra networks and digital watermarking (Szyller et al., 2021; Charfeddine et al., 2022). Adversarial training aims to strengthen ML/DL models by including adversarial samples during the training process. Preprocessing techniques involve carefully planned data transformations that limit the impact of adversarial perturbations. Adding more networks uses external models to identify samples that have never been seen before, improving the system's ability to detect adversarial attacks. Embedding digital watermarks during training allows models owners to identify them in the event of an adversarial attack. Furthermore, additional methods for detecting adversarial samples have been proposed, such as using subnetworks as detectors or using confidence scores to identify out-of-class data. Moreover, defensive techniques based on generative adversarial networks (GANs) have been developed to enhance the robustness of IDSs against certain types of attacks (Alotaibi and Rassam, 2023).
5. There are certain additional challenges with IDSs, particularly in terms of system reliability. Cybersecurity specialists now usually agree on IDS guidelines, so the system's forecasts should be comprehensible. As a result, their increasing sophistication is a significant disadvantage given the high accuracy levels achieved by such systems; they cannot include information about why they make decisions. As a result, some details about the causes that underpin IDS forecasts must be provided, as well as some clarification on the intrusions discovered by cybersecurity professionals. Few studies have described these new trends and developments in IDSs (Younisse et al., 2022; Pande and Khamparia, 2023).

   These research works propose systems based on Shapley additive explanations (SHAPs) to overcome these drawbacks and provide a more accurate interpretation of IDS. SHAP offers a solid theoretical foundation for both shallow and deep-trained models. The authors in Pande and Khamparia (2023) define a system that delivers both local and global interpretations to improve the generalizability of all IDSs. Local descriptions include knowledge that each function value reduces or increases the anticipated likelihood. Global interpretations examine the relationships between the importance of functions and specific types of threats by extracting key attributes from each IDS. These systems lead to a better understanding of IDS forecasts and ultimately aim to instill confidence in IDSs for cyber-users. They enable cybersecurity professionals to better detect cyber attacks.
6. DL models cannot perform well when small training datasets are used, or when there is a discrepancy or inconsistency in data distribution between training and test data. Hence, the quantity and quality of features are important in improving classification because they help the DL model understand their significance and correlation. If only a few features are used, classification quality suffers, resulting in overfitting; if too many are used, generalization suffers, resulting in underfitting. Deep Transfer Learning was introduced to address the issues, which are primarily caused by data scarcity and inconsistency (Kheddar et al., 2023; Latif et al., 2024). It is based on the principle of feeding target model knowledge from a pre-trained source model, so that the target model begins with patterns learned while completing a related task of the source model rather than starting from the beginning. ML and DL techniques include multi-task learning, domain adaptation, multiple and/or cross-modalities, and the use of multiple datasets. They can be viewed as a method of fusing information from multiple sources to improve the overall performance of the model. ML and DL enable more effective information fusion and can produce better

results than training models from scratch. These advantages motivate the development of ML and DL-based IDS models to solve many problems in a wide range of applications and to detect attacks and intrusions that traditional methods may miss.

7. Cybersecurity research is vulnerable to a multitude of problems with infrastructure, making it more difficult to operate in a real-time environment due to a variety of concerns, particularly processing overhead. Furthermore, as technology advances, there is a possibility of attacks on polymorphic systems, with new attacks emerging each time. Traditional IDS databases do not include these new attacks. Thus, real-time intrusion detection systems are required to detect and prevent attacks as soon as they occur. This can be accomplished by continuously monitoring system activities and detecting intrusions in real-time.

   It is undeniable that Machine Learning gained popularity in IDS due to its ability to detect unknown threats. However, classical machine learning-based algorithms are too slow to handle many Gbps of traffic and thus cannot be used in high throughput networks. A possible solution to this problem is to propose two levels of classifiers: one for per-packet detection and another for per-flow detection to compensate for performance and accuracy. The level 1 classifier extracts some selected features from the packet first, allowing for faster classification and real-time attack detection. The level 2 classifier only works with flows not classified by the level 1 classifier (Seo and Pak, 2021).

7.4 Practical and managerial implications of ML and DL-based intrusion detection systems

Practically, ML and DL-based IDS may assist cybersecurity teams focus their attention on genuine threats, reducing the risk of alert fatigue and allowing for more effective incident response strategies. Organizations may mitigate the impact of breaches and potential damages by identifying and containing security incidents as soon as they occur. This scalability is critical for organizations operating in dynamic and changing cyber threat landscapes. While ML and DL-based IDS may offer advanced capabilities for detecting and mitigating cyber threats, effective implementation necessitates careful consideration of managerial implications.

In reality, integrating ML and DL-based intrusion detection systems necessitates a significant investment of resources, including machine learning experts, computational infrastructure, and ongoing maintenance. Managers must allocate adequate resources to ensure the effectiveness and efficiency of these systems. Furthermore, managers must invest in training programs to improve the skills of cybersecurity personnel in ML and DL techniques. This includes understanding how to effectively interpret and act on the systems' outputs. In addition, executives must carefully evaluate vendors that offer ML and DL-based IDS solutions. The robustness of the algorithms, scalability, interoperability with existing systems, and experience dealing with emerging threats are all important considerations.

In addition, compliance requirements such as GDPR, HIPAA, or industry-specific regulations may necessitate certain data handling and privacy considerations when implementing ML and DL-based IDS. Further, compliance requirements such as GDPR (Mohammad Amini et al., 2023), HIPAA (Humphrey, 2021), or industry-specific regulations may require particular data handling and privacy considerations when implementing ML and DL-based IDS. Managers must ensure that these systems follow applicable regulations and standards. While ML and DL-based IDS provide advanced threat detection capabilities, they may introduce new risks such as model bias, adversarial attacks, and interpretability challenges. They must effectively assess and mitigate these risks to ensure that the IDS is reliable and trustworthy. Managers should evaluate the performance of various algorithms and feature sets to improve detection capabilities. Moreover, they should ensure that these systems are seamlessly integrated into the network infrastructure to reduce latency and enhance responsiveness. By effectively addressing these factors, organizations may leverage these advancements to improve their cybersecurity posture and mitigate risks.

8 Comparison with related studies

Several scientific studies on IDS were published in recent years. To evaluate our research, we included a Table 7 comparing our findings to those of other studies. This table serves to delineate the distinctions between our proposed methodology and the findings from existing surveys.

Table 7 compares our proposed study to other surveys on ML and DL-based IDS. We noticed that all of the surveys investigated IDS classification and intelligent techniques in IDS. However, the specific techniques used may vary between surveys. Our study, along with those conducted by Saranya et al. (2020); Si-Ahmed et al. (2023), discussed detection methods in IDS and reviewed IDS-related datasets. Only Saranya et al. (2020), Maseno et al. (2022), and Yi et al. (2023) conducted surveys on specific IDS applications such as IoT, Smart City, fog, and Big Data. Our study, along with those of Maseno et al. (2022) and Yi et al. (2023), examined the metrics and indicators used in IDS. However, only our paper with Haji and Ameen (2021) have focused on the various types of attacks in IDS. Nonetheless, we have not addressed the challenges associated with unbalanced data categories or the processing of high-dimensional mass data in IDS, as Si-Ahmed et al. (2023) and Yi et al. (2023) did in their surveys.

Except for the study by Yi et al. (2023), all of the surveys tackled research challenges related to Machine Learning (ML) and Deep Learning (DL)-based IDS. Our research, as well as that of Haji and Ameen (2021) and Maseno et al. (2022), have identified and discussed new trends in machine learning and deep learning-based IDS. The practical and managerial implications of ML and DL-based IDS have been discussed in our work, in addition to those by Saranya et al. (2020), Maseno et al. (2022), and Yi et al. (2023). Only our survey, as well as those of Haji and Ameen (2021) and Si-Ahmed et al. (2023), compared their findings to other related studies or approaches. Overall, the comparison table address various aspects of IDS. The specific focus and depth of coverage may differ between surveys, highlighting various perspectives and

TABLE 7 Comparison with other similar review articles.

	Our Survey	Saranya et al. (2020)	Haji and Ameen (2021)	Yi et al. (2023)	Maseno et al. (2022)	Si-Ahmed et al. (2023)
IDS classification	X	X	X	X	X	X
Detection method	X	X				X
Intelligent technique Comparison	X	X	X	X	X	X
Discussing specific application of IDS (Iot, Smart City, Fog, Big Data)		X		X	X	
Datasets review	X	X				X
Metrics and indicators review	X			X	X	
Attacks review	X		X			
Unbalanced data categories Processing of high-dimensional mass data				X		X
Research challenges for ML and DL-Based IDS	X	X	X		X	X
New trends in ML and DL-based IDS	X		X		X	
Practical and managerial Implications of ML and DL-Based IDS	X	X		X	X	
Comparison with other	X		X			X

areas of emphasis in the field of IDS research. According to the findings, while all surveys provide valuable insights, our survey stands out because it covers a wide range of concerns. Furthermore, the surveys by Saranya et al. (2020), Maseno et al. (2022), and (Si-Ahmed et al., 2023) are noteworthy studies.

9 Conclusions

This paper presents a comprehensive survey of modern ML and DL-based intrusion detection algorithms, including recent solutions, datasets, metrics, indicators, and detected attacks, to provide valuable insights to researchers in this field. A systematic approach was used to select relevant and recent articles about AI-based IDS. The concept of IDS was extensively discussed, along with its various classification schemes based on the reviewed literature. Furthermore, each article's methodology was examined and their strengths and weaknesses were highlighted regarding intrusion detection capabilities and model complexity. This analysis revealed that recent developments favor DL-based approaches for improving the performance and effectiveness of IDS by increasing accuracy rates and decreasing false alarm rates. DL schemes have outperformed ML-based methods' ability to independently learn features and fit complex models. However, the complexity of DL algorithms requires significant computing resources for processing power and storage capabilities, posing challenges for real-time implementation of intrusion detection systems. Furthermore, the study found that 41% of proposed methodologies were tested on outdated datasets such as KDD Cup'99 and NSL-KDD, limiting their effectiveness in detecting modern network attacks in real-time environments. Addressing these challenges is important for meeting real-time requirements and enhancing IDS performance. It is crucial for AI-based IDS methods to be regularly tested with updated datasets to achieve accurate intrusion detection. Thus, the paper effectively tackled these challenges and projected future developments in ML and DL-based IDS systems. Besides, the proposed survey is evaluated by comparing it to other studies, identifying differences and similarities between our suggested methodology and existing surveys. According to this comparison, while all surveys provided useful information, ours stood out for addressing an extensive variety of concerns.

We noticed through this research study that there are still research gaps, such as improving model performance for specific attacks in real-world environments and finding efficient solutions to reduce complexity. Therefore, future research could focus on developing a lightweight and effective IDS framework that relies on less complex DL algorithms and efficient detection mechanism.

Author contributions

AA: Formal analysis, Writing – original draft, Writing – review & editing. BA: Formal analysis, Resources, Writing – review & editing. MC: Conceptualization, Data curation, Writing – review & editing. BH: Methodology, Project administration, Writing – review & editing. FA: Methodology, Validation, Writing – review & editing. AA: Investigation, Methodology, Writing – review & editing. AH: Funding acquisition, Investigation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research has been supported by the Ministry of Higher Education and Scientific Research of Tunisia under grant agreement number LR11ES48 and the UK Engineering and Physical Sciences Research Council (EPSRC) Grants Ref. EP/T021063/1, EP/T024917/1.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abou Khamis, R., Shafiq, M. O., and Matrawy, A. (2020). "Investigating resistance of deep learning-based ids against adversaries using min-max optimization," in *ICC 2020–2020 IEEE International Conference On Communications (ICC)* (IEEE), 1–7. doi: 10.1109/ICC40277.2020.9149117

Agrawal, S., Sarkar, S., Aouedi, O., Yenduri, G., Piamrat, K., Alazab, M., et al. (2022). Federated learning for intrusion detection system: concepts, challenges and future directions. *Comput. Commun.* 195, 346–361. doi: 10.1016/j.comcom.2022.09.012

Al Tobi, A. M., and Duncan, I. (2018). KDD 1999 generation faults: a review and analysis. *J. Cyber Secur. Technol.* 2, 164–200. doi: 10.1080/23742917.2018.1518061

Alavizadeh, H., Alavizadeh, H., and Jang-Jaccard, J. (2022). Deep q-learning based reinforcement learning approach for network intrusion detection. *Computers* 11:41. doi: 10.3390/computers11030041

Al-Emadi, S., Al-Mohannadi, A., and Al-Senaid, F. (2020). "Using deep learning techniques for network intrusion detection," in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)* (IEEE), 171–176. doi: 10.1109/ICIoT48696.2020.9089524

Alenezi, N., and Aljuhani, A. (2023). Intelligent intrusion detection for industrial internet of things using clustering techniques. *Comput. Syst. Sci. Eng.* 46:36657. doi: 10.32604/csse.2023.036657

Alhajjar, E., Maxwell, P., and Bastian, N. (2021). Adversarial machine learning in network intrusion detection systems. *Exp. Syst. Applic.* 186:115782. doi: 10.1016/j.eswa.2021.115782

Ali Hussein, A., and Boudour Ammar, M. C. B. B. H. (2024b). "Enhanced intrusion detection based hybrid meta-heuristic feature selection," in *16th International Conference on Computational Collective Intelligence.*

Ali Hussein, A., and Maha Charfeddine, B. A. B. B. H. (2024a). "Intrusion detection schemes based on synthetic minority oversampling technique and machine learning models," in *Conference 27th IEEE International Symposium on Real-Time Distributed Computing* (IEEE).

Al-Omari, M., Rawashdeh, M., Qutaishat, F., Alshira, H., M., and Ababneh, N. (2021). An intelligent tree-based intrusion detection model for cyber security. *J. Netw. Syst. Manag.* 29, 1–18. doi: 10.1007/s10922-021-09591-y

Alotaibi, A., and Rassam, M. A. (2023). Adversarial machine learning attacks against intrusion detection systems: a survey on strategies and defense. *Fut. Internet* 15:62. doi: 10.3390/fi15020062

Al-Qatf, M., Lasheng, Y., Al-Habib, M., and Al-Sabahi, K. (2018). Deep learning approach combining sparse autoencoder with SVM for network intrusion detection. *IEEE Access* 6, 52843–52856. doi: 10.1109/ACCESS.2018.2869577

Alsarhan, A., Alauthman, M., Alshdaifat, E., Al-Ghuwairi, A.-R., and Al-Dubai, A. (2021). Machine learning-driven optimization for SVM-based intrusion detection system in vehicular ad hoc networks. *J. Ambient Intell. Hum. Comput.* 14, 6113–6122. doi: 10.1007/s12652-021-02963-x

Alzahrani, A. O., and Alenazi, M. J. (2021). Designing a network intrusion detection system based on machine learning for software defined networks. *Fut. Internet* 13:111. doi: 10.3390/fi13050111

Andresini, G., Appice, A., Di Mauro, N., Loglisci, C., and Malerba, D. (2020). Multi-channel deep feature learning for intrusion detection. *IEEE Access* 8, 53346–53359. doi: 10.1109/ACCESS.2020.2980937

Azizjon, M., Jumabek, A., and Kim, W. (2020). "1D CNN based network intrusion detection with normalization on imbalanced data," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)* (IEEE), 218–224. doi: 10.1109/ICAIIC48513.2020.9064976

Bakhsh, S. A., Khan, M. A., Ahmed, F., Alshehri, M. S., Ali, H., and Ahmad, J. (2023). Enhancing iot network security through deep learning-powered intrusion detection system. *Internet Things* 24:100936. doi: 10.1016/j.iot.2023.100936

Belgrana, F. Z., Benamrane, N., Hamaida, M. A., Chaabani, A. M., and Taleb-Ahmed, A. (2021). "Network intrusion detection system using neural network and condensed nearest neighbors with selection of nsl-kdd influencing features," in *2020 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)* (IEEE), 23–29. doi: 10.1109/IoTaIS50849.2021.9359689

Bhosale, K. S., Nenova, M., and Iliev, G. (2020). "Intrusion detection in communication networks using different classifiers," in *Techno-Societal 2018: Proceedings of the 2nd International Conference on Advanced Technologies for Societal Applications* (Springer), 19–28. doi: 10.1007/978-3-030-16962-6_3

Binbusayyis, A., and Vaiyapuri, T. (2021). Unsupervised deep learning approach for network intrusion detection combining convolutional autoencoder and one-class SVM. *Appl. Intell.* 51, 7094–7108. doi: 10.1007/s10489-021-02205-9

Borkar, A., Donode, A., and Kumari, A. (2017). "A survey on intrusion detection system (ids) and internal intrusion detection and protection system (IIDPS)," in *2017 International Conference on Inventive Computing and Informatics (ICICI)* (IEEE), 949–953. doi: 10.1109/ICICI.2017.8365277

Catillo, M. A. P., and Villano, U. (2023). CPS-GUARD: Intrusion detection for cyber-physical systems and IOT devices using outlier-aware deep autoencoders. *Comput. Secur.* 129:103210. doi: 10.1016/j.cose.2023.103210

Chandra, A., Khatri, S. K., and Simon, R. (2019). "Filter-based attribute selection approach for intrusion detection using k-means clustering and sequential minimal optimization techniq," in *2019 Amity International Conference on Artificial Intelligence (AICAI)* (IEEE), 740–745. doi: 10.1109/AICAI.2019.8701373

Charfeddine, M., Kammoun, H. M., Hamdaoui, B., and Guizani, M. (2024). Chatgpt's security risks and benefits: offensive and defensive use-cases, mitigation measures, and future implications. *IEEE Access* 12, 30263–30310. doi: 10.1109/ACCESS.2024.3367792

Charfeddine, M., Mezghani, E., Masmoudi, S., Amar, C. B., and Alhumyani, H. (2022). Audio watermarking for security and non-security applications. *IEEE Access* 10, 12654–12677. doi: 10.1109/ACCESS.2022.3145950

Choraś, M., and Pawlicki, M. (2021). Intrusion detection approach based on optimised artificial neural network. *Neurocomputing* 452, 705–715. doi: 10.1016/j.neucom.2020.07.138

Das, S., Saha, S., Priyoti, A., Roy, E., Sheldon, F., Haque, A., et al. (2021). Network intrusion detection and comparative analysis using ensemble machine learning and feature selection. *IEEE Trans. Netw. Serv. Manag.* 19, 4821–4833. doi: 10.1109/TNSM.2021.3138457

Deng, Z., Choi, K.-S., Jiang, Y., Wang, J., and Wang, S. (2016). A survey on soft subspace clustering. *Inf. Sci.* 348, 84–106. doi: 10.1016/j.ins.2016.01.101

Dini, P., Elhanashi, A., Begni, A., Saponara, S., Zheng, Q., and Gasmi, K. (2023). Overview on intrusion detection systems design exploiting machine learning for networking cybersecurity. *Appl. Sci.* 13:7507. doi: 10.3390/app13137507

Einy, S., Oz, C., and Navaei, Y. D. (2021). The anomaly-and signature-based ids for network security using hybrid inference systems. *Mathem. Problems Eng.* 2021, 1–10. doi: 10.1155/2021/6639714

Establishment, C. S. (2023). *Communications Security Establishment Annual Report 2022–2023*. Available online at: https://www.cse-cst.gc.ca/en/accountability/transparency/reports/communications-security-establishment-annual-report-2022-2023

Fki, Z., Ammar, B., and Ayed, M. B. (2023). Towards automated optimization of residual convolutional neural networks for electrocardiogram classification. *Cogn. Comput.* 2023, 1–11. doi: 10.1007/s12559-022-10103-6

Frank, N., and Nancy (2019). A critical review on adversarial attacks on intrusion detection systems. *Transl. Eng.* 5:3.

Gómez, Á. L. P., Maimó, L. F., Celdrán, A. H., Clemente, F. J. G., and Cleary, F. (2021). Crafting adversarial samples for anomaly detectors in industrial control systems. *Proc. Comput. Sci.* 184, 573–580. doi: 10.1016/j.procs.2021.03.072

Gu, J., and Lu, S. (2021). An effective intrusion detection approach using svm with naïve bayes feature embedding. *Comput. Secur.* 103:102158. doi: 10.1016/j.cose.2020.102158

Guezzaz, A., Benkirane, S., Azrour, M., and Khurram, S. (2021). A reliable network intrusion detection approach using decision tree with enhanced data quality. *Secur. Commun. Netw.* 2021, 1–8. doi: 10.1155/2021/1230593

Haji, S. H., and Ameen, S. Y. (2021). Attack and anomaly detection in iot networks using machine learning techniques: a review. *Asian J Res. Comput. Sci.* 9, 30–46. doi: 10.9734/ajrcos/2021/v9i230218

Hameed, A., Fourati, R., Ammar, B., Ksibi, A., Alluhaidan, A. S., Ayed, M. B., et al. (2024). Temporal-spatial transformer based motor imagery classification for BCI using independent component analysis. *Biomed. Signal Proc. Control* 87:105359. doi: 10.1016/j.bspc.2023.105359

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., et al. (2024). Interpreting black-box models: a review on explainable artificial intelligence. *Cogn. Comput.* 16, 45–74. doi: 10.1007/s12559-023-10179-8

He, K., Kim, D. D., and Asghar, M. R. (2023). Adversarial machine learning for network intrusion detection systems: a comprehensive survey. *IEEE Commun. Surv. Tutor.* 25, 538–566. doi: 10.1109/COMST.2022.3233793

Hindy, H., Brosset, D., Bayne, E., Seeam, A. K., Tachtatzis, C., Atkinson, R., et al. (2020). A taxonomy of network threats and the effect of current datasets on intrusion detection systems. *IEEE Access* 8, 104650–104675. doi: 10.1109/ACCESS.2020.3000179

Hnamte, V., Nhung-Nguyen, H., Hussain, J., and Hwa-Kim, Y. (2023). A novel two-stage deep learning model for network intrusion detection: LSTM-AE. *IEEE Access*. 11, 37131–37148. doi: 10.1109/ACCESS.2023.3266979

Humphrey, B. A. (2021). *Data privacy vs. innovation: A quantitative analysis of artificial intelligence in healthcare and its impact on HIPAA regarding the privacy and security of protected health information*. Robert Morris University ProQuest Dissertation & Theses.

Jatti, S. A. V., and Sontif, V. K. (2019). Intrusion detection systems. *Int. J. Recent Technol. Eng.* 8, 3976–3983. doi: 10.35940/ijrte.B1540.0982S1119

Javed, A. R., Saadia, A., Mughal, H., Gadekallu, T. R., Rizwan, M., Maddikunta, P. K. R., et al. (2023). Artificial intelligence for cognitive health assessment: state-of-the-art, open challenges and future directions. *Cogn. Comput.* 15, 1767–1812. doi: 10.1007/s12559-023-10153-4

Jia, Y., Wang, M., and Wang, Y. (2019). Network intrusion detection algorithm based on deep neural network. *IET Inf. Secur.* 13, 48–53. doi: 10.1049/iet-ifs.2018.5258

Jiang, K., Wang, W., Wang, A., and Wu, H. (2020). Network intrusion detection combined hybrid sampling with deep hierarchical network. *IEEE Access* 8, 32464–32476. doi: 10.1109/ACCESS.2020.2973730

Jony, A. I., and Arnob, A. K. B. (2024). A long short-term memory-based approach for detecting cyber attacks in IOT using cic-iot2023 dataset. *J. Edge Comput.* 3:48. doi: 10.55056/jec.648

Karatas, G., Demir, O., and Sahingoz, O. K. (2020). Increasing the performance of machine learning-based idss on an imbalanced and up-to-date dataset. *IEEE Access* 8, 32150–32162. doi: 10.1109/ACCESS.2020.2973219

Kasongo, S. M. (2023). A deep learning technique for intrusion detection system using a recurrent neural networks based framework. *Comput. Commun.* 199, 113–125. doi: 10.1016/j.comcom.2022.12.010

Kavitha, S., and Manikandan, J. (2022). Design of a bottleneck layered dnn algorithm for intrusion detection system. *Methods* 3, 242–258. doi: 10.36548/jsws.2021.4.004

Khalid Albulayhi, Q. A. A.-H. (2023). Adversarial deep learning in anomaly based intrusion detection systems for IOT environments. *Int. J. Wirel. Microw. Technol.* 13, 1–10. doi: 10.5815/ijwmt.2023.04.01

Khan, M. A., and Kim, J. (2020). Toward developing efficient conv-ae-based intrusion detection system using heterogeneous dataset. *Electronics* 9:1771. doi: 10.3390/electronics9111771

Khan, R. U., Zhang, X., Alazab, M., and Kumar, R. (2019). "An improved convolutional neural network model for intrusion detection in networks," in *2019 Cybersecurity and Cyberforensics Conference (CCC)* (IEEE), 74–77. doi: 10.1109/CCC.2019.000-6

Kheddar, H., Himeur, Y., and Awad, A. I. (2023). Deep transfer learning for intrusion detection in industrial control networks: a comprehensive review. *J. Netw. Comput. Applic.* 220:103760. doi: 10.1016/j.jnca.2023.103760

Kim, S., Razi, A., Stringhini, G., Wisniewski, P. J., and De Choudhury, M. (2021). A human-centered systematic literature review of cyberbullying detection algorithms. *Proc. ACM Hum. Comput. Inter.* 5, 1–34. doi: 10.1145/3476066

Kumar, S., and Pathak, N. K. (2022). Evaluation of machine learning algorithms for intrusion detection utilizing UNSW-NB15 dataset. *J. Pharm. Negat. Results* 13, 4819–4832. doi: 10.1109/SILCON59133.2023.10404204

Kunhare, N., and Tiwari, R. (2018). "Study of the attributes using four class labels on KDD99 and NSL-KDD datasets with machine learning techniques," in *2018 8th International Conference on Communication Systems and Network Technologies (CSNT)* (IEEE), 127–131. doi: 10.1109/CSNT.2018.8820244

Kurniawan, Y. I., Razi, F., Nofiyati, N., Wijayanto, B., and Hidayat, M. L. (2021). Naive bayes modification for intrusion detection system classification with zero probability. *Bull. Electr. Eng. Inf.* 10, 2751–2758. doi: 10.11591/eei.v10i5.2833

Lansky, J., Ali, S., Mohammadi, M., Majeed, M. K., Karim, S. H. T., Rashidi, S., et al. (2021). Deep learning-based intrusion detection systems: a systematic review. *IEEE Access* 9, 101574–101599. doi: 10.1109/ACCESS.2021.3097247

Latif, S., Boulila, W., Koubaa, A., Zou, Z., and Ahmad, J. (2024). DTL-IDS: an optimized intrusion detection framework using deep transfer learning and genetic algorithm. *J. Netw. Comput. Applic.* 221:103784. doi: 10.1016/j.jnca.2023.103784

Li, F., Shinde, A., Shi, Y., Ye, J., Li, X.-Y., and Song, W. (2019). System statistics learning-based IOT security: feasibility and suitability. *IEEE Internet Things J.* 6, 6396–6403. doi: 10.1109/JIOT.2019.2897063

Liu, C., Gu, Z., and Wang, J. (2021). A hybrid intrusion detection system based on scalable k-means+ random forest and deep learning. *IEEE Access* 9, 75729–75740. doi: 10.1109/ACCESS.2021.3082147

Madwanna, Y., B., A., R., R. A., R., et al. (2023). "YARS-IDS: a novel ids for multi-class classification," in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, 1–6. doi: 10.1109/I2CT57861.2023.10126301

Martins, N., Cruz, J. M., Cruz, T., and Abreu, P. H. (2020). Adversarial machine learning applied to intrusion and malware scenarios: a systematic review. *IEEE Access* 8, 35403–35419. doi: 10.1109/ACCESS.2020.2974752

Maseer, Z. K., Yusof, R., Bahaman, N., Mostafa, S. A., and Foozy, C. F. M. (2021). Benchmarking of machine learning for anomaly based intrusion detection systems in the CICIDS2017 dataset. *IEEE Access* 9, 22351–22370. doi: 10.1109/ACCESS.2021.3056614

Maseno, E. M., Wang, Z., and Xing, H. (2022). A systematic review on hybrid intrusion detection system. *Secur. Commun. Netw.* 2022:9663052. doi: 10.1155/2022/9663052

Mbona, I., and Eloff, J. H. (2022). Detecting zero-day intrusion attacks using semi-supervised machine learning approaches. *IEEE Access* 10, 69822–69838. doi: 10.1109/ACCESS.2022.3187116

Michelena, Á., Aveleira-Mata, J., Jove, E., Bayón-Gutiérrez, M., Novais, P., Romero, O. F., et al. (2024). A novel intelligent approach for man-in-the-middle attacks detection over internet of things environments based on message queuing telemetry transport. *Expert Syst.* 41:e13263. doi: 10.1111/exsy.13263

Miller, D. J., Xiang, Z., and Kesidis, G. (2020). Adversarial learning targeting deep neural network classification: a comprehensive review of defenses against attacks. *Proc. IEEE* 108, 402–433. doi: 10.1109/JPROC.2020.2970615

Mittal, M., Iwendi, C., Khan, S., and Rehman Javed, A. (2021). Analysis of security and energy efficiency for shortest route discovery in low-energy adaptive clustering hierarchy protocol using levenberg-marquardt neural network and gated recurrent unit for intrusion detection system. *Trans. Emerg. Telecommun. Technol.* 32:e3997. doi: 10.1002/ett.3997

Mohammad Amini, M., Jesus, M., Fanaei Sheikholeslami, D., Alves, P., Hassanzadeh Benam, A., and Hariri, F. (2023). Artificial intelligence ethics and challenges in healthcare applications: a comprehensive review in the context of the european gdpr mandate. *Mach. Learn. Knowl. Extr.* 5, 1023–1035. doi: 10.3390/make5030053

Mohammadi, M., Rashid, T. A., Karim, S. H. T., Aldalwie, A. H. M., Tho, Q. T., Bidaki, M., et al. (2021). A comprehensive survey and taxonomy of the svm-based intrusion detection systems. *J. Netw. Comput. Applic.* 178:102983. doi: 10.1016/j.jnca.2021.102983

Naseer, S., Saleem, Y., Khalid, S., Bashir, M. K., Han, J., Iqbal, M. M., et al. (2018). Enhanced network anomaly detection based on deep neural networks. *IEEE access* 6, 48231–48246. doi: 10.1109/ACCESS.2018.2863036

Nguyen, T. T., and Reddi, V. J. (2023). Deep reinforcement learning for cyber security. *IEEE Trans. Neural Netw. Learn. Syst.* 34, 3779–3795. doi: 10.1109/TNNLS.2021.3121870

Ogundokun, R. O., Awotunde, J. B., Sadiku, P., Adeniyi, E. A., Abiodun, M., and Dauda, O. I. (2021). An enhanced intrusion detection system using particle swarm optimization feature extraction technique. *Proc. Comput. Sci.* 193, 504–512. doi: 10.1016/j.procs.2021.10.052

Oprea, S.-V., Bâra, A., Puican, F. C., and Radu, I. C. (2021). Anomaly detection with machine learning algorithms and big data in electricity consumption. *Sustainability* 13:10963. doi: 10.3390/su131910963

Panagiotou, P., Mengidis, N., Tsikrika, T., Vrochidis, S., and Kompatsiaris, I. (2021). Host-based intrusion detection using signature-based and ai-driven anomaly detection methods. *Inf. Secur.* 50, 37–48. doi: 10.11610/isij.5016

Pande, S., and Khamparia, A. (2023). Explainable deep neural network based analysis on intrusion detection systems. *Comput. Sci.* 24:4551. doi: 10.7494/csci.2023.24.1.4551

Prasath, S., Sethi, K., Mohanty, D., Bera, P., and Samantaray, S. R. (2022). Analysis of continual learning models for intrusion detection system. *IEEE Access* 10, 121444–121464. doi: 10.1109/ACCESS.2022.3222715

Rahman, M. A., Asyhari, A. T., Wen, O. W., Ajra, H., Ahmed, Y., and Anwar, F. (2021). Effective combining of feature selection techniques for machine learning-enabled iot intrusion detection. *Multim. Tools Applic.* 80, 31381–31399. doi: 10.1007/s11042-021-10567-y

Riyaz, B., and Ganapathy, S. (2020). A deep learning approach for effective intrusion detection in wireless networks using cnn. *Soft Comput.* 24, 17265–17278. doi: 10.1007/s00500-020-05017-0

RM, S. P., Maddikunta, P. K. R., Parimala, M., Koppu, S., Gadekallu, T. R., Chowdhary, C. L., et al. (2020). An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IOMT architecture. *Comput. Commun.* 160, 139–149. doi: 10.1016/j.comcom.2020.05.048

Saba, T., Rehman, A., Sadad, T., Kolivand, H., and Bahaj, S. A. (2022). Anomaly-based intrusion detection system for IOT networks through deep learning model. *Comput. Electr. Eng.* 99:107810. doi: 10.1016/j.compeleceng.2022.107810

Saheed, Y. K., Usman, A. A., Sukat, F. D., and Abdulrahman, M. (2023). A novel hybrid autoencoder and modified particle swarm optimization feature selection for intrusion detection in the internet of things network. *Front. Comput. Sci.* 5:997159. doi: 10.3389/fcomp.2023.997159

Salo, F., Injadat, M., Moubayed, A., Nassif, A. B., and Essex, A. (2019). "Clustering enabled classification using ensemble feature selection for intrusion detection," in *2019 International Conference on Computing, Networking and Communications (ICNC)* (IEEE), 276–281. doi: 10.1109/ICCNC.2019.8685636

Salvakkam, D. B., Saravanan, V., Jain, P. K., and Pamula, R. (2023). Enhanced quantum-secure ensemble intrusion detection techniques for cloud based on deep learning. *Cogn. Comput.* 15, 1593–1612. doi: 10.1007/s12559-023-10139-2

Saranya, T., Sridevi, S., Deisy, C., Chung, T. D., and Khan, M. A. (2020). Performance analysis of machine learning algorithms in intrusion detection system: a review. *Proc. Comput. Sci.* 171, 1251–1260. doi: 10.1016/j.procs.2020.04.133

Seo, W., and Pak, W. (2021). Real-time network intrusion prevention system based on hybrid machine learning. *IEEE Access* 9, 46386–46397. doi: 10.1109/ACCESS.2021.3066620

Si-Ahmed, A., Al-Garadi, M. A., and Boustia, N. (2023). Survey of machine learning based intrusion detection methods for internet of medical things. *Appl. Soft Comput.* 140:110227. doi: 10.1016/j.asoc.2023.110227

Singhal, A., Maan, A., Chaudhary, D., and Vishwakarma, D. (2021). "A hybrid machine learning and data mining based approach to network intrusion detection," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)* (IEEE), 312–318. doi: 10.1109/ICAIS50930.2021.9395918

Sultana, N., Chilamkurti, N., Peng, W., and Alhadad, R. (2019). Survey on SDN based network intrusion detection system using machine learning approaches. *Peer-to-Peer Network. Applic.* 12, 493–501. doi: 10.1007/s12083-017-0630-0

Sumaiya Thaseen, I., Saira Banu, J., Lavanya, K., Rukunuddin Ghalib, M., and Abhishek, K. (2021). An integrated intrusion detection system using correlation-based attribute selection and artificial neural network. *Trans. Emerg. Telecommun. Technol.* 32:e4014. doi: 10.1002/ett.4014

Süzen, A. A. (2021). Developing a multi-level intrusion detection system using hybrid-DBN. *J. Ambient Intell. Human. Comput.* 12, 1913–1923. doi: 10.1007/s12652-020-02271-w

Szyller, S., Atli, B. G., Marchal, S., and Asokan, N. (2021). "DAWN: dynamic adversarial watermarking of neural networks," in *Proceedings of the 29th ACM International Conference on Multimedia*, 4417–4425. doi: 10.1145/3474085.3475591

Talukder, M. A., Hasan, K. F., Islam, M. M., Uddin, M. A., Akhter, A., Yousuf, M. A., et al. (2023). A dependable hybrid machine learning model for network intrusion detection. *J. Inf. Secur. Applic.* 72:103405. doi: 10.1016/j.jisa.2022.103405

Tan, X., Su, S., Zuo, Z., Guo, X., and Sun, X. (2019). Intrusion detection of UAVS based on the deep belief network optimized by PSO. *Sensors* 19:5529. doi: 10.3390/s19245529

Tang, T. A., McLernon, D., Mhamdi, L., Zaidi, S. A. R., and Ghogho, M. (2019). "Intrusion detection in sdn-based networks: deep recurrent neural network approach," in *Deep Learning Applications for Cyber Security* (Springer), 175–195. doi: 10.1007/978-3-030-13057-2_8

Ullah, I., and Mahmoud, Q. H. (2020). "A technique for generating a botnet dataset for anomalous activity detection in iot networks," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 134–140. doi: 10.1109/SMC42975.2020.9283220

Vangipuram, R., Gunupudi, R. K., Puligadda, V. K., and Vinjamuri, J. (2020). A machine learning approach for imputation and anomaly detection in iot environment. *Expert Syst.* 37:e12556. doi: 10.1111/exsy.12556

Wang, Y.-C., Houng, Y.-C., Chen, H.-X., and Tseng, S.-M. (2023). Network anomaly intrusion detection based on deep learning approach. *Sensors* 23:2171. doi: 10.3390/s23042171

Wester, P., Heiding, F., and Lagerström, R. (2021). "Anomaly-based intrusion detection using tree augmented naive bayes," in *2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW)*, 112–121. doi: 10.1109/EDOCW52865.2021.00040

Wisanwanichthan, T., and Thammawichai, M. (2021). A double-layered hybrid approach for network intrusion detection system using combined naive bayes and svm. *IEEE Access* 9, 138432–138450. doi: 10.1109/ACCESS.2021.3118573

Wu, Z., Wang, J., Hu, L., Zhang, Z., and Wu, H. (2020). A network intrusion detection method based on semantic re-encoding and deep learning. *J. Netw. Comput. Applic.* 164:102688. doi: 10.1016/j.jnca.2020.102688

Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., et al. (2018). Machine learning and deep learning methods for cybersecurity. *IEEE Access* 6, 35365–35381. doi: 10.1109/ACCESS.2018.2836950

Xu, C., Shen, J., Du, X., and Zhang, F. (2018). An intrusion detection system using a deep neural network with gated recurrent units. *IEEE Access* 6, 48697–48707. doi: 10.1109/ACCESS.2018.2867564

Yang, Y., Zheng, K., Wu, B., Yang, Y., and Wang, X. (2020). Network intrusion detection based on supervised adversarial variational auto-encoder with regularization. *IEEE Access* 8, 42169–42184. doi: 10.1109/ACCESS.2020.2977007

Yi, T., Chen, X., Zhu, Y., Ge, W., and Han, Z. (2023). Review on the application of deep learning in network attack detection. *J. Netw. Comput. Applic.* 212:103580. doi: 10.1016/j.jnca.2022.103580

Younisse, R., Ahmad, A., and Abu Al-Haija, Q. (2022). Explaining intrusion detection-based convolutional neural networks using shapley additive explanations (shap). *Big Data Cogn. Comput.* 6:126. doi: 10.3390/bdcc6040126

Yu, Y., and Bian, N. (2020). An intrusion detection method using few-shot learning. *IEEE Access* 8, 49730–49740. doi: 10.1109/ACCESS.2020.2980136

Zhang, X., Chen, J., Zhou, Y., Han, L., and Lin, J. (2019). A multiple-layer representation learning model for network-based attack detection. *IEEE Access* 7, 91992–92008. doi: 10.1109/ACCESS.2019.2927465

Zhu, J., and Liu, X. (2024). An integrated intrusion detection framework based on subspace clustering and ensemble learning. *Comput. Electr. Eng.* 115:109113. doi: 10.1016/j.compeleceng.2024.109113

Zou, L., Luo, X., Zhang, Y., Yang, X., and Wang, X. (2023). HC-DTTSVM: a network intrusion detection method based on decision tree twin support vector machine and hierarchical clustering. *IEEE Access* 11, 21404–21416. doi: 10.1109/ACCESS.2023.3251354