Securing IoT Device, Machine learning based Intrusion detection system

[1] Mrs. Tejshri Niranjan Shevate , [2] Dr. Balendra Kumar Garg , [3]Dr R. D. Kumbhar

[1]Research Scholar, [2]Assistant Professor , [3]Assistant Professor

[1,22]Department of Computer Science and Application Technology
[1,2]MATS university Raipur Chhattisgarh, India
[3] KBPIMSR Varye Satara, India

Abstract-

Connected devices have become deeply embedded in modern lifestyles under the paradigm known as the Internet of Things (IoT). These technologies are now widely deployed across diverse application domains, including industrial automation, smart cities, energy systems, healthcare services, and residential environments. Despite their benefits, IoT systems introduce significant security challenges arising from inherent limitations such as constrained computational resources, heterogeneous architectures, and infrequent firmware updates. These vulnerabilities make IoT environments attractive targets for cyberattacks. To address these concerns, research efforts have increasingly focused on the development of Intrusion Detection Systems (IDS) tailored to IoT networks. The effectiveness of IDS can be substantially enhanced through the incorporation of machine learning (ML) techniques, which enable improved detection accuracy, adaptability to evolving attack patterns, identification of previously unseen threats, and reduction of false alarm rates.

This study presents a comprehensive review of recent research that integrates machine learning approaches into intrusion detection mechanisms for IoT systems. The surveyed methodologies are systematically evaluated using multiple criteria, including feature selection strategies, learning algorithms, detection performance, false positive rates, and the types of security threats addressed. A critical assessment of each approach is provided to highlight strengths, limitations, and practical applicability. Additionally, the paper examines widely used benchmark datasets that contain intentionally generated attacks within experimental IoT network environments. Particular emphasis is placed on anomaly-based intrusion detection techniques employing machine learning models, offering an in-depth analysis of their design and performance. Finally, a comparative discussion of existing methods is conducted to identify unresolved challenges and future research directions. To facilitate clarity and comparison, concise summary tables are included to outline the key characteristics of each proposed model.

Keywords— Internet of Things (IoT), Intrusion Detection System (IDS), Machine Learning, Deep Learning, Network Security, DoS/DDoS Attacks, Feature Selection, Dataset Preprocessing, NSL-KDD, Performance Evaluation Metrics

I.INTRODUCTION

Rapid advancements across multiple technological domains—including sensor technologies, embedded systems, high-speed internet, and next-generation wireless communications such as 5G and the emerging 6G, which is expected to deliver data rates up to one hundred times faster than its predecessor—have enabled the widespread adoption of intelligent devices in everyday life. The interconnection of these devices through the Internet has led to the emergence of the Internet of Things (IoT) paradigm. According to statistical projections reported by Statista, the global number of connected devices is anticipated to reach approximately 30.9 billion by 2025 [1]. Furthermore, the IoT market is projected to experience sustained growth between 2024 and 2028, with expected revenues reaching $145.4 billion, corresponding to an increase of nearly 54.75% [2]. IoT technologies are now deployed across a wide range of sectors, including industrial systems, smart cities, energy management, healthcare services, smart residential environments, and agricultural applications [3].

Despite these benefits, the rapid expansion of IoT infrastructures has significantly increased the attack surface of digital systems, giving rise to serious security concerns. Cyberattacks targeting critical infrastructures—such as defense installations and nuclear facilities—have become increasingly frequent and sophisticated, posing severe risks not only to national security but also to environmental safety [4]. Moreover, the heterogeneity of IoT ecosystems, characterized by diverse communication protocols, hardware platforms, and standards, presents substantial challenges for conventional security mechanisms, which often struggle to adapt effectively to such dynamic environments.

To mitigate these risks, a variety of security solutions have been proposed, including cryptographic approaches such as homomorphic encryption [5], robust authentication mechanisms [6], and trust management techniques designed to ensure reliable interactions among IoT components [7]. While these methods contribute to strengthening system security, they remain largely inadequate in addressing network-centric attacks, particularly those that evolve rapidly or exploit unknown vulnerabilities. This limitation highlights the need for an additional defense layer capable of detecting and responding to malicious activities in a timely and proactive manner. Intrusion Detection Systems (IDS) offer such a capability by continuously monitoring network traffic and system behavior to identify abnormal or unauthorized activities. However, traditional IDS solutions are often unsuitable for IoT environments due to resource constraints and their inability to detect previously unseen attack patterns. Consequently, recent research has increasingly focused on anomaly-based IDS approaches enhanced with machine learning (ML) techniques, which leverage data-driven models to identify novel threats and adapt to evolving attack behaviors.

This study presents a comprehensive review of recent research focusing on the integration of Intrusion Detection Systems (IDS) with machine learning (ML) techniques for IoT environments. The main contributions of this work can be summarized as follows. First, an overview of the IDS landscape is provided by systematically classifying existing IDS types and outlining commonly adopted detection strategies. Second, the study examines the key components involved in IDS model evaluation by identifying widely used benchmark datasets that include intentionally generated attacks on experimental IoT networks, along with a discussion of feature reduction methods and performance evaluation metrics. Third, an in-depth analysis of state-of-the-art intrusion detection approaches published between 2017 and 2023 is conducted, with particular emphasis on anomaly-based IDS models that employ machine learning techniques. Finally, a comparative discussion of the reviewed methods is presented to highlight unresolved challenges in IoT-oriented IDS development and to outline promising directions for future research.

II. INTRUSION DETECTION SYSTEMS (IDS) ECOSYSTEM: AN OVERVIEW

Intrusion detection refers to the systematic and continuous observation of network traffic and system activities with the objective of identifying abnormal or unauthorized behavior. When implemented as a software-based solution, this process is commonly referred to as an Intrusion Detection System (IDS). IDS constitutes a critical component of network security, as it aims to detect and report potential security threats before they result in service disruption, unauthorized system access, or data breaches. IDS solutions are generally classified into two main categories: Host-based Intrusion Detection Systems (HIDS) [8] and Network-based Intrusion Detection Systems (NIDS) [9].

Host-based IDS operate at the individual device level by monitoring system-specific activities, including modifications to system and application files, memory usage patterns, and local process behavior. This approach requires deployment on each host within the network, enabling the detection of malicious activities that target a particular device. In contrast, Network-based IDS are positioned at strategic points within the network infrastructure to passively inspect traffic flowing across network segments. NIDS solutions may be implemented as either hardware appliances or software applications and are designed to interface with various networking technologies, such as Ethernet and Fiber Distributed Data Interface (FDDI), depending on vendor specifications. Typically, NIDS employ dual network interfaces, where one interface operates in promiscuous mode to capture and analyze network packets, while the second interface is used for management, control, and alert reporting. The fundamental distinction between these two IDS types lies in their monitoring scope: HIDS concentrates on safeguarding individual hosts through internal system analysis, whereas NIDS focuses on detecting anomalous behavior across the network by examining aggregate traffic patterns. In addition to deployment-based classifications, Intrusion Detection Systems can also be categorized according to their detection strategies. One commonly adopted approach is the signature-based detection method [10], which relies on a predefined database of known attack patterns. This technique continuously inspects network or host-level activities and compares observed behavior against stored signatures generated by either host-based or network-based IDS solutions. While signature-based detection is effective in accurately identifying previously known attacks with low false alarm rates, its primary limitation lies in its inability to detect novel or zero-day threats. Consequently, the effectiveness of this approach depends heavily on the frequent updating of the signature repository, and delays in updates may leave systems vulnerable to emerging attacks that evade pattern matching mechanisms.

Another widely used strategy is anomaly-based intrusion detection [11], which operates by establishing a model of normal system or network behavior and subsequently monitoring ongoing activities for significant deviations from this baseline. Any detected anomaly is treated as a potential security threat, triggering alerts or automated response mechanisms within an Intrusion Detection and Prevention System (IDPS). Owing to its behavior-based nature, anomaly detection is capable of identifying previously unseen attack patterns that do not correspond to known signatures. Moreover, this approach is inherently adaptive, as it can learn from evolving traffic patterns and update the baseline model accordingly. Despite these advantages, anomaly-based IDS faces notable challenges, particularly in accurately defining and maintaining a representative baseline in heterogeneous and dynamic network environments. In addition, false negatives may occur when malicious traffic closely resembles legitimate behavior, potentially allowing attacks to go undetected.

To address the limitations of individual detection strategies, hybrid IDS approaches have been proposed that integrate both signature-based and anomaly-based techniques. By combining the precision of signature matching with the adaptability of anomaly detection, hybrid IDS models aim to achieve improved detection performance and enhanced resilience against a wider range of attack scenarios.

3.KEY ELEMENTS FOR ML MODEL ANALYSIS

Before examining the various machine learning techniques applied to intrusion detection systems, it is essential to address three fundamental aspects that significantly influence IDS performance: the selection of appropriate datasets, dataset preprocessing procedures, and performance evaluation metrics. This section primarily emphasizes data preprocessing, with particular attention to feature selection and feature extraction techniques, and subsequently discusses the metrics commonly employed to assess model effectiveness.

3.1 Common Datasets

The evaluation of machine learning–based intrusion detection models requires the use of carefully selected datasets that offer both sufficient data volume and high-quality, representative features, while accurately reflecting the attack scenarios and system behaviors under investigation. The choice of dataset plays a critical role in determining the reliability, generalizability, and robustness of the proposed detection models. As discussed in [12], an effective dataset should encompass realistic traffic patterns, diverse attack types, and well-defined ground truth labels. In this context, the following subsection presents some of the most widely used benchmark datasets within the IoT intrusion detection research domain.

One of the most widely adopted benchmark datasets in intrusion detection research is the **CICIDS-2017 dataset**, developed by the Canadian Institute for Cybersecurity. This dataset is publicly available for academic use and is designed to reflect realistic network traffic conditions. It contains 15 labeled classes derived from seven major attack categories, including Denial of Service (DoS), Port Scanning, and Brute Force attacks. A total of 80 traffic features are extracted from diverse network services such as FTP, HTTP, SSH, and email protocols, making the dataset suitable for a wide range of intrusion detection experiments. Due to its richness and diversity, CICIDS-2017 has been extensively used in recent IDS studies [13, 14].

Another prominent dataset is **UNSW-NB15**, generated by the University of New South Wales in Australia. This dataset was created to overcome the limitations of older intrusion detection benchmarks by incorporating contemporary attack behaviors. It includes nine distinct attack categories, such as Fuzzers, Denial of Service, and Reconnaissance attacks. A total of 49 features are extracted from raw network traffic, enabling effective application of machine learning algorithms. With approximately 2.5 million records, UNSW-NB15 provides a substantial volume of data for training and evaluating robust network-based intrusion detection systems and has been widely employed in the literature [15–17].

The **NSL-KDD dataset** is an improved version of the original KDD Cup 1999 dataset, specifically designed to address issues related to redundancy and bias present in the original data. It consists of approximately 150,000 instances described by 41 features while maintaining the same attack taxonomy. The attacks are grouped into four primary categories: Denial of Service (DoS), User-to-Root (U2R), Remote-to-Local (R2L), and Probe attacks. Despite its age, NSL-KDD remains a popular benchmark for comparative analysis and performance evaluation in IDS research [13, 14, 18].

The **IoT-23 dataset**, developed by the Network Security Lab at the University of Colorado Boulder, is specifically tailored for IoT security research. It comprises 20 extracted features and includes traffic generated

from 20 distinct malware samples, such as Mirai, Torii, Gagfyt, and Kenjiro, executed on real IoT devices. In addition to malicious traffic, the dataset also contains benign traffic captures from three IoT devices, enabling effective evaluation of anomaly detection techniques in IoT environments. This dataset has been utilized in several recent studies focusing on IoT malware detection [19, 20].

The **Kyoto dataset** consists of real network traffic collected on a daily basis between 2006 and 2009. It includes 24 features extracted from network sessions and provides a large-scale dataset comprising over 50 million normal traffic sessions, more than 43 million confirmed attack sessions, and approximately 425,000 sessions labeled as unknown attacks. Due to its scale and real-world nature, the Kyoto dataset is valuable for studying long-term network behavior and intrusion detection under realistic operating conditions.

The **CSE-CIC-IDS-2018 dataset**, also developed by the Canadian Institute for Cybersecurity, extends earlier CIC datasets by incorporating more recent attack scenarios and traffic patterns. It contains 80 features and includes a broad range of attacks, such as Brute Force, Port Scanning, Botnet activity, Denial of Service, and Distributed Denial of Service attacks. The dataset is publicly available and is frequently used for evaluating advanced machine learning and deep learning–based IDS models.

Finally, the **AFDA dataset**, developed by the University of New South Wales, is specifically designed to address advanced and emerging threats. It focuses on sophisticated attack types, including zero-day exploits, stealth-based attacks, and webshell intrusions. This dataset is particularly relevant for evaluating IDS solutions aimed at detecting previously unseen or highly evasive attack behaviors.

3.2 Preprocessing phase: Feature selection (FS)

Feature Selection (FS) constitutes a critical preprocessing stage in machine learning–based intrusion detection systems, as it aims to identify the most informative features while eliminating irrelevant or redundant attributes. By reducing the dimensionality of the dataset, FS helps to lower computational overhead, mitigate overfitting, and improve model interpretability. The primary objective of FS is to construct a compact feature subset that adequately captures the underlying characteristics of the detection problem. Commonly adopted FS approaches include machine learning–based methods, filter-based techniques, and wrapper-based strategies. Wrapper methods evaluate candidate feature subsets by directly assessing their impact on the performance of a chosen learning algorithm, thereby enabling the selection of features that maximize classification accuracy. In contrast, filter-based methods operate independently of any specific learning model and rely on statistical measures to rank features, making them computationally efficient for eliminating duplicated, highly correlated, or irrelevant attributes. However, filter methods may be limited in addressing complex feature dependencies and multicollinearity.

A number of studies in the literature have demonstrated that applying FS techniques can significantly enhance detection accuracy while reducing false positive rates. In [21], the authors selected 23 features from the NSL-KDD dataset based on their variability under abnormal conditions, indicating their relevance for intrusion detection. A Random Forest (RF) algorithm was then employed to estimate the importance of these features. Using a wrapper-based preprocessing approach, the method ultimately identified the five most influential features for classification. Nevertheless, such approaches do not guarantee that excluded features are entirely non-contributory for all datasets or attack scenarios. In another study, the authors of [22] proposed a wireless network threat detection framework that integrates deep learning with a wrapper-based FS technique to reduce feature dimensionality. From an initial set of 154 attributes, the wrapper method selected 26 features, leading to improved detection performance. Additionally, the work presented in [23] introduced a Tree-Seed Algorithm (TSA) for feature extraction, aiming to remove redundant attributes and reduce data dimensionality. The proposed approach successfully reduced the feature set from 41 to 6 features, achieving a classification accuracy of 87.34%, thereby demonstrating the effectiveness of FS in enhancing model performance.Overall, the reviewed studies consistently indicate that the application of feature selection techniques contributes positively to improving the accuracy and efficiency of machine learning–based intrusion detection models.

4. NEXT-GEN NIDS FOR IOT: ML-BASED APPROACH

Machine learning–based approaches have demonstrated strong effectiveness in analyzing real-time network traffic and consistently outperform many traditional intrusion detection techniques. These methods operate by learning characteristic patterns associated with normal system behavior, enabling the identification of

deviations that may indicate malicious activity. In this section, machine learning–driven intrusion detection techniques are categorized into three primary classes: supervised learning approaches, unsupervised learning approaches, and deep learning–based methods. Each category is examined in terms of its underlying principles and applicability to intrusion detection in IoT environments.

4.1. Most ML algorithms used in IDS

This section outlines the two most commonly adopted categories of machine learning techniques used in intrusion detection, namely supervised and unsupervised learning methods. Within the supervised learning paradigm, widely employed algorithms include Artificial Neural Networks (ANN) [14] and Support Vector Machines (SVM) [18]. Artificial Neural Networks are computational frameworks inspired by the biological neural architecture of the human brain. They consist of interconnected processing units, referred to as neurons, arranged in multiple layers. Each neuron processes incoming data through weighted connections and activation functions before transmitting the output to subsequent layers. Although ANNs are capable of modeling complex, non-linear relationships, they are often associated with notable challenges, such as high computational complexity, significant training time, and the requirement for large volumes of labeled data. Furthermore, ANNs are susceptible to overfitting, which may limit their ability to generalize effectively to previously unseen data.Support Vector Machines are frequently adopted as an alternative supervised learning approach due to their strong generalization capabilities. The primary objective of SVM is to determine an optimal separating hyperplane that maximizes the margin between different classes within the feature space. By focusing on boundary samples, known as support vectors, SVM exhibits robustness to noise and performs well in high-dimensional settings, making it particularly suitable for intrusion detection tasks involving complex feature spaces.

In the context of unsupervised learning methods, Deep Neural Networks (DNN) and Convolutional Neural Networks (CNN) have gained increasing attention for their effectiveness in intrusion detection applications. Convolutional Neural Networks [16], originally developed for image and pattern recognition tasks, extend traditional neural network architectures by incorporating convolutional and pooling layers. These layers enable CNNs to automatically extract salient features from structured input data by applying learnable filters and reducing dimensionality, respectively, before classification is performed through fully connected layers. This hierarchical feature extraction capability makes CNNs well suited for capturing spatial and temporal patterns in network traffic data. Deep Neural Networks [36], in contrast, are characterized by the presence of multiple hidden layers, distinguishing them from shallow neural networks that typically contain only one or two hidden layers. The increased depth of DNN architectures allows for the hierarchical learning of abstract and high-level features directly from raw input data. This capability enables DNNs to model complex relationships and subtle variations in network behavior, which is particularly advantageous for detecting sophisticated and previously unseen intrusion patterns.

4.2. Supervised Learning Methods

Supervised machine learning techniques rely on labeled data to learn discriminative patterns that enable the classification of previously unseen instances. Numerous supervised approaches have been explored in the literature for intrusion detection in IoT environments, including Artificial Neural Networks (ANN) [24, 25], Support Vector Machines (SVM) [15, 18, 26], Linear Regression (LR) [18], Random Forest (RF) [13], Autoencoders (AE) [27], and Naive Bayes (NB) classifiers [14]. Among these, ANN-based methods have received considerable attention due to their ability to model complex, non-linear relationships within network traffic data.

ANNs are computational frameworks inspired by the structure and functionality of biological neural systems, composed of interconnected neurons arranged in layered architectures. Each neuron processes input signals through weighted connections and activation functions before forwarding the output to subsequent layers. Despite their strong representational power, ANN-based approaches suffer from several limitations, including high architectural complexity, long training times, and the need for large volumes of labeled data. These challenges can limit their scalability and real-time applicability in resource-constrained IoT environments. Recent studies have nevertheless employed ANN models as classifiers for intrusion detection in IoT networks. For instance, the work presented in [25] proposes an ANN-based intrusion detection system (ANNIDS) designed to identify two routing attacks in RPL-based IoT networks, namely the DODAG Information Solicitation (DIS) attack and the Version attack. The proposed ANNIDS architecture is based on

a five-layer multilayer perceptron (MLP), comprising three hidden layers. The model was trained using 80% of the available traffic data, corresponding to 59,104 packets that included both attack and benign instances, while the remaining 20% was reserved for testing. Experimental results demonstrated high true positive rates, precision, recall, and F1-score values. However, the study lacks critical methodological details, particularly regarding the dataset origin and the feature selection process. The absence of this information raises questions about the reproducibility of the results and the generalizability of the proposed model. Support Vector Machine–based approaches have also been widely investigated for IoT intrusion detection due to their robustness in high-dimensional feature spaces. In [28], the authors address denial-of-service (DoS) attacks that disrupt network traffic intensity in IoT systems. The proposed SVM-based model relies exclusively on packet arrival rate statistics at individual nodes. A filter-based feature selection technique was applied, resulting in the selection of three features: the minimum, maximum, and median packet arrival rates. Traffic exceeding a predefined threshold was classified as malicious. Due to the absence of suitable real-world datasets containing packet arrival rate attributes, the authors generated a synthetic dataset using MATLAB 2018b, modeling traffic behavior based on a Poisson distribution. The proposed approach was evaluated against existing models, including GA-SVM [29], A-IDS [30], and WFS-IDS [31], and achieved an accuracy of 98.35%, a false positive rate below 2%, and a total training and testing time of approximately 16.36 seconds. While these results indicate strong performance for DoS detection, the model's effectiveness against other attack types remains uncertain. A notable limitation of this approach is its inability to detect attacks that do not significantly affect traffic intensity, thereby restricting its applicability to a narrow class of threats.

In another study, the authors in [18] investigated intrusion detection mechanisms for smart irrigation systems in precision agriculture environments. Three supervised learning models—Linear Regression, Support Vector Machines, and Random Forest—were developed and compared to assess their efficiency and robustness against IoT-related attacks that could compromise irrigation accuracy. The experiments were conducted using the NSL-KDD dataset, with Principal Component Analysis (PCA) employed for feature selection. A total of 20% of the dataset, corresponding to 25,192 records, was used for training, while the remaining 80% (100,781 records) was allocated for testing. Experimental results demonstrated that the SVM-based model significantly outperformed the other classifiers, achieving an accuracy of 98%, whereas the LR and RF models achieved accuracy values below 78%. However, the evaluation was limited to a single dataset, which raises concerns regarding the adaptability and robustness of the proposed solutions when applied to different datasets or real-world IoT scenarios. Overall, these studies highlight the effectiveness of supervised learning techniques, particularly ANN and SVM, in IoT intrusion detection. Nevertheless, limitations related to dataset diversity, feature transparency, and attack generalization remain open challenges that warrant further investigation.

Table 1. Summary of ML methods.

Category	**Year**	**Reference**	**Approach**	**Attacks**	**Dataset**	**Accuracy**	**FS**	**FP**
Supervised	2019	Anitha et al. [24]	ANN	DIS attack, Version attack	–	100%	No info	0%
Supervised	2019	Hanif et al. [25]	ANN	Encompassing fuzzers, shellcode, worms	UNSW-NB15	84%	–	8%
Supervised	2019	Jan et al. [28]	SVM	DoS attack	Custom dataset	98.35%	Filter	<2%
Supervised	2022	Raghuvanshi et al. [18]	SVM, LR, RF	Irrigation system attack	NSL-KDD	98% (SVM) / 78%	PCA	–
Supervised	2021	Keserwani et al. [13]	GWO, PSO, RF	DoS, Probe, R2L, U2R	KDDCup99, NSL-KDD, CICIDS-2017	99.66%	GWO-PSO	–
Unsupervised	2017	Al-Yaseen et al. [32]	SVM, K-Means, ELM	DoS, U2R, R2L	KDD Cup 1999	95.17%	K-Means	1.87%

Unsupervised	2018	Nõmm et al. [26]	IF, LOF, SVM	DoS	BASHLITE, Mirai	>97%	LCC	–
Unsupervised	2021	Odu-muyiwa et al. [27]	AE, RBM, K-Means, EM	SYN-Flood, UDP-Lag	CICDDoS2019, Mirai, BASHLITE	99%, 80%, 50%	Min–Max	–
Unsupervised	2022	Shitharth et al. [14]	ADC, DBSCANPPGO, LNB	DoS, Probe, R2L, U2R	NSL-KDD, CICIDS-2017, Bot-IoT	93.89%	PPGO	4.7%
Unsupervised	2021	Liu et al. [15]	OCSVM	Encompassing fuzzers, shellcode, worms	UNSW-NB15	86.68%	PSO, LightGBM	10.62 %

4.3. Unsupervised Learning Methods

Unsupervised learning enables intrusion detection without relying on labeled data, making it suitable for identifying previously unseen attacks. In [27], several unsupervised models were evaluated for detecting transport-layer DDoS attacks, specifically SYN-Flood and UDP-Lag attacks. The authors compared Autoencoders (AE), Restricted Boltzmann Machines (RBM), K-Means, and Expectation–Maximization (EM) using the CICDDoS2019, Mirai, and BASHLITE datasets, with Min–Max normalization applied during preprocessing. Among the evaluated methods, AE achieved the highest performance, exceeding 99% accuracy on the Mirai and BASHLITE datasets, while the remaining techniques achieved accuracy levels between 50% and 80%. Despite these promising results, the study lacks clarity regarding training data proportions, false positive rates, and the suitability of Min–Max normalization across all evaluated algorithms.

In another study, the authors of [32] proposed a multi-level hybrid intrusion detection framework combining Support Vector Machines (SVM) and Extreme Learning Machines (ELM), supported by a modified K-Means clustering technique. The customized clustering approach reduced dataset size by generating smaller representative subsets, which were then used to train the SVM–ELM classifiers. Experiments conducted on the KDD Cup 1999 dataset demonstrated reduced training time, lower false positive rates, and improved detection of both known and unknown attacks. However, the evaluation was limited to a single, outdated dataset, raising concerns regarding the model's robustness and applicability to modern IoT environments.

4.4. Deep Learning Methods

Deep learning [34] is a subfield of ML that uses DNNs to model complex relationships between input data and output predictions [35]. Various DL methods, such as DNN [36] and CNN [16], enable efficient handling of diverse intrusion types and enhance detection accuracy. A compelling approach is introduced in [37]. The method employs two modules: Spider Monkey Optimization (SMO) for feature selection and Stacked Deep Polynomial Networks (SDPN) for attack classification. Trained on the NSL-KDD dataset, the framework outperforms DT, GBT, and KNN with 99% accuracy, precision, recall, and F-score. However, lacking validation on other datasets, and training on 85% of the data, its real-world reliability is questionable. In addition, [38] employs a feedforward neural network (FNN) for binary and multiclass classification, focusing on DDoS, DoS, Reconnaissance, and Information Theft attacks. Detection time is approximately 5 minutes for a single attack type and 20 minutes for multiple attacks. Trained on the Bot IoT dataset, the system achieves 98% accuracy across various attacks and 88% accuracy for data exfiltration due to its resemblance to reconnaissance.

Table 2. Summary of Deep Learning Methods.

Year	Reference	Approach	Attacks	Dataset	Accuracy	FS	FP
2022	Otoum et al. [37]	S-DPN	DoS, DDoS, R2L, Probe	NSL-KDD	99.02%	SMO	–
2019	Ge et al. [38]	FNN	DoS, DDoS, Reconnaissance, Information theft	Bot-IoT	98%	–	<2%
2019	Thamilrasu et al. [36]	DBN, DNN	Sinkhole, DDoS, blackhole, opportunistic service, wormhole	Custom dataset	99%	–	–
2022	Siddharthan et al. [39]	SVM, GB, KNN, NB, RF	SYN Flood	SENMQTT-SET	99%	LR	1.3%
2022	Sayed et al. [16]	MyCNN, IoTCNN	Backdoor, DoS, fuzzers, reconnaissance, shellcode, worms, exploits	NF-UNSW-NB15-v2	99%	–	<2%
2023	Chen et al. [17]	DT, WE	Attacks in (HeIoT) networks	KDD Cup 99, NSL-KDD, UNSW-NB15	>98.5%	–	–

5. DISCUSSION

An analysis of datasets used in studies published between 2017 and 2023 shows a strong preference for the NSL-KDD dataset due to its improved feature representation and better class balance compared to the older KDD Cup 99 dataset. This trend suggests that many proposed methods primarily focus on detecting DoS-related attacks in IoT environments. Although some studies rely on real-world traffic data, KDD Cup 99 continues to be used despite its inability to reflect emerging and evolving attack behaviors. Most datasets emphasize DoS, R2L, and U2R attacks, while largely neglecting modern threats such as Common Vulnerabilities and Exposures (CVE), zero-day exploits, and hardware-level attacks. Across the reviewed literature, widely adopted algorithms include SVM, RF, ANN, DNN, and CNN, typically achieving accuracy and precision in the range of 95%–98%. In particular, SVM-, ANN-, and CNN-based approaches demonstrate high classification performance, often exceeding 98%. However, each technique presents inherent limitations: ANN models struggle with rare attack detection and scalability, SVMs incur high computational and memory costs for large datasets, and CNNs demand substantial labeled data and computational resources to achieve optimal performance.

6. CONCLUSION AND PERSPECTIVES

This paper presents a comprehensive review of recent advancements in machine learning–based intrusion detection systems for IoT environments between 2017 and 2023. A structured taxonomy is introduced, categorizing existing approaches according to learning algorithms, feature selection techniques, datasets, and evaluation metrics. Although notable progress has been achieved, several challenges persist, particularly in detecting sophisticated attacks that closely resemble normal traffic patterns. In addition, reducing false alarm rates and computational overhead remains a critical requirement for practical deployment. These findings indicate that further research is necessary to enhance the robustness and efficiency of ML-based IoT intrusion detection systems. As future work, we aim to update the NSL-KDD dataset by incorporating contemporary attack scenarios, enabling more realistic training and evaluation for industrial IoT applications. Furthermore, improvements to IDS architectures will be explored to better address scalability, accuracy, and real-time performance requirements.

REFERENCES

[1] “Global IoT and non-IoT connections 2010-2025: https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/

[2] “Europe: internet of things market revenue 2018-2028: https://www.statista.com/forecasts/1283723/revenue-from-internet-of-things-in-europe

[3] Y. Perwej, al., “The Internet of Things (IoT) and its Application Domains,” IJCA, Apr. 2019,

[4] E. Ronen, al., “IoT Goes Nuclear: Creating a ZigBee Chain Reaction,” IEEE Symposium on Security and Privacy (SP), May 2017

[5] B. Chah, al., “H3PC: Enhanced Security and Privacy-Preserving Platoon Construction Based on Fully Homomorphic Encryption,”IEEE 26th International Conference on Intelli-gent Transportation Systems (ITSC), Sep. 2023,

[6] T. Shah, al., “Authentication of IoT Device and IoT Server Using Secure Vaults,” in 2018 IEEE International Conference. 2018,

[7] K. M. Sadique, al., “Towards Security on Internet of Things: Applications and Challenges in Technology,” Procedia Computer Science, 2018,

[8] H.-J. Liao, C.-H. Richard Lin, Y.-C. Lin, and K.-Y. Tung, “Intrusion detection system: A comprehensive review,” Journal of Network and Computer Applications, Jan. 2013,

[9] O. Chaieb, al., “Machine Learning-Based Intrusion Detection System: Review and Taxono-my,” in Proceedings International Conference on Big Data and Internet of Things, 2023

[10] M. Masdari, al.,“A survey and taxonomy of the fuzzy signature-based Intrusion Detection Systems,” Applied Soft Computing, Jul. 2020

[11] A. Patcha, al., “An overview of anomaly detection techniques: Existing solutions and latest technological trends,” Computer Networks. 2007

[12] C. Alex, al., “A Comprehensive Survey for IoT Security Datasets Taxonomy, Classification and Machine Learning Mechanisms,” Computers & Security, Sep. 2023,

[13] P. K. Keserwani, al., “A smart anomaly-based intrusion detection system for the Internet of Things (IoT) network using GWO–PSO–RF model,” J Reliable Intell Environ, Mar. 2021

[14] S. Shitharth, al., “An Innovative Perceptual Pigeon Galvanized Optimization (PPGO) Based Likelihood Naïve Bayes (LNB) Classification Approach for Network Intrusion Detection System,” IEEE Access, 2022

[15] J. Liu, al., “Research on Intrusion Detection Based on Particle Swarm Optimization in IoT,” IEEE Access, 2021

[16] N. Sayed, Augmenting IoT Intrusion Detection System Performance Using Deep Neural Network. Computers, Materials and Continua, 2022

[17] D. Chen, al., “Heterogeneous IoT Intrusion Detection Based on Fusion Word Embedding Deep Transfer Learning,” IEEE Transactions on Industrial Informatics, Aug. 2023

[18] A. Raghuvanshi, al., “Intrusion Detection Using Machine Learning for Risk Mitigation in IoT-Enabled Smart Irrigation in Smart Farming,” Journal of Food Quality, Feb. 2022

[19] A. K. Sahu, al., “Internet of Things attack detection using hybrid Deep Learning Model,” Computer Communications, Aug. 2021

[20] I. Ullah, al., “Network Traffic Flow Based Machine Learning Technique for IoT Device Identification,” IEEE International Systems Conference (SysCon), Apr. 2021

[21] M. Zolanvari, al., “Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things,” IEEE Internet of Things Journal,.2019

[21] M. Zolanvari, al., “Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things,” IEEE Internet of Things Journal,.2019

[22] S. M. Kasongo, al., “A deep learning method with wrapper based feature extraction for wire-less intrusion detection system,” Computers & Security2020

[23] F. Chen, al., “A Feature Selection Approach for Network Intrusion Detection Based on Tree-Seed Algorithm and K-Nearest Neighbor,” IDAACS-SWS, Sep. 2018

[24] A. A. Anitha, al. Annids: artificial neural network based intrusion detection system for in-ternet of things,” Int. J. Innov. Technol. Exp, 2019

[25] S. Hanif, al., “Intrusion Detection in IoT Using Artificial Neural Networks on UNSW-15 Dataset,” presented at the HONET-ICT, 2019,

[26] S. Nõmm, al., “Unsupervised Anomaly Based Botnet Detection in IoT Networks,”17th IEEE International Conference on Machine Learning and Applications (ICMLA), Dec. 2018

[27] V. Odumuyiwa , al., “DDOS Detection on Internet of Things Using Unsupervised Algorithms, Journal of Cyber Security and Mobility, 2021

[28] S. U. Jan, al., “Toward a Lightweight Intrusion Detection System for the Internet of Things,” IEEE Access, 2019

[29] P. Tao, al., “An Improved Intrusion Detection Algorithm Based on GA and SVM,” IEEE Access, 2018

[30] S. Aljawarneh, al., “Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model,” Journal of Computational Science, Mar. 2018

[31] Y. Li, al., “Building lightweight intrusion detection system using wrapper-based feature se-lection mechanisms,Computers & Security. 2009

[32] W. L. Al-Yaseen, al., “Multi-level hybrid support vector machine and extreme learning ma-chine based on modified K-means for intrusion detection system,” Expert Systems with Ap-plications, Jan. 2017

[33] M. Tavallaee, al., “A detailed analysis of the KDD CUP 99 data set,” IEEE SCISDA, 2009

[34] Y. LeCun, al., “Deep learning,” Nature, May 2015

[35] W. Samek, al., “Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications,” Proceedings of the IEEE,. 2021

[36] G. Thamilarasu, al., “Towards Deep-Learning-Driven Intrusion Detection for the Internet of Things,” Sensors, Jan. 2019

[37] Y. Otoum, al.,“DL-IDS: a deep learning–based intrusion detection framework for securing IoT,” Transactions on Emerging Telecommunications Technologies, 2022

[38] M. Ge, X. Fu, N. Syed, Z. Baig, G. Teo, and A. Robles-Kelly, “Deep Learning-Based Intru-sion Detection for IoT Networks,” IEEE. 2019

[39] H. Siddharthan, “SENMQTT-SET: An Intelligent Intrusion Detection in IoT-MQTT Networks Using Ensemble Multi Cascade Features,” IEEE, 2022