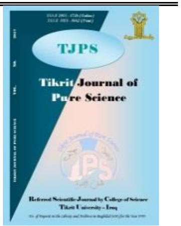
https://doi.org/10.25130/tjps.v29i3.1553

IASJ
المجلات الاكاديمية العراقية

Tikrit Journal of Pure Science

ISSN: 1813 – 1662 (Print) --- E-ISSN: 2415 – 1726 (Online)

Journal Homepage: http://tjps.tu.edu.iq/index.php/j



Using Machine Learning Algorithms in Intrusion Detection Systems: A Review

Mazin S. Mohammed [1] **, Hasanien Ali Talib** [2]

[1] *Department of Graduate Studies, University of Mosul, Mosul, Iraq*

[2] *Department of Computer, College of Education for Pure Sciences, University of Mosul, Mosul, Iraq*

ARTICLE INFO.

Article history:

-Received:	**12 / 9 / 2023**
-Received in revised form:	**13 / 12 / 20 23**
-Accepted:	**18 / 12 / 20 23**
-Final Proofreading:	**22 / 3 / 2024**
-Available online:	**25 / 6 / 2024**

Keywords: Supervised learning, IoT intrusion detection systems, Machine learning algorithms, Feature selection, Future research directions.

Corresponding Author:

Name: **Mazin S. Mohammed**

E-mail: **mazinsalm@uomosul.edu.iq**

Tel:

©2024 THIS IS AN OPEN ACCESS ARTICLE UNDER THE CC BY LICENSE http://creativecommons.org/licenses/by/4.0/

cc BY

ABSTRACT

Intrusion Detection Systems (IDS) are essential for identifying and mitigating security threats in Internet of Things (IoT) networks. This paper explores the unique challenges of IoT environments and presents machine learning (ML) algorithms as powerful solutions for IoT-IDS, encompassing supervised, unsupervised, and semi-supervised learning. Notable algorithms, including decision trees, random forests, support vector machines, and deep learning architectures, are discussed. Emphasis is placed on the critical role of feature selection in developing efficient IDS, addressing challenges such as heterogeneity, limited resources, real-time detection, privacy concerns, and adversarial attacks. Future research directions include advanced ML algorithms for IoT data, integration of anomaly detection, exploration of federated learning, and combining ML with other cybersecurity techniques. The paper advocates for benchmark datasets and evaluation frameworks to standardize the assessment of ML-based IoT-IDS approaches, ultimately contributing to heightened security and integrity in IoT systems..

استخدام خوارزميات التعلم الآلي في أنظمة كشف التسلل: مراجعة

مازن سالم محمد [1]، حسنين علي طالب

[1]*قسم الدراسات العليا ، جامعة الموصل ، الموصل ، العراق*

[2]*قسم الحاسوب ، كلية التربية للعلوم الصرفة ، جامعة الموصل ، الموصل ، العراق*

الملخص

تعتبر أنظمة كشف التسلل (IDS) ضرورية لتحديد وتخفيف التهديدات الأمنية في شبكات إنترنت الأشياء (IoT). تستكشف هذه الورقة التحديات الفريدة لبيئات إنترنت الأشياء وتقدم خوارزميات التعلم الآلي (ML) كحلول قوية لـ IoT-IDS، والتي تشمل التعلم الخاضع للإشراف وغير الخاضع للإشراف وشبه الخاضع للإشراف. تمت مناقشة الخوارزميات البارزة، بما في ذلك أشجار القرار والغابات العشوائية وآلات ناقل الدعم وهياكل التعلم العميق. يتم التركيز على الدور الحاسم لاختيار الميزات في تطوير IDS فعال، ومعالجة التحديات مثل عدم التجانس، والموارد المحدودة، والكشف في الوقت الحقيقي، ومخاوف الخصوصية، والهجمات العدائية. تتضمن اتجاهات البحث المستقبلية خوارزميات التعلم الآلي المتقدمة لبيانات إنترنت الأشياء، وتكامل الكشف عن الحالات الشاذة، واستكشاف التعلم الموحد، والجمع بين التعلم الآلي وتقنيات الأمن السيبراني الأخرى. في النهاية تدعو

https://doi.org/10.25130/tjps.v29i3.1553

هذه الورقة إلى إنشاء مجموعات بيانات مرجعية وأطر تقييم لتوحيد تقييم مناهج IoT-IDS القائمة على التعلم الآلي، مما يساهم في نهاية المطاف في زيادة الأمن والنزاهة في أنظمة إنترنت الأشياء.

1. Introduction

1.1 Background

The Internet of Things (IoT) has revolutionized how physical devices communicate and interact, forming interconnected systems that enable seamless information exchange [1, 2]. With the increasing prevalence of IoT devices in various domains such as healthcare, transportation, manufacturing, and intelligent cities, robust security measures are paramount. These devices, from medical and healthcare devices to driverless vehicles, industrial robots, smart T.V.s, wearables, and smart city infrastructures, often handle sensitive information, including personal data [2-4]. As IoT devices proliferate, the attack surface area expands, increasing the likelihood of cyber-attacks. Safeguarding the communication and data exchange facilitated by IoT technologies necessitates the development of effective IoT intrusion de007Atection systems (IDS) [5, 6]. Ensuring the security of IoT applications has become a critical aspect of their implementation. In recent years, advancements in Artificial Intelligence (A.I.), particularly machine learning and deep learning techniques, have been leveraged to enhance IoT IDS. Various studies have explored applying these techniques using diverse datasets to validate the development of IoT IDS [7-9]. However, there remains a lack of clarity regarding which datasets and A.I. techniques are most effective for building efficient IoT IDS.

Additionally, evaluating some IDS techniques often overlooks the time consumed in the building and testing phases, despite its critical role in the effectiveness of "online" IDSs. This research paper aims to provide an up-to-date taxonomy and critical review of recent work in IoT IDS. It offers a comprehensive overview of existing IoT IDSs, classifying them based on the proposed taxonomy. By examining the key aspects of IoT IDS, this paper facilitates a quick understanding of the field for researchers. Furthermore, it critically reviews machine learning and deep learning techniques employed in building IoT IDS, exploring detection methods, validation strategies, deployment approaches, and evaluation techniques. The paper delves into the complexity of different detection techniques, intrusion deployment strategies, and their evaluation, providing valuable insights and suggesting the best techniques based on the nature of the IoT IDS. Additionally, the challenges faced by current IoT IDSs are discussed, shedding light on areas that require further attention and improvement.

1.2 Problem Statement

Integrating ML techniques into IoT-IDS (IDS) presents a promising approach to enhancing the security and resilience of IoT networks. This section highlights the critical problem statements associated with the application of ML in IoT-IDS between 2018 and 2023.

1. Limited labeled datasets: Developing accurate and robust ML models for IoT-IDS requires large-scale, labeled datasets that capture the diversity of IoT network traffic and attack scenarios.

2. F.S. for IoT-IDS: IoT networks generate vast amounts of data from various sources, including sensors, actuators, and communication protocols. Selecting relevant features from this high-dimensional data is crucial to improve the efficiency and effectiveness of ML models in IoT-IDS.

3. Adaptability to dynamic IoT environments: IoT networks are highly dynamic, with devices joining and leaving the network, changing their behaviors, and encountering new attack patterns. ML algorithms used in IoT-IDS must be able to adapt to these dynamic environments and continuously update their models to detect emerging threats. Ensuring real-time adaptability and scalability while maintaining high detection accuracy is a complex problem.

4. Scalability and computational constraints: IoT environments consist of many interconnected devices with limited computational capabilities. Deploying resource-intensive ML algorithms on resource-constrained IoT devices may result in performance degradation and energy inefficiency. Developing lightweight and energy-efficient ML models that can operate within the constraints of IoT devices is a critical challenge.

5. Interpretability and explainability: ML models used in IoT-IDS often exhibit complex decision-making processes, making it challenging to interpret and explain the reasoning behind their predictions.

1.3 Objectives

This study survey's main goal is to present a thorough review of the studies on the use of machine learning in IoT-IDS (IDS) that were carried out between 2018 and 2023. The following are the specific goals of this survey:

1. To identify and examine the cutting-edge machine learning methods used in IoT-IDS: The purpose of this survey is to examine the various machine learning techniques and algorithms applied to IoT-IDS throughout the given time period. It will examine the benefits, drawbacks, and suitability of various methods in relation to Internet of Things networks.

2. To investigate into F.S. approaches for IoT-IDS: F.S. is important since it helps make ML models in IoT-IDS more effective and efficient. This review examines the F.S. techniques used in the literature within the given time period and assesses how well they choose pertinent characteristics for IoT-IDS.

3. To examine the assessment techniques applied to ML-based IoT-IDS: In order to assess the effectiveness of ML models in IoT-IDS, the right metrics, datasets, and assessment techniques are needed. This review attempts to examine the assessment techniques used in the literature throughout the designated period of time and determine whether or not they are appropriate for evaluating the effectiveness of ML-based IoT-IDS.

4. To give a thorough rundown of the advantages and disadvantages of the current methods: This review attempts to determine the benefits and drawbacks of the ML-based techniques applied in IoT-IDS by looking at the relevant research works in the field. It will emphasize the main accomplishments, creative concepts, and difficulties of each technique.

5. To identify potential future research directions: Based on the analysis of existing approaches, this survey aims to identify the critical research gaps and suggest potential future directions for advancing ML in IoT-IDS. It will outline the research areas requiring further exploration and propose innovative ideas to address the challenges of applying ML in IoT-IDS.

2. Intrusion Detection System in Internet of Things

In the Internet of Things (IoT) context, IDS aims to detect and mitigate security threats and attacks within IoT networks. Traditional IDS, designed for traditional computer networks, are not directly applicable to IoT environments' unique characteristics and challenges [10]. IoT-IDS require specialized approaches to handle the large-scale deployment, heterogeneity, resource constraints, and dynamic nature of IoT networks [11]. IoT-IDS typically involve monitoring and analyzing network traffic, device behavior, and communication patterns to identify potential security breaches [12]. They rely on various techniques, including rule-based systems, anomaly detection, and ML, to detect and respond to security incidents [13]. See Fig 1.

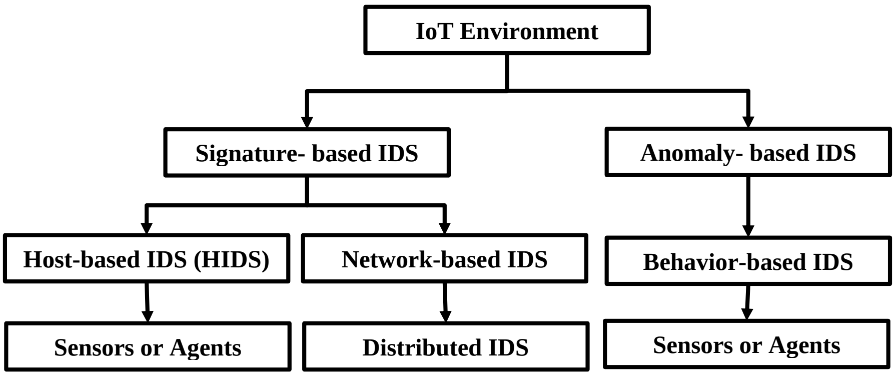


Fig 1: Classification of intrusion detection systems in the Internet of Things (IoT-IDS)

ML algorithms used in IoT-IDS can be categorized into supervised, unsupervised, and semi-supervised learning. Supervised learning algorithms leverage labeled data to learn normal and malicious behavior patterns, while unsupervised learning algorithms identify anomalies based on data distribution. Semi-supervised learning algorithms combine labeled and unlabeled data to enhance detection capabilities [14]. Deep learning techniques, such as deep neural networks and recurrent neural networks, have also shown promise in detecting complex and sophisticated attacks in IoT networks. Implementing effective IDS in IoT networks has several challenges [2, 15]. The following challenges are particularly relevant in the context of applying ML techniques in IoT-IDS:

1. Heterogeneity and scalability: IoT networks comprise various devices, communication protocols, and data formats. Developing ML models that can handle the heterogeneity of IoT data and scale to large-scale IoT deployments is a challenge. The models should be adaptable to various devices, communication technologies, and network architectures [16-18].

2. Limited computational resources: IoT devices often have limited computational power, memory, and energy resources. Designing lightweight ML algorithms that can operate efficiently on resource-constrained IoT devices is crucial. These algorithms should balance detection accuracy and computational overhead to ensure practical implementation in IoT environments [17, 19].

3. Real-time detection and response: IoT networks operate in real time, and timely detection and response to security incidents are critical. ML-based IDS should be capable of processing and analyzing data in real-time to detect and respond to attacks on time. Real-time detection requires efficient algorithms and optimized computational processes to handle IoT data streams' high volume and velocity [20].

4. Privacy and data protection: IoT devices collect and transmit sensitive data, making privacy and data protection essential considerations in IoT-IDS. ML algorithms should be designed to respect privacy

https://doi.org/10.25130/tjps.v29i3.1553

requirements and ensure secure data handling. Additionally, models should be robust against attacks targeting privacy-sensitive information[21, 22].

5. Adversarial attacks and model robustness: IoT networks are vulnerable to adversarial attacks, where malicious actors intentionally manipulate or evade detection mechanisms. ML models used in IoT-IDS should be robust against adversarial attacks and resilient to adversarial perturbations. Developing techniques to enhance the robustness of ML models in the presence of adversarial threats is an ongoing challenge [23].

3. ML Techniques for IoT-IDS

To guarantee the security and integrity of IoT systems, intrusion detection in networks is essential. In Internet of Things (IoT) contexts, machine learning (ML) approaches have become effective instruments for identifying and reducing intrusions [24, 25]. See Fig. 2.

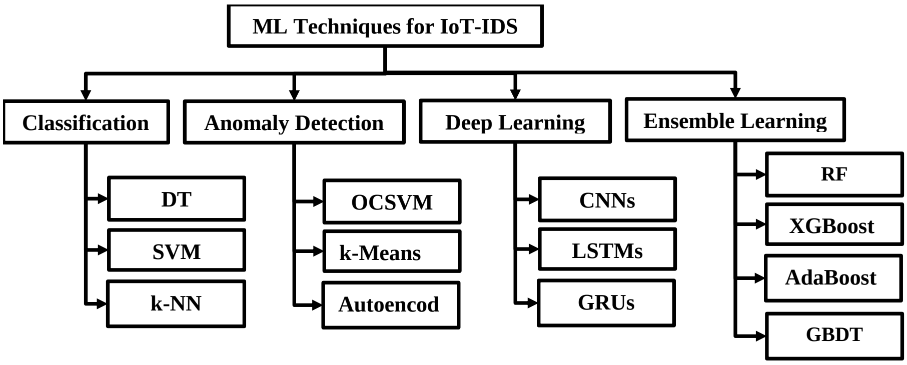


Fig 2. Taxonomy of ML Techniques for IoT-IDS

The several ML approaches that have been used to IoT intrusion detection are covered in this section.

3.1 Classification Algorithms

Intrusion detection systems (IDS) frequently use classification algorithms to categorize network traffic into distinct classifications, such as malicious or benign. These algorithms categorize unseen instances by using patterns they have learned from labeled training data. Many classification algorithms have been applied in the context of IoT-IDS, such as decision trees (DT) [25, 26, 27, 28, 29, 30], random forests (RF) [25, 7, 26, 27, 31, 8, 24], support vector machines (SVM) [25, 27, 31, 29, 24, 32], k-nearest neighbors (KNN) [27, 33, 30,32], and XGBoost [26,28, 24]. Decision trees are straightforward yet powerful algorithms that generate a feature-based decision tree model. Multiple decision trees are combined to create random forests, which are well-known for their strong classification performance and capacity to handle high-dimensional data. The SVM algorithm is a binary classification technique that determines the best hyperplane to divide several groups. KNN classifies instances based on their proximity to labeled instances in the feature space. XGBoost is an ensemble learning algorithm that combines weak classifiers to form a strong classifier.

3.2 Anomaly Detection Algorithms

Anomaly detection algorithms are particularly useful for identifying unknown and novel attacks in IoT-IDS. These algorithms learn the normal behavior of the system and flag instances that deviate significantly from the learned patterns as anomalies. Popular anomaly detection algorithms used in IoT-IDS include one-class support vector machines (OCSVM), k-means clustering, and autoencoders. OCSVM is a variant of SVM that learns a boundary around the normal instances in the feature space [25, 33]. It can then detect deviations from this boundary as anomalies. K-means clustering partitions the data into k clusters, where instances that do not belong to any cluster are considered anomalies. Autoencoders are deep learning models that aim to reconstruct the input data from a compressed representation. Instances that have a high reconstruction error are identified as anomalies.

3.3 Deep Learning Techniques

Deep learning techniques, specifically neural networks, have gained significant attention in IoT-IDS due to their ability to learn complex patterns from high-dimensional data. Convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and gated recurrent units (GRUs) are commonly used deep learning architectures for IoT-IDS [34,35, 36]. CNNs are effective in capturing spatial dependencies in data, making them suitable for analyzing IoT network traffic. LSTM and GRU networks are recurrent neural networks (RNNs) that excel at capturing temporal dependencies [3, 25, 27]. They can effectively model sequential data, making them well-suited for analyzing time-series data in IoT-IDS.

3.4 Ensemble Learning Approaches

Ensemble learning combines multiple models to improve the overall performance and robustness of the intrusion detection system. Bagging and boosting are two popular ensemble learning approaches used in IoT-IDS [24]. Bagging, short for bootstrap aggregating, involves training multiple models on

https://doi.org/10.25130/tjps.v29i3.1553

different subsets of the training data and combining their predictions through voting or averaging. This helps reduce the impact of individual model biases and improves the overall accuracy. Random forests and XGBoost are examples of bagging-based ensemble methods. Boosting, on the other hand, focuses on iteratively training weak models and giving more weight to misclassified instances [26, 28]. This allows the ensemble to emphasize the difficult instances and improve the overall classification performance. AdaBoost and gradient boosting are well-known boosting algorithms used in IoT-IDS.4. Feature Selection in IoT-IDS

4. Feature Selection in IoT-IDS

4.1 Importance of Feature Selection

Feature selection is a critical step in developing an effective Intrusion Detection System (IDS) for Internet of Things (IoT) networks. With the rapid growth of IoT and the increasing number of connected devices, the amount of data generated has also increased significantly [37]. However, not all features or attributes of the data contribute equally to the detection of intrusions or anomalies. In fact, including irrelevant or redundant features can introduce noise and negatively impact the performance of the IDS. By determining the most pertinent and instructive features for intrusion detection, feature selection is essential to enhancing the efficacy and efficiency of IoT-IDS [7]. The intrusion detection system (IDS) can concentrate on critical information and minimize the computational complexity involved in handling vast volumes of data by choosing a subset of characteristics with strong discriminating power. As a result, the IDS performs better overall and detects threats more quickly and with fewer false positives [38]. Additionally, feature selection aids in resolving IoT devices' resource limitations. These gadgets frequently feature constrained memory, processor, and energy capacities. The computational load on IoT devices may be minimized by choosing a smaller number of characteristics, allowing them to effectively carry out intrusion detection duties within their limited resources.

4.2 FS Techniques in IoT-IDS

Supervised learning methods that may be applied to feature selection (FS) in Internet of Things intrusion detection systems (IoT-IDS) based on the data presented in Table 1:

1. Manual Feature Selection:

Used methods for manual feature selection in their studies. Manual feature selection is the process of choosing pertinent features using subject skills and knowledge. With this method, researchers may concentrate on particular characteristics that are most likely to aid in the intrusion detection process [25,35].

2. Hybrid Feature Selection:

used Information Measure of Feature (IMF) and Uncertainty Measure of Feature (UMF) hybrid feature selection approaches. Several feature selection techniques are used in hybrid feature selection to take advantage of each one's unique advantages. It can help improve detection performance and offer a more thorough examination of feature relevance [7].

3. Network Profiling

Employed a feature selection method based on the correlation coefficient. The linear link between traits is measured by the correlation coefficient, which shows how dependent they are on one another. The algorithm can concentrate on the most pertinent data for precise identification by choosing characteristics that have a strong correlation to the goal variable (intrusion or normal behavior) [31,24].

4. Correlation Coefficient

employed a feature selection method based on the correlation coefficient. The linear link between traits is measured by the correlation coefficient, which shows how dependent they are on one another. The algorithm can concentrate on the most pertinent data for precise identification by choosing characteristics that have a strong correlation to the goal variable (intrusion or normal behavior) [31,24].

5. Supervised Learning Algorithms in IoT-IDS

5.1 Survey of Research Works

The domain of IoT Intrusion Detection Systems (IoT-IDS) has witnessed a notable surge in interest in supervised learning methods. These algorithms use labeled data to train models that are capable of precisely identifying and classifying intrusions in Internet of Things networks. We provide a review of studies that have investigated the use of supervised learning techniques in IoT-IDS in this section.

The application of deep learning algorithms for intrusion detection in Internet of Things networks was the subject of one research by Banaamah and Ahmad [35]. Using a typical dataset for intrusion detection in the Internet of Things, they evaluated the effectiveness of many deep learning models, including convolutional neural networks (CNNs), long short-term memory (LSTM), and gated recurrent units (GRUs). When compared to current methods, their suggested method showed better accuracy.

A comparative research of machine learning methods for IoT network intrusion detection was carried out by Marwa Baich et al. [31]. They examined how different machine learning techniques, such as decision trees, performed when applied to a dataset that had both binary and multi-class categorization. According to the study, the Fisher score Decision Tree algorithm performed the best, achieving high accuracy and short forecast times.

Another work by Bouazza et al. [29] used an intrusion detection system based on machine learning to identify routing assaults in the Internet of Things. Using machine learning techniques and a dataset of IoT assaults produced via simulations, they created an

https://doi.org/10.25130/tjps.v29i3.1553

anomaly-based intrusion detection system. They suggested a technique that detected routing-based attacks with more accuracy and precision by adding additional sensitive characteristics and balancing the dataset.

Arhore [24] concentrated on machine learning-based intrusion detection in Internet of Things platforms. The study examined a number of machine learning approaches and assessed how well they performed using measures including F1 score, recall, precision, and classification accuracy. The goal of the study was to provide an appropriate algorithm that can identify network intrusions effectively and efficiently, with 99% accuracy and high efficiency.

A machine learning-based distributed intrusion detection solution for Internet of Things networks was proposed by Gad et al. [28]. To train and evaluate several machine learning techniques, they used the ToN-IoT dataset, which represents data from multiple levels of the IoT system. Their suggested model proved the effectiveness of the XGBoost strategy for intrusion detection in IoT networks by using ML algorithms in binary and multi-class classification tasks.

A number of IoT dangers were discovered, and Islam et al. [25] talked about both shallow and deep machine learning-based intrusion detection systems in IoT environments. They used benchmark datasets to assess these models' performance and discovered that deep machine learning performed better at identifying IoT threats than shallow machine learning.

Ayub et al. [30] created an intelligent intrusion detection system for smart city networks using machine learning in a different research. They used a variety of supervised machine learning methods, such as decision trees, XGBoost, k-nearest neighbors (KNN), linear and quadratic discriminant analysis, and XGBoost, and compared the outcomes. The KNN algorithm offered a quick, safe, and clever intrusion detection system (IDS) solution. It also demonstrated the greatest accuracy, followed by XGBoost and decision trees.

For Internet of Things systems, Siham and Kerem [33] suggested a novel detection approach based on deep learning and machine learning techniques. To find abnormalities in IoT networks, they ran trials and contrasted several AI models. Their research demonstrated how ML and DL algorithms may be used to identify different kinds of assaults on Internet of Things platforms.

The issue of large dimensionality in IoT intrusion detection systems was tackled by Albulayhi et al. [7]. They used set theory and entropy-based techniques to present a unique feature extraction and selection strategy. Their strategy led to the selection of a subset of pertinent attributes that successfully gathered the data needed for intrusion detection. Their method increased the intrusion detection system's effectiveness and performance by lowering the dataset's dimensionality.

A well-known paper by Rose et al. [39] suggests an anomaly-based approach to intrusion detection that integrates machine learning and network monitoring methods. All networked IoT devices are dynamically profiled and monitored by the system, which looks for abnormal network transactions and tampering attempts. Any departure from the specified device profile is viewed as an assault and is examined more closely. The authors analyze raw traffic and find possible assaults using a machine learning classifier. Cyber-Trust testbed experimental findings show encouraging results, with a low false-positive rate of 0.98% and an overall accuracy of 98.35%.

To identify cyber threats in IoT networks, Kothari et al. [34] present intelligent intrusion detection systems (IDS) models based on deep learning approaches. They create deep learning algorithms that can identify malware in IoT networks and categorize stolen programs using the TensorFlow framework. To train and assess their models, the authors make use of email datasets and the Google Code Jam dataset. Their method offers an effective way to find harmful assaults in Internet of Things infrastructures.

The problem of creating a multi-class attack detection and classification system for Internet of Things networks is addressed by Othman and Abdullah [32]. They suggest an intelligent intrusion detection system that takes advantage of machine learning techniques' categorization capabilities, including support vector machines, artificial neural networks, and K-Nearest Neighbor. To train and evaluate their models, the authors utilize the IoT23 dataset, which contains millions of examples of both benign and harmful activity from IoT-connected devices. The outcomes show how well the suggested IDS can identify and categorize assaults.

A hybrid intrusion detection system (HID) is suggested by Alghayadh and Debnath [26] for smart home security in Internet of Things environments. Their technology analyzes user activity and finds intrusions by combining machine learning methods such as Xgboost, random forest, decision tree, K-nearest neighbors, and abuse detection tool. For smart homes, the HID system offers improved security and privacy by adjusting to user behavior and surroundings.

The problem of creating lightweight intrusion detection systems for Internet of Things networks is addressed by Ozer et al. [27]. They provide a method that focuses on choosing the best and most effective feature pairs from datasets in order to facilitate the creation of lightweight IDS. The BoT-IoT (2018) dataset and machine learning methods are used by the authors to create and contrast feature-pair-based and full-feature-based intrusion detection systems. Their results demonstrate that feature-pair-based intrusion detection systems (IDS) may achieve excellent detection accuracy.

An improved dynamic SBPSO (Sticky Binary Particle Swarm Optimization) is the foundation of Sarwar et al.'s [8] proposed enhanced anomaly detection system for the Internet of Things. To improve the searchability of SBPSO for feature selection, they add dynamic parameters and a dynamic search space reduction technique. When compared to traditional PSO-based feature selection techniques, the suggested system exhibits better accuracy, lower computing costs, and shorter prediction times. It is tested on two IoT network datasets.

By combining Principal Component Analysis (PCA) and Mayfly Optimization (MAO) for dimensionality reduction, Borderline Synthetic Minority Oversampling Technique (BSMOTE) for data balancing, and Long Short-Term Memory (LSTM) for classification, Karamollaoğlu et al.[3] present a novel IDS for IoT environments. The suggested model outperforms conventional machine learning techniques in identifying assaults with high accuracy (99.51%) in high-dimensional, complicated, and unbalanced data.

These methods have been used with diverse datasets, feature selection strategies, and machine learning algorithms in IoT-IDS research projects. The effectiveness and precision of these methods differ based on the particular application and assessment criteria applied in every research project. When choosing and evaluating the efficacy of these strategies, it is crucial to take the unique needs and features of the IoT system into account.

5.2 Comparison and Analysis of Existing Approaches

Numerous studies have been carried out in the area of intrusion detection systems (IDS) for Internet of Things networks, as listed in Table 1. With the use of various datasets, feature selection strategies, machine learning methods, and reporting accuracy metrics, each research focuses on a distinct component of IDS.

Table 1: Summary of Research Works

Paper	year	Dataset	Feature Selection	ML Algorithm	Acc
N. Islam et al. [25]	2021	NSL-KDD, IoTDevNet, DS2OS, IoTID20, and IoT Botnet	manual	Bi-LSTM, DT, RF, and SVM	Bi-LSTM = 99.04%
K. Albulayhi et al. [7]	2021	IoTID20 and NSL-KDD	hybrid feature selection (IMF, UMF)	RF, MLP, J48, and IBk	RF = 99.98%
J. R. Rose et al. [39]	2021	Cyber-Trust	Network profiling	MobileNetV3	98.35%
T. Kothari et al. [34]	2021	custom dataset	Colour graphics construction from raw binary data	DCNN	CNN+LSTM (97.16)
Alghayadh and Debnath [26]	2021	CSE-CIC-IDS2018 NSL-KDD	full features	RF, Xgboost, DT, K-NN, and misuse detection technique	Xgboost = 98.6%
E. Özer et al. [27]	2021	BoT-IoT (2018)	feature-pair-based	KNN RBF SVM Gaussian Process DT RF ANN AdaBoost NB	RF = 99.9
A. M. Banaamah et al. [35]	2022	Bot-IoT	manual	CNN, GRN and LSTM	GRN = 0.998
M. Baich et al. [31]	2022	NSL-KDD	Pearson correlation Fisher Score	DT, SVM, NB, and RF	DT = 99.26%
A. Bouazza et al. [29]	2022	custom dataset	full features	DT, SVM, NB, and RF	RF = 0.999
A. R. Gad et al. [28]	2022	ToN-IoT	Chi2	RF, XGboost and DT	XGboost = 0.999
Siham and Kerem [33]	2022	UNSW-NB15	random forest	NB, kNN, LR, DT,	RF = 87.09%
Karamollaoğlu et al [3]	2022	IoTID20	PCA-MAO	LSTM	99.51%
Sarwar et al. [8]	2022	IoTID20 and UNSW-NB15	IDSBPSO	RF	99%
S. A. Arhore [24]	2023	IoT NID	correlation coefficient	RF, XGboost and SVM	RF = 99.42%
M. Y. Ayub et al. [30]	2023	UNSW-NB15	full features	XG Boost, KNN and DT	KNN
Othman and Abdullah [32]	2023	IoT23	correlations coefficient	KNN, SVM, and ANN	KNN = 0.99
B. Mansi et al. [36]	2023	IoTID20	correlation	PCC-CNN	99%

https://doi.org/10.25130/tjps.v29i3.1553

An overview of several research studies on intrusion detection systems (IDS) for Internet of Things networks is given in Table 1. The paper's details, publication year, dataset, feature selection methods, machine learning (ML) algorithms, and accuracy claims are all included in the table. NSL-KDD, IoTDevNet, DS2OS, IoTID20, IoT Botnet, Cyber-Trust, CSE-CIC-IDS2018, BoT-IoT, UNSW-NB15, and bespoke datasets are only a few of the datasets covered by the mentioned articles. Each article uses a different feature selection strategy, such as chi-square, correlation coefficient, manual selection, network profiling, hybrid feature selection (IMF, UMF), Pearson correlation, Fisher score, feature-pair-based selection, and manual selection. Bi-LSTM, decision tree (DT), random forest (RF), support vector machine (SVM), multilayer perceptron (MLP), J48, IBk, MobileNetV3, deep convolutional neural network (DCNN), XGBoost, K-nearest neighbors (KNN), Gaussian Process, artificial neural network (ANN), AdaBoost, and Naive Bayes (NB) are just a few of the models that are used in the ML algorithms. The suggested methods' efficacy is shown by the stated accuracy numbers, several of which achieve high accuracy rates. The accuracy statistics show that ML algorithms have been successfully applied for IoT intrusion detection, ranging from 87.09% to 99.98%.

3.5 Datasets for IoT-IDS

In order to create and assess intrusion detection systems (IDS) for Internet of Things networks, scientists use a variety of datasets that are intended to reflect the features and difficulties of IoT settings. We go over a number of frequently used datasets for IoT-IDS in this part, along with the attributes associated with each.

1. KDDCup-99 Dataset (1998):

A well-known dataset that is frequently used in the field of network intrusion detection is KDDCup-99. It offers a thorough collection of network traffic data and is frequently used as a standard for assessing IDS performance, despite not being designed with IoT-IDS in mind. Many attack methods, such as Denial of Service (DoS), Probe, Remote to Local (R2L), and User to Root (U2R), are included in the dataset [40].

2. Kyoto_2006 Dataset (2006):

The Kyoto_2006 dataset, which is based on actual network traffic data gathered at Kyoto University, is primarily concerned with IoT network intrusion detection. It is appropriate for assessing the effectiveness of anomaly detection techniques in IoT-IDS as it encompasses both known and unknown threat types [41].

3. NSL-KDD Dataset (2009):

An improved version of the KDDCup-99 dataset is called NSL-KDD. It improves upon the previous dataset's shortcomings and duplications to more accurately depict contemporary network activity. Numerous network connections classified as regular or with certain attack kinds, such as DoS, Probe, R2L, and U2R, make up NSL-KDD. It is frequently used to assess how well ML algorithms function in IoT-IDS [42].

4. UNSW-NB15 Dataset (2015):

A large-scale dataset created especially for network intrusion detection research is the UNSW-NB15 dataset. It includes data from attack and regular traffic that was produced in an actual Internet of things scenario. Numerous attack types are covered by the dataset, such as Reconnaissance, Shellcode, Analysis, Backdoor, DoS, Exploits, Fuzzers, Generic, Normal, and Worms. It is often used to assess the performance of ML algorithms in IoT-IDS [43].

5. AWID Dataset (2016):

The goal of the AWID dataset is to identify intrusions into wireless networks. It covers a range of attack scenarios, including denial-of-service (DoS), man-in-the-middle, and key cracking. The dataset is used to assess how well IDS performs in wireless IoT networks by capturing the unique issues associated with these networks [44].

6. CIC_IDS2017 Dataset (2017):

A comprehensive collection of network traffic data produced by several attack methods, such as Brute Force, HeartBleed, Botnet, DoS, DDoS, Web, and Infiltration, is included in the CIC_IDS2017 dataset. The dataset's purpose is to assess how well machine learning algorithms identify and categorize Internet of Things network threats [45].

7. CSECIC_IDS2018 Dataset (2018):

Similar to the CIC_IDS2017 dataset, the CSECIC_IDS2018 dataset also includes attack types such Web, HeartBleed, and Infiltration. Its goal is to offer a wide range of attack scenarios so that IDS performance in IoT networks may be assessed [46].

8. LITNET_2020 Dataset (2020):

The LITNET_2020 dataset comprises a wide range of attack types, such as Smurf, ICMP Flood, UDP Flood, SYN flood, HTTP Flood, LAND, W32.Blaster, Code Red, SPAM, Reaper Worm, Scan, and Packet Fragmentation, and is primarily focused on network intrusion detection in Internet of Things settings. In order to assess ML-based IDS, it offers a realistic IoT network traffic scenario [47].

9. BOUN_DDoS Dataset (2020):

Specifically, distributed denial-of-service (DDoS) assaults on Internet of Things networks are the focus of the BOUN_DDoS dataset. It contains both regular and DDoS attack traffic, enabling the assessment of machine learning methods for identifying and averting DDoS assaults in Internet of Things settings [48].

10. IoTID20 Dataset (2020):

The purpose of the IoTID20 dataset is to assess intrusion detection in Internet of Things networks. It covers a variety of attack types, including ARP spoofing, HTTP flooding, UDP flooding, Brute Force, and Syn flooding. The dataset offers a thorough assessment platform for ML-based IDS and attempts to capture the distinctive features of IoT assaults [49].

These datasets are useful tools for ML algorithm testing, training, and benchmarking in IoT-IDS. They are used by researchers to assess the effectiveness, precision, and efficiency of intrusion detection systems in identifying different kinds of assaults in Internet of Things networks. Researchers can aid in the creation of more reliable and efficient IDS solutions for protecting IoT devices by using representative datasets. Researchers can view the datasets and their attributes in Table 2.

Table 2: datasets and properties

Ref	Dataset name	Year	No. of Classes	Attack Classes
[40]	KDDCup-99	1998	4	Normal ,DoS, Probe, R2L, U2R
[41]	Kyoto_2006	2006	2	Attacks, not Attacks
[42]	NSL_KDD	2009	4	Normal, DoS, Probe, R2L, U2R
[43]	UNSW_NB15	2015	9	Analysis, Backdoor, DoS, Exploits, Fuzz, Generic, Normal, Rec, Shell, Worms
[44]	AWID	2016	4	Keycracking, Key stream retrieving, Dos, Man in the M
[45]	CIC_IDS2017	2017	7	Brute, Heart Bleed, Bot net, DoS, D DoS, Webs, Infiltration
[45]	CSECIC_IDS2018	2018	7	Heart Bleed, DoS, Botnet, DDoS, Force, Infiltration, Web
[46]	LITNET_2020	2020	12	Smurf, ICMP Flood, UDP Flood, SYN flood, HTTP Flood, LAND, W32, Code Red, SPAM, Reaper Worm, Scan, Packet Frag
[47]	BOUN_DDoS	2020	2	Attacks, not Attacks
[48]	IoTID20	2020	9	Normal, Syn Flooding, Brute Force, HTTP Flooding, UDP Flooding ARP Spoofing Host Port, OS Normal,Syn Flooding,Brute Force, HTTP Flooding, UDP FloodingARP SpoofingHost Port, OS

Table 2 shows how the data utilized in IoT-IDS is oriented to give thorough analysis and performance testing of intrusion detection systems in Internet of Things networks. This data represents real-world difficulties in this environment and covers a range of attack types directed towards IoT networks. The history of this data extends from 1998 to 2020, and this shows the development that has occurred in the field of intrusion detection over the years as well as the emergence and development of IoT technologies. This data provides a variety of attack types such as Denial of Service (DoS), Probe, Remote to Local (R2L), User to Root (U2R), Analysis, Backdoor, Exploits, Fuzzers, Generic, Reconnaissance, Shellcode, Worms, Key cracking, Keystream retrieving, Man-in-the-Middle, Brute Force, HeartBleed, Botnet, DDoS, Web, Infiltration, and others. This can help researchers evaluate the efficiency and accuracy of intrusion detection systems in detecting a wide range of potential attack types. Some of the statements focus on simulating specific IoT network challenges, such as wireless communications, IoT attacks, and distributed DDoS attacks. This helps in evaluating the performance of intrusion detection systems in this specific environment. Various intrusion detection techniques are used, including intrusion detection, anomaly detection, machine learning, and network analysis. This enhances the diversity of tools and techniques used to develop intrusion detection systems in IoT networks.

6. Challenges and Future Directions

6.1 Challenges in Applying ML in IoT-IDS

Applying ML (ML) in IoT-IDS (IDS) poses several challenges. Firstly, IoT environments' diverse and dynamic nature makes it difficult to create effective and generalizable ML models. Secondly, large-scale and heterogeneous IoT data require preprocessing techniques and F.S. methods tailored to IoT-specific characteristics. Thirdly, the limited computational resources of IoT devices restrict the complexity and size of ML models that can be deployed. Additionally, real-time intrusion detection is crucial in IoT systems, demanding low-latency ML algorithms. Ensuring the security and privacy of IoT data and handling high dimensionality and noise in IoT data are further challenges in ML-based IoT-IDS. Lastly, the scarcity of labeled training data for IoT-specific attacks hinders the development of accurate and robust ML models.

6.2 Potential Future Research Directions

In applying ML in IoT-IDS, several potential future research directions can be explored. Firstly, developing advanced ML algorithms that are specifically designed to handle the unique characteristics of IoT data, such as heterogeneity, high dimensionality, and dynamicity, can enhance the performance of IDS. Secondly, integrating anomaly detection techniques with ML models to detect emerging and previously unseen attacks in real time is an important area of research. Thirdly, exploring federated learning approaches that enable

https://doi.org/10.25130/tjps.v29i3.1553

collaborative learning among distributed IoT devices while preserving data privacy can address the challenges of limited computational resources and data privacy. Additionally, investigating the use of explainable A.I. techniques to enhance the transparency and interpretability of ML-based IDS in IoT can facilitate trust and adoption. Additionally, investigating the integration of machine learning (ML) with other cybersecurity methods like encryption and secure communications can offer all-encompassing security solutions for Internet of Things environments. Last but not least, developing benchmark datasets and assessment frameworks especially for machine learning-based IoT-IDS can facilitate the standardization of assessment and comparison of various methodologies, encouraging more developments in the field.

7. Conclusion

In conclusion, machine learning (ML) has become a viable method for Internet of Things (IoT) intrusion detection. The research examined in this paper shows how well machine learning (ML) algorithms work to identify intrusions and improve the security of Internet of Things (IoT) systems. To increase the precision and effectiveness of IDS (IDS) in IoT, several machine learning (ML) approaches have been used, including ensemble learning and F.S. methods. The dynamic nature of IoT data and the constrained computing capacity of IoT devices are two obstacles that still need to be addressed. Notwithstanding these difficulties, this field of study is still developing. Upcoming paths include creating sophisticated machine learning algorithms specific to the properties of IoT data, incorporating anomaly detection methods, and investigating explainable A.I. and federated learning strategies. ML-based intrusion detection systems (IDSs) in the Internet of Things (IoT) can enhance security and ensure the secure and dependable functioning of IoT systems by tackling these obstacles and venturing into novel research directions.

References

[1]. Guo, H., Goodchild, M. F., & Annoni, A. (2020). Internet of Things :In Manual of Digital Earth. Springer, Singapore, pp. 253-270. https://doi.org/10.1007/978-981-32-9915-3_11

[2]. Bellini, P., Nesi, P., & Pantaleo, G. (2022). IOT-enabled Smart Cities: A review of concepts, frameworks and Key Technologies. Applied Sciences, 12(3), 1607. https://doi.org/10.3390/app12031607

[3]. Syed, A. S., Sierra-Sosa, D., Kumar, A., & Elmaghraby, A. (2021). IOT in Smart Cities: A Survey of Technologies, practices and challenges. Smart Cities, 4(2), 429–475. https://doi.org/10.3390/smartcities4020024

[4]. Bhushan, B., Kumar, A., Agarwal, A. K., Kumar, A., Bhattacharya, P., & Kumar, A. (2023). Towards a secure and sustainable internet of medical things (IOMT): Requirements, design challenges, security techniques, and future trends. Sustainability, 15(7), 6177. https://doi.org/10.3390/su15076177

[5]. Tariq, U., Ahmed, I., Bashir, A. K., & Shaukat, K. (2023). A critical cybersecurity analysis and future research directions for the internet of things: A comprehensive review. Sensors, 23(8), 4117. https://doi.org/10.3390/s23084117

[6]. E. Altulaihan, M.A. Almaiah, and A. Aljughaiman. (2022). Cybersecurity Threats, Countermeasures and Mitigation Techniques on the IoT: Future Research Directions, Electronics, vol. 11, no. 11, p. 3330,. https://doi.org/10.3390/electronics11203330

[7]. Altulaihan, E., Almaiah, M. A., & Aljughaiman, A. (2022). Cybersecurity Threats, Countermeasures and Mitigation Techniques on the IoT: Future Research Directions. Electronics, 11 (11), 3330. https://doi.org/10.3390/electronics11203330

[8]. Gyamfi, E., & Jurcut, A. (2022). Intrusion Detection in Internet of Things Systems: A Review on Design Approaches Leveraging Multi-Access Edge Computing, Machine Learning, and Datasets. Sensors, 22 (10), 3744. https://doi.org/10.3390/s22103744

[9]. Arshad, J., Azad, M. A., Abdeltaif, M. M., & Salah, K. (2020). An Intrusion Detection Framework For Energy Constrained IoT Devices. Mech. Syst. Signal Process., 136 , 106436. https://doi.org/10.1016/j.ymssp.2019.106436

[10]. Mazhar, T., Talpur, D. B., Shloul, T. A., Ghadi, Y. Y., Haq, I., Ullah, I., ... & Hamam, H. (2023). Analysis of IoT Security Challenges and Its Solutions Using Artificial Intelligence. Brain Sciences, 13(4), 683.

[11]. Saheed, Y. K., Abiodun, A. I., Misra, S., Holone, M. K., & Colomo-Palacios, R. (2022). A machine learning-based intrusion detection for detecting internet of things network attacks. Alexandria Engineering Journal, 61 (12), 9395-9409.

[12]. Kaur, B., Dadkhah, S., Shoeleh, F., Neto, E. C. P., Xiong, P., Iqbal, S., ... & Ghorbani, A. A. (2023). Internet of things (IoT) security dataset evolution: Challenges and future directions. Internet of Things, 100780.

[13]. Benkhelifa, E., Welsh, T., & Hamouda, W. (2018). A Critical Review of Practices and Challenges in Intrusion Detection Systems for IoT: Toward Universal and Resilient Systems. IEEE Communications Surveys & Tutorials, 20 (4), 3496-3509. https://doi.org/10.1109/COMST.2018.2844742

[14]. Sicato, J. C. S., Singh, S. K., Rathore, S., & Park, J. H. (2020). A comprehensive analyses of intrusion detection system for IoT environment.

https://doi.org/10.25130/tjps.v29i3.1553

Journal of Information Processing Systems, 16 (4), 975-990.

[15]. Diro, A., Chilamkurti, N., Nguyen, V.-D., & Heyne, W. (2021). A Comprehensive Study of Anomaly Detection Schemes in IoT Networks Using Machine Learning Algorithms. Sensors, 21 , 8320. https://doi.org/10.3390/s21248320

[16]. Zehra, S., Faseeha, U., Syed, H. J., Samad, F., Ibrahim, A. O., Abulfaraj, A. W., & Nagmeldin, W. (2023). Machine Learning-Based Anomaly Detection in NFV: A Comprehensive Survey. Sensors, 23 , 5340. https://doi.org/10.3390/s23115340

[17]. Zikria, Y. B., Afzal, M. K., Kim, S. W., Marin, A., & Guizani, M. (2020). Deep learning for intelligent IoT: Opportunities, challenges and solutions. Computer Communications, 164 , 50-53.

[18]. Gerodimos, A., Maglaras, L., Ferrag, M. A., Ayres, N., & Kantzavelou, I. (2023). IoT: Communication protocols and security threats. Internet of Things and Cyber-Physical Systems, 3, 1–13.

[19]. Al-Amiedy, T. A., Anbar, M., Belaton, B., Kabla, A. H. H., Hasbullah, I. H., & Alashhab, Z. R. (2022). A Systematic Literature Review on Machine and Deep Learning Approaches for Detecting Attacks in RPL-Based 6LoWPAN of Internet of Things. Sensors, 22 , 3400. https://doi.org/10.3390/s22093400

[20]. Rodríguez-Rodríguez, I., Campo-Valera, M., Rodríguez, J.-V., & Frisa-Rubio, A. (2023). Constrained IoT-Based Machine Learning for Accurate Glycemia Forecasting in Type 1 Diabetes Patients. Sensors, 23, 3665. https://doi.org/10.3390/s23073665

[21]. Asharf, J., Moustafa, N., Khurshid, H., Debie, E., Haider, W., & Wahab, A. (2020). A Review of Intrusion Detection Systems Using Machine and Deep Learning in Internet of Things: Challenges, Solutions and Future Directions. Electronics, 9 (7), 1177. https://doi.org/10.3390/electronics9071177

[22]. Shahid, J., Ahmad, R., Kiani, A. K., Ahmad, T., Saeed, S., & Almuhaideb, A. M. (2022). Data Protection and Privacy of the Internet of Healthcare Things (IoHTs). Appl. Sci., 12 (4), 1927. https://doi.org/10.3390/app12041927

[23]. Taherdoost, H. (2023). Security and Internet of Things: Benefits, Challenges, and Future Perspectives. Electronics, 12 (8), 1901. https://doi.org/10.3390/electronics12081901

[24]. Arhore, S. A. (2022). Intrusion Detection in IoT Systems using Machine Learning (Doctoral dissertation, Dublin, National College of Ireland).

[25]. Islam, N., Farhin, F., Sultana, I., Kaiser, M. S., Rahman, M. S., Mahmud, M., ... & Cho, G. H. (2021). Towards Machine Learning Based Intrusion Detection in IoT Networks. Computers, Materials & Continua, 69 (2).

[26]. Alghayadh, F., & Debnath, D. (2021). A hybrid intrusion detection system for smart home security based on machine learning and user behavior. Advances in Internet of Things, 11 (1), 10-25.

[27]. Özer, E., İskefiyeli, M., & Azimjonov, J. (2021). Toward lightweight intrusion detection systems using the optimal and efficient feature pairs of the Bot-IoT 2018 dataset. International Journal of Distributed Sensor Networks, 17(10), 15501477211052202.

[28]. Gad, A. R., Haggag, M., Nashat, A. A., & Barakat, T. M. (2022). A Distributed Intrusion Detection System using Machine Learning for IoT based on ToN-IoT Dataset. International Journal of Advanced Computer Science and Applications, 13(6).

[29]. Amouri, A., Alaparthy, V. T., & Morgera, S. D. (2020). A machine learning based intrusion detection system for mobile Internet of Things. Sensors, 20(2), 461.

[30]. Ayub, M. Y., Haider, U., Haider, A., Tashfeen, M. T. A., Shoukat, H., & Basit, A. (2023). An Intelligent Machine Learning based Intrusion Detection System (IDS) for Smart cities networks. EAI Endorsed Transactions on Smart Cities, 7(1), e4-e4.

[31]. Baich, M., Hamim, T., Sael, N., & Chemlal, Y. (2022). Machine Learning for IoT based networks intrusion detection: a comparative study. Procedia Computer Science, 215, 742-751.

[32]. Othman, T. S., & Abdullah, S. M. (2023). An Intelligent Intrusion Detection System for Internet of Things Attack Detection and Identification Using Machine Learning. ARO-THE SCIENTIFIC JOURNAL OF KOYA UNIVERSITY, 11(1), 126-137.

[33]. Amarouche, s., & küçük, k. (2022). Machine and deep learning-based intrusion detection and comparison in internet of things. Journal of naval sciences and engineering, 18(2), 333-361.

[34]. Wang, Y., Sun, T., Li, S., Yuan, X., Ni, W., Hossain, E., & Poor, H. V. (2023). Adversarial Attacks and Defenses in Machine Learning-Powered Networks: A Contemporary Survey. ArXiv, abs/2303.06302.

[35]. Banaamah, A. M., & Ahmad, I. (2022). Intrusion Detection in IoT Using Deep Learning. Sensors, 22(21), 8417.

[36]. Bhavsar, M., Roy, K., Kelly, J., & Olusola, O. (2023). Anomaly-based intrusion detection system for IoT application. Discover Internet of Things, 3(1), 5.

[37]. Nimbalkar, Pushparaj & Kshirsagar, Deepak. (2021). Feature selection for intrusion detection system in Internet-of-Things (IoT). ICT Express. 7. 10.1016/j.icte.2021.04.012.

[38]. Rodríguez, M., Alesanco, Á., Mehavilla, L., & García, J. (2022). Evaluation of Machine Learning Techniques for Traffic Flow-Based Intrusion Detection. Sensors, 22(23), 9326.

[39]. Rose, J. R., Swann, M., Bendiab, G., Shiaeles, S., & Kolokotronis, N. (2021, June). Intrusion detection using network traffic profiling and machine learning for IoT. 2021 IEEE 7th International

https://doi.org/10.25130/tjps.v29i3.1553

Conference on Network Softwarization (NetSoft) (pp. 409-415).

[40]. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, 1-6. https://doi.org/10.1109/CISDA.2009.5356528

[41]. Song, J., Takakura, H., Okabe, Y., Eto, M., Inoue, D., & Nakao, K. (2011). Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation. Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS '11) (pp. 29–36). https://doi.org/10.1145/1978672.1978676

[42]. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A Detailed Analysis of the KDD CUP 99 Data Set. Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.

[43]. Moustafa, N., & Slay, J. (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). 2015 Military Communications and Information Systems Conference (MilCIS), 1-6. https://doi.org/10.1109/MilCIS.2015.7348942

[44]. Chen, J., Yang, T., He, B., & He, L. (2021). An analysis and research on wireless network security dataset. 2021 International Conference on Big Data Analysis and Computer Science (BDACS), 80-83. https://doi.org/10.1109/BDACS53596.2021.00025

[45]. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSp, 1, 108-116.

[46]. Damasevicius R, Venckauskas A, Grigaliunas S, Toldinas J, Morkevicius N, Aleliunas T, Smuikys P. (2020). LITNET-2020: An Annotated Real-World Network Flow Dataset for Network Intrusion Detection. Electronics, 9(5), 800. https://doi.org/10.3390/electronics9050800

[47]. Derya Erhan. (2019). Boğaziçi University DDoS Dataset. IEEE Dataport. https://dx.doi.org/10.21227/45m9-9p82

[48]. Ullah, I., & Mahmoud, Q. H. (2020, May). A scheme for generating a dataset for anomalous activity detection in iot networks. In Canadian conference on artificial intelligence (pp. 508-520). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-47358-7_52.