International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I4P128

Volume 73 Issue 4, 341-367, April 2025

Review Article

A Comprehensive Review: Performance Analysis of Machine and Deep Learning Techniques for Intrusion Detection Using Synthetic Data

Neha[1], Abhishek Kajal[2]

[1,2]*Department of Computer Science, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India.*

[*]*Corresponding Author: nehasharma31066@gmail.com*

Received: 12 November 2024 Revised: 21 March 2025 Accepted: 01 April 2025 Published: 26 April 2025

Abstract *- In current times, the security of information for critical infrastructures has become extremely crucial. The most common threat faced by CIIs is in the form of frequent network intrusions. This article provides a comprehensive analysis of various intrusion detection techniques and cybersecurity measures for critical infrastructures. The main aim is to present a comparative analysis of network performance in the presence of DDoS, DoS, malware, APT and Ransomware attacks while also analyzing solutions to mitigate these challenges in different sectors of critical infrastructures like healthcare, government, defense, energy, and other online platforms. The study evaluates the effectiveness of emerging ML and DL approaches such as DT, RF, SVM, CNN, LSTM, GRU, RNN, etc. The most widely used datasets, such as KDD-Cup99, NSL-KDD, CICIDS2017-18-19, BOT-IoT, and TON-IoT, were also analyzed for evaluating the efficiency proposed by researchers for safeguarding CIIs. The dataset analysis investigates the performance dependence against the features used in feature engineering, followed by feature selection and feature extraction techniques. This review study also provides an overview of NIDS, Anomaly, behavior-based detection and IPS. This article analyzes the recent papers and highlights the significance of thorough testing on large datasets and the need for real-time situation comparisons to understand the effectiveness of these methods in protecting IDS for CIIs. Most of the information in the article is extracted from reputed database depositories and research articles retrieved from 2005 to 2024. In the end, the various challenges and recommendations will be outlined to be helpful in future research directions. Moreover, these ML and DL techniques were implemented on a synthetic dataset against three cyber-attack types: DDoS, SQL Injection and ransomware. We observed that the accuracy of DL-based techniques improved with the increase in the number of data samples, getting 98.8 accuracy for CNN against a sample of 10 lakh instances, including 20 attributes.*

Keywords *- Machine Learning, Deep Learning, Critical Information Infrastructures, Publicly available datasets, Syntactic dataset, IDS.*

1. Introduction

In our society, certain resources are extremely important for essential tasks, and these can be either physical (like buildings and facilities) or related to computer systems. Because of their crucial roles, authors call these resources critical infrastructures. These infrastructures have specific network setups and regularly handle repeated transactions among different points [1]. Authors have to merge or combine technologies into these critical infrastructures to improve their productivity and meet society's demands as per increases the societal needs as our population grows day by day. Our computer systems are linked to these infrastructures, which are called critical information infrastructures. These systems manage data as well as supervise other physical systems. It can impact services provided by the physical devices if these information systems are attacked by the attackers [2]. The usage of technology in critical infrastructures has rendered them vulnerable to cyber-attacks. Intruders can attack CII [3] weaknesses on both the network and the application layers. Such intrusions affect the security, availability, and integrity of the information stored inside these infrastructures. In simple terms, intruders can remotely hack and be controllers of these information systems, resulting in hardware failures, software malfunctions, and potentially dangerous situations [4]. It is important to secure and maintain the stability of interlinked data or information infrastructures. This is important for the easy operation of a nation's critical systems, as marked by the latest studies [5]. The security of CII (critical information infrastructures) [6] is of overriding value for individuals, governments and organizations. Critical information infrastructures represent the necessary systems, networks, and assets that hold a nation's functioning, including those in the domains of healthcare, energy, finance, transportation, industries, water and many more [7]. The uninterrupted growth of digital techniques and the growing

belief in the internet have made these infrastructures more prone to cyber threats, necessitating robust cybersecurity measures to secure them. One vital element of this cybersecurity [8] framework is that IDS opposes various kinds of web-based attacks [9]. One of the most noticeable Web-based intrusive attacks [10] is DDoS. Some other threatful web-based attacks are cross-site scripting and SQL injection [11], which may stance critical dangers to CIIs and could affect in the loss of personal secured information and financial harm [12].

1.1. Critical Information Infrastructure

Critical information infrastructure [13,14] refers to the essential devices, resources, and networks necessary for the operation and comfort of a nation, its budget, and its citizens. These infrastructures embrace various sectors, including energy grids [15], telecommunications, transportation systems, financial institutions, healthcare facilities, smart cities [16] and government agencies. An attack on any of these sectors can have terrible consequences, disrupting services and vulnerable, sensitive data and possibly causing important economic and societal damage [17].

1.1.1. Challenges in Securing CII RQ1

The growing digitization of CII has opened these sectors to an increasing threat picture [18]. Cyberattacks, ranging from ransomware [19] and data breaches to intelligent nation-state-sponsored intrusions, have become more frequent and intelligent. As a result, guaranteeing the security of critical infrastructure has become a high priority for governments and organizations [15].

1.2. Benefits of IDS in CII Security

1.2.1. Early Threat Detection

IDS systems can identify possible threats in their early stages, enabling fast response and mitigation.

1.2.2. Real-time Threat Detection

IDS endlessly supervise network traffic, supplying real-time alerts and informing, allowing efficient reaction to expected threats.

1.2.3. Incident Response

IDS can provide important data for incident response, helping organizations realize the nature and scope of a cyber-attack [20].

1.2.4. Regulatory Compliance

In many sectors, agreement with cyber-security regulations and standards is compulsory [21] [22]. IDS can aid in meeting these requirements.

1.2.5. Reduced Response Time

With the power to rapidly determine and categorize threats, IDS decrease the reaction time to security incidents, minimizing the impact of attacks and expected harm.

1.2.6. Enhanced Network Visibility

IDS gives precious insight into network traffic, distinguishing trends, patterns, and expected danger. This content aids in increasing network security by implementing essential cautiousness.

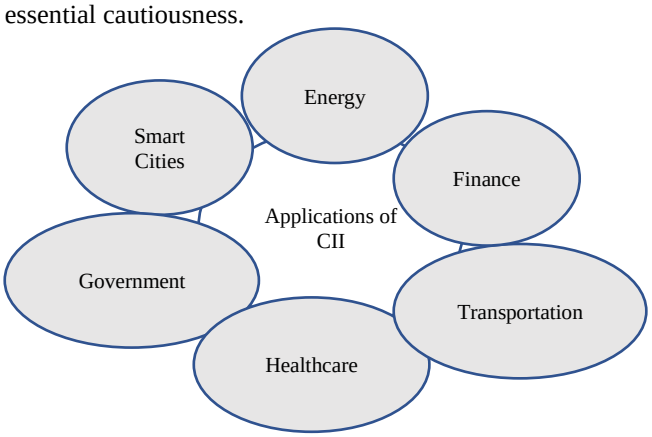

Fig. 1 Applications of CII

1.3. Application Across Different Sectors

1.3.1. Government

Governments use IDS to defend sensitive information, national security assets, and critical infrastructure such as transportation networks and power grids.

1.3.2. Finance

IDS systems are essential for protecting financial institutions, which are attractive targets for cybercriminals trying to steal funds or sensitive client data [19].

1.3.3. Healthcare

In the healthcare sector, IDS can help protect patient data, medical devices, and critical systems like electronic health records forties [23][24].

1.3.4. Energy

The energy sector relies heavily on IDS to defend power grids, oil and gas facilities, and other infrastructure from cyber threats [25] [26] [27].

1.3.5. Transportation

Transportation systems, including air traffic control and railways, use IDS to maintain the safety and reliability of their operations. And many more like: IoT's vulnerability to secret communication threats in areas such as Health and Manufacturing IoT. It analyzes the IEEE802.15.4 protocol, importantly DSME behavior, to evaluate risks and proposes functional security present [28]. The use of Model-Based Systems Engineering (MBSE) in the space industry underscores the need for robust cybersecurity in progressively tangled cyber-physical space systems [8].

1.4. Intrusion Detection in CII RQ2

Intrusion Detection Systems (IDS) [29] [30] date back

several decades and has acquired significantly over a period of time. IDS is a hypercritical element of cybersecurity, planned to supervise systems activities for the mark of unauthorized access, misuse, or other malicious activities, such as web-based attacks [12]. Combining Intrusion Detection Systems into critical information infrastructure is a foundation of modern cyber-security. It improves the ability to monitor, detect, and respond to cyber threats effectively, protecting the essential systems that support our daily lives. As technology continues to advance and threats become more intelligent, the role of IDS in protecting CII becomes increasingly critical [31].

1.5. Purpose of IDS

Intrusion detection systems function as a progressive security state, assisting companies in securing their digital properties and sensitive information by quickly identifying and responding to security occurrences [32]. It benefits other security measures like antivirus and firewall software by concentrating on detecting odd or suspicious activities within a network.

1.6. Detection Techniques of IDS

IDS employ several methods for detecting intrusions:

1.6.1. Signature-based Detection

This technique relies on already existing signatures or patterns of known attacks. When the network detects a pattern match, it generates an alert [33].

1.6.2. Anomaly-based Detection

This type of detection searches for departures from predetermined ranges of typical network activity. It improves alerts when activities deviate significantly from the norm [34].

1.6.3. Behavioral-based Detection

This method notices user and network behavior over time to detect deviations from well-known patterns [35].

1.6.4. Heuristic-based Detection

Heuristics relates to using formulas and algorithms to detect attainable malicious activity based on known threats, tricks and patterns.

1.7. Components

An IDS typically consists of the following components [16]:

1.7.1. Sensors or Data Collectors

Sensors are the devices that gather information and data from different sources, such as other systems and applications logs and network traffic.

1.7.2. Analysis Engine

It basically does work or processes on the collected data and information and then applies detection methods to detect achievable intrusions.

1.7.3. Alerting Mechanism

In this mechanism, when an intrusion is detected in the network, the IDS generates alerts and sends them to the administrators or SOC (Security Operation Center).

1.7.4. Reporting and Logging

It may generate a summary for analysis and compliance purposes, and the intrusion detection system manages logs of detected activities.

1.8. Types of IDS

1.8.1. NIDS (Network-based IDS)

Monitors network traffic at various points within a network, typically at network boundaries [36] [37].

1.8.2. Host-based IDS (HIDS)

Installed on personal devices, it traces activities on the host itself, searching for suspicious behavior.

1.8.3. Hybrid IDS

Merge both NIDS and HIDS for more comprehensive security protection.

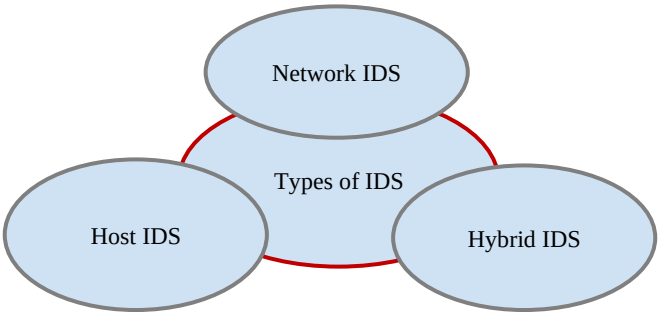

Fig. 2 Types of ID

1.9. Challenges and Threats to Intrusion Detection Systems

IDSs are a captious factor of recent cyber-security schemes. They assist companies in proactively identifying and responding to potential threats, increasing overall security posture and mitigating the effect of cyberattacks. However, they should be part of a layered security strategy that includes firewalls, antivirus software, and personnel training to provide strong protection against changing cyber threats. IDSs face many threats that can make them ineffective and uncompromising security systems [15].

1.9.1. Evasive Tactics

As attackers become more skilful at dodging detection, NIDS faces situations when distinguishing and analyzing malicious traffic patterns. Attackers often use bafflement methods like encryption, tunneling, and polymorphic malware to get around conventional signature-based detection techniques. Moreover, so-called" zero-day" exploits, which exploit vulnerabilities unknown to network administrators

[38] or security vendors, exist as an additive obstruction to NIDS. To justify these threats, NIDS can combine advanced anomaly detection algorithms confident in knowing normal network behavior. ML, AI, and behavioral analysis methods can improve NIDS' ability by detecting abnormal from normal traffic patterns and distinguishing expected threats based on dynamic behavior. In addition, real-time threat intelligence feeds can supply up-to-date content on emerging attack vectors, allowing NIDS to change rapidly to new threats.

1.9.2. Distributed Denial of Service (DDoS) Attacks

Most of the methods used to overpower NIDS (Network Intrusion Detection System) and go near its ability are introduced by DDoS attacks. These attacks overwhelm web and system resources, rendering the NIDS ineffective owing to the sheer volume of malicious traffic. Attackers commonly launch large DDoS attacks[39] using botnets, which usually consist of compromised IoT devices[40, 41, 38, 1]. In order to reduce the impact of these attacks, organizations should be proactive and invest in cloud-based DDoS protection services, develop redundant NIDS infrastructure [42], and properly partition their systems. Moreover, utilizing traffic technology methods such as traffic filtering, rate limitation, and anomaly detection can help recognize DDoS attacks without using NIDS resources excessively. The hybrid approach, which combines the Neural Networks and ABC (Artificial Bee Colony), demonstrated superior performance for detecting and mitigating cyber threats compared to stand-alone approaches. It was observed that the hybrid approach achieved lower delay and higher throughput with the help of simulation, indicating its effectiveness in identifying and responding to several types of cyber-attacks. These results recommend that mixing AI and swarm intelligence techniques can significantly improve network security against growing threats [43].

1.9.3. Advanced Persistent Threats APTs

APT means permanent because APTs are more secretive than other cyber threats and are a more advanced threat to cybersecurity. APTs are complicated attacks that generally go after valuable things and take advantage of faults. APTs have performed long-term smart operations that are usually missed by NIDS. An intruder may use various techniques like lateral movement, backdoor trojans, and increased privileges, which make it challenging for NIDS to recognize their existence [43]. The system of government should place a multi-dimensional defense method to prevent advanced persistent threats, in which systems must be modified and updated often, firm access controls must be applied, and progressive threat-hunting methods [44] must be used for ongoing monitoring. The NIDS's ability to recognize and respond to advanced persistent threats can be enhanced by utilizing SEIM (Security Information and Event Management) systems [45], which consist of real-time monitoring, event correlation, and log management. As one obtains a digital viewpoint, network intrusion detection systems are tackled with a continually evolving variety of threats. It is skilled in recognizing destructive traffic patterns and pleasing effective action, thanks to its advanced technologies, which consist of AI, ML [46] and behavior-based study. Furthermore, a comprehensive defensive strategy should be used by enterprises and combining DDoS protection services, network partitioning and SIEM solution to support their complete security position. To secure our digital infrastructures, NIDS can continue adapting and giving new features to the ever-changing cyber-attack dynamics.

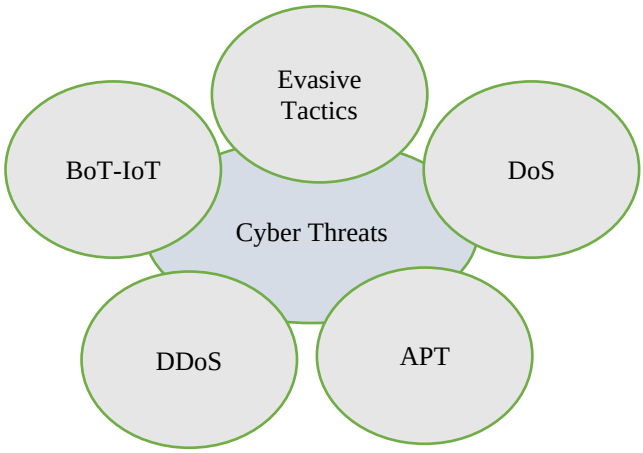

Fig. 3 Some Cyber Security Threats

1.9.4. Malware

Malware constantly [47] changes and evolves in order to avoid detection. Polymorphic malware modifies its code structure, making it difficult for traditional detection approaches to identify and classify such attacks [48, 49, 50, 51, 52].

1.9.5. Ransomware

Ransomware is a popular attack that targets the diverse sectors of critical infrastructures, including public companies, healthcare, telecom, transportation, etc., but this attack hits deeply on industrial areas and produces major financial losses to the users and victims. It can effectively run in several environments with a smaller number of processing resources and minimal memory. The ransomware attack has two types of behavior, i.e., locking the services or encrypting the data, which may affect the victim's operation and data, which is why it's required to be proactive due to the irreversible damage [19].

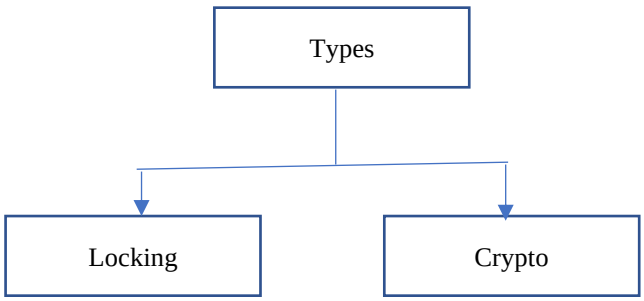

Fig. 4 Types of Ransomware Attack

1.10. Significance of Intrusion Detection Systems

1.10.1. Protecting Confidentiality

An intrusion detection system helps in securing sensitive data or information and also prevents unauthorized access to critical information. Intrusion detection systems can eliminate the expected gaps and preserve confidentiality.

1.10.2. Ensuring Integrity

IDS identify attempts to change data, ensuring the system's integrity. IDS detects illegal changes to system packages by continuously evaluating them and comparing them to established rules, alerting administrators about modifications by unauthorized persons.

1.10.3. Preserving Availability

IDS help to manage network availability by detecting patterns of DoS attacks that might flood network resources. By rapidly identifying these types of attacks, IDS enables network managers to take critical measures to ensure continuous operation.

1.11. Features of Intrusion Detection Systems

Features of IDS are given below [40]:

1.11.1. Signature-based Detection

This sort of IDS employs a repository of already existing attack patterns, called signatures, to determine predicted threats. By comparing network packets that contradict these signatures, IDS can spot and alert administrators about known attack patterns.

1.11.2. Anomaly-based Detection

These IDSs launch a standard of normal system behavior and endlessly supervisor for any abnormality from that standard. If the state falls outside the normal range, it is considered anomalous and triggers an alert.

1.11.3. Host-based and Network-based IDS

Host-based IDS analyzes actions on a single machine, whereas network-based IDS examines network traffic to identify potential risks. Some IDS combine both techniques to provide full protection.

1.12. Intrusion Prevention System (IPS)

IPS, a delay of IDS, not only identifies and informs on harmful activity but also takes automated steps to prevent it from harming the network [53].

1.13. Research Gap

While numerous studies focus on traditional ML and DL techniques, they often lack a comprehensive evaluation of feature engineering techniques, class balancing strategies, and explainability methods. Additionally, limited research integrates ML and DL-based architectures for intrusion detection, making it a promising area for exploration. Our work addresses how it advances beyond prior studies and improves detection performance by demonstrating an artificial intrusion detection dataset to show the impact of techniques and workflow for the researcher to know the process and how efficiently to learn the data to enhance the performance of models. With the help of this review, researchers will get complete practical knowledge about working ML and DL approaches in the field of intrusion detection systems for CII because, in the previous research works, researchers did not clearly provide the relevant information for every step in a single paper. Hereby, our work clearly addresses how it advances beyond prior studies and improves detection performance.

2. Related Work

Numerous papers have been published on using ML and DL-based approaches to detect attacks on critical infrastructure. This section examines existing research studies on cyber-security by examining previous research and conclusions. The summaries are arranged by ML/DL [57] approaches, followed by cybersecurity challenges for IDS and affect the important infrastructures. Naseer et al. (2018) proposed Improved Network Anomaly Detection using DNNs on NSL-KDD with the help of KNN, Decision tree, CNN and LSTM for intrusion detection. The anomaly detection system gives promising results in real-world applications using deep IDS models [9]. In this Hatcher et al. (2018) survey about platforms, Applications and Emerging Research Trends like neural networks, the Internet of Things and cyber-physical systems in deep learning [5]. After that, Xin Yang et al. (2018) Discuss the methods of IDS for cybersecurity in ML and DL and some commonly used datasets and also discuss some challenges using ML/DL and give suggestions for research directions in the cybersecurity field [21].

R. Vinaya Kumar et al. (2019) proposed malware detection in robust intelligence using deep learning and achieved zero-day malware detection with the help of hybrid deep learning and scalable framework for real-time deployments on virtual box and cuckoo sandbox datasets [47]. Derhab et al. (2020) proposed based on TCNN and efficient feature engineering of intrusion detection for IoT using a combination of LSTM and CNN on the BoT-IoT dataset and achieved 99.9986% accuracy for multi-class traffic detection. Design another principle for testing IDS against adversarial attacks [63].

In the field of cybersecurity, Guangming Xian et al. (2020) proposed cyber-intrusion prevention on a big scale using a semi-supervised discriminative deep belief network in DL- based on local and non-local regularization and improve the performance of intrusion prevention system and time decreases as the number of hidden layers increases on KDD Cup99 and NSL-KDD datasets [85]. Review the approaches by Akeem Alimi et al. (2020) for Power System Security and Stability with ML approaches like ANN, DT, and SVM and studies about reinforcement and deep reinforcement learning

techniques for transient stability and give directions for future work [86]. Susilo et al. (2020) proposed an algorithm for detecting DoS attacks using DL using Python language with the scikit learn package TensorFlow and Seaborn to increase accuracy and effectively mitigate attacks on IoT networks. Authors combine several algorithms of ML and DL in order to mitigate attacks on NIDS in real time [20].

In this work, Yadigar et al. (2020) compare and discuss the cyber-attack detection for the modem development state by using DL, which was investigated using a combination of CNN, DBN, RNN, LSTM, Autoencoder using a different KDD99, NSL-KDD, etc. This work highlights the primary problem for minor classes of IDS complexity and their low accuracy. By applying the hybrid method, authors can improve these issues [22]. Proposed the Imbalanced Network Traffic by Liu Lan et al. (2020) for intrusion detection using ML and DL on NSL-KDD and CSE-CIC-IDS2018 and also proposes a novel Difficult Set Sampling Technique and this improves the imbalanced original training set and provides targeted augmented data for minor class [87]. And in the field of IoT, Idrissi et al. (2021) proposed IoT against botnet attacks for IDS using deep learning with CNN on the BoT-IoT dataset and compared with other algorithms like RNN, GRU and LSTM and achieved 99.94% accuracy and 0.58% validation loss and predicted execution time is less than 0. 34ms. Try to use self-supervised learning to produce an updated and powerful model for autonomous IDS in the future [88].

Fatani et al. (2021) Proposed Enhanced Transient Search Optimization and an efficient AI mechanism for IDS in IoT using CNN and TSODE on KDDCup99, NSL-KDD, Bot-IoT and CICIDS-2017 and achieves improved accuracy as compared to existing approaches [89]. After that, Alkahtani et al. (2021) proposed an advanced Internet of Things infrastructure for intrusion detection system using deep learning techniques CNN, LSTM and CNN-LSTM and achieved an accuracy of 96.60%,99.82% and 98.80%, respectively, on the IoTID20 dataset and improve the security of IoT environment [48].

Followed by Ullah et al. (2021) proposed a deep learning-based model and design for anomaly detection in IoT networks using CNN in D, 2D and 3D and validating BoT-IoT, MQTT-IoT-IDS2020 and IoT-23 intrusion detection datasets and then transfer learning used to implement binary and multiclass classification and achieved improved accuracy, F1-score, precision and recall as compared to existing one. Authors can use RNN, GAN, and FFN to detect anomalies in the future [90]. But in this research, Zhu Tianqing et al. (2022) workaround privacy by applying Differential Privacy in key areas of AI in multi-agent systems using DL and ML; this study delivers a fresh view of possibilities for enhancing AI performance [12]. Proposed a new approach by Bar Rotem et al. (2022) at the packet level for traffic detection by using SimCSE inspired by NLP on the USTC-TFC2016 dataset and achieved 99.99% accuracy and 99.98% of the ISCXVPN2016 dataset. This approach effectively detects network traffic and is robust in zero-day attack detection. Authors can apply different encryption protocols on different datasets [31]. In this, Halbouni et al. (2022) review ML and DL approaches for cybersecurity for intrusion detection systems. This discussion discusses the algorithms, learning approach, applications, datasets and network implementation [32].

The work in this paper by researchers Qadir et al. (2022) focuses on availability and its dependent factors like reliability, timeliness, and accessibility. Then, they measure the impact of DoS attacks using empirical security research. As a result, the accessibility is degraded, which leads to the DoS attack and the DoS impact on response time [4]. And Wu Zihan et al. (2022) proposed an approach based on a Robust Transformation for intrusion detection system using fuzzy neural network and LSTM and achieved 99.17% of the F1-score on the CICIDS2017 dataset was 98.48% of F1-score on CIC-DDoS2019 and authors also studied about RNN in deep learning and SVM in machine learning and authors improve the speed of transformer technique for quick response in the future [50].

Ferrag et al. (2022) proposed new comprehensive Realistic datasets and applications of IoT and IIoT in cyber security for Federated and centralized learning of Edge-IIoTest using deep learning, and these datasets are publicly accessed on five attacks like DoS/DDoS, Man in the middle, information gathering, malware and injection attacks [53]. Sogut et al. (2023) Propose a multi-model for the classification and detection of DDoS attacks on SCADA Systems using CNN, LSTM, RF, DT, KNN, Naive Bayes, a hybrid of CNN+LSTM techniques of ML and DL using datasets from tested and achieve 98% accuracy by hybrid and security improved effectively. In the future, the effect of DDoS attacks will be reduced using different techniques and environments and by detecting different types of attacks [57].

Sadhwani et al. (2023) Propose a Lightweight Model to detect DDoS attacks using machine learning techniques like RF, Naive Bayes, ANN, KNN, and Logistic regression on limited features of BoT-IoT and TON-IoT datasets and achieve 100% accuracy with Naive Bayes and RF respectively. In future, researchers will implement different IoT datasets or implement using deep learning techniques, and to detect any attack, authors will create a front-end application [58]. A dynamic IDS for critical information infrastructure using m-SVM machine learning technique in real time and use principal component analysis for feature reduction and achieve 97.64% accuracy with a 99.20 detection rate; this work proposed by Adejimi et al. (2023) and authors will enhance the performance of the system by using deep learning techniques in the future [59]. Alqudhaibi et al. (2023) Propose a Proactive Approach based on Attacker Motivations for predicting cybersecurity threats in critical infrastructure for

Industry 4.0. Achieve 66% FPR with trained and tested datasets using machine learning techniques and detecting malware, DDoS attacks and Jamming and Spoofing. Accuracy depends on the datasets [91]. Akmal Khalid et al. (2023) Review Game Theory Approaches to detect and defend against Advanced Persistent Threats and outlines the challenges and new opportunities to bypass tactics and techniques for defenses [43].

Henry et al. (2023) a DL technique was used to develop an IDS in this research. They used CNN with the GRU framework. The dataset considered in this work proposes a variety of attacks and a huge number of samples, which were used to verify the recommended model. The suggested model achieved a low FP rate and a classification accuracy rate of 98.73%. They used nearly half of the total attributes compared with all other classifiers. Authors can improve performance by optimizing techniques [67].

Alnifie et al. (2023), In this research paper, the work focuses on releasing the estimation of optimism bias and its impact on human perception using a meta-analysis due to inappropriate risk perception. Optimism bias has a huge impact on overall cyber security, and humans are more targeted than others. In the future, researchers can develop effective interventions to reduce the risk of security breaches [92].

Wang et al. (2023) Authors focus this paper on a case with ROP payload detection for tackling imbalanced data in cybersecurity using a transfer learning approach in deep learning with the help of an imbalance dataset. As a result of this approach, the number of false positives is reduced, and detected malicious samples have also decreased. In the future, researchers can use samples of minority-class data and high-quality source data [93].

Alghamdi et al. (2023) In this research work, the authors focus on IoT of an ensemble Deep learning-based IDS using lambda architecture with LSTM and convolution neural network and artificial neural network classifiers that give high accuracy with less useful processing time in a multi-pronged classification strategy as a result and in future authors can use automated machine learning techniques for tuning the hyper-parameters [94].

Adesina et al. (2023) Review Wireless communication using RF data on Adversarial machine learning for wireless security and discusses AML attack types and discusses the problems as well as defense mechanisms of wireless communication [81]. Park et al. (2023) proposed a classification system for CAN Protocol using Graph-based Intrusion detection in Vehicle security and achieved an improved accuracy of 9% compared to existing intrusion detection in combined attacks and enhancing the security of in-vehicle networks [95].

Gehlot et al. (2022) research describes a neural network-based intrusion detection system (IDS) that protects vital infrastructure from cyber assaults. It underlines the importance of intrusion detection systems (IDS) in quickly detecting and minimizing assaults, particularly in critical industries. The suggested system uses deep learning approaches such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks to improve detection accuracy. The system's effectiveness is demonstrated by experimental results, which show improvements in reaction time, latency, and resource utilization for real-time intrusion detection in emergency scenarios [75].

Abbas et al. (2021) This research investigates the use of machine learning (ML) techniques in Intrusion Detection Systems (IDS), with an emphasis on the ensemble approach for improved accuracy. It discusses Artificial Neural Networks (ANN) and Deep Learning (DL), supporting ensemble techniques due to their adaptability and ability to minimize errors. The proposed ensemble model, which comprises logistic regression, naive Bayes, and decision trees, uses the CICIDS2017 dataset and offers better accuracy than DL approaches while requiring less CPU power [83].

Ahmad et al. (2021) Ahmad and associates (2021) Benchmark datasets like KDD Cup '99 and NSL-KDD are used to test the precision, recall, false alarm rate, true negative rate, and accuracy of the assessment metrics for ML and DL-based IDS. In order to increase network security, future trends indicate the development of effective NIDS frameworks, solutions for complex model architectures, and research into DL approaches for IoT and cyber-physical systems. Challenges include dataset imbalance and real-world performance [96].

Genge et al. (2023) Utilizing the Vinyl-Acetate Monomer (VAM) process as a case study, the researchers explain the E-APT Detect technique for detecting Advanced Persistent Threats (APTs) in industrial processes. It highlights the system's ability to identify abnormalities generated by APTs, even when attackers seek to hide their existence by compromising observable parameters. Empirical evaluations emphasize the significance of accuracy in the attacker's manipulation technique and propose possible applications in real-world industrial systems [97]. Ntafloukas et al. (2022) The analysis revealed a significant reduction in cyber-physical risk with implementing Integrated Control Barriers (ICBs). Importance indexes were crucial in accurately assessing risk, with a sensitivity analysis showing a 58.05% increase in risk when attacks were carried out by terrorist organizations. Stakeholder involvement is vital in the assessment process to address the complex nature of cyber-physical risks in IoT-enabled transportation infrastructure [98]. He, Yanjie and Li Wei (2022). The proposed method makes use of spatial aspects of network traffic that have been transformed into

pictures using Convolutional Neural Networks (CNN). Two datasets, Shadow Socks-Regular and ISCX VPN-non-VPN, were used, each with different characteristics such as packet size and inter-arrival time sequences. In experiments, the technique showed great accuracy, precision, recall, and F1 scores, notably for VPN traffic identification, where the F1 score was 99.8% [99].

Chen Ling and Lai Lin (2023) the researchers looked at several DDoS assaults utilizing datasets from virtual machines on the CDX 3.0 cloud platform. It proposed a Poisson distribution-based detection technique and evaluated it against the Radial Basis Function Network, Support Vector Machine, Bagging, and J48 Decision Tree. The results revealed that the suggested technique outperformed other methods in terms of accuracy (96.13%) and false positive rate (0.005%) [100].

Al Taleb and Saqib Nazar (2022) The proposed study employs a hybrid deep learning model integrating 1D CNNs and QRNNs for cyber threat intelligence in smart cities, which is tested on the BoT- IoT and TON-IoT datasets. The model achieves exceptional accuracy, precision, recall, and F1 score, with an accuracy of 99.99%, precision and recall of 100%, and F1 score of 100% on both datasets, outperforming cutting-edge models and demonstrating its effectiveness in real-time threat detection with low false positive rates [69].

Dwairi et al. (2022) Dwairi and associates (2022) The suggested Self-Healing Version-Aware Ransomware Recovery Approach (SH-VARR) combines an access-control module with a decentralized version-aware control system to defend XML-based files from ransomware attacks. The SH-VARR framework was evaluated using zip, bzip2, and zip compression on a dataset 500.odt files. The findings demonstrated the effectiveness of the zip approach in ransomware recovery and file protection by demonstrating that it achieved the best balance of performance characteristics, including storage overhead, time required, CPU use, and memory consumption [101].

Vitorino et al. (2022) This study evaluates A2PM, a technique for creating realistic hostile cases in cybersecurity, using the CIC-IDS2017 and IoT-23 datasets. A2PM modifies attributes, including connection flags, packet inter-arrival time, and flow duration, in order to evade detection. It performs better than rival methods like One-Pixel and JSMA. The findings demonstrate that models—especially RF models—trained using A2PM remain more resilient to adversarial attacks in terms of accuracy, precision, recall, and F1-score [102].

Li Yanmiao et al (2022) Yanmiao Li and others (2022) Three datasets were used in the proposed work: MAWI Lab, CIC- IDS 2017, and NSL-KDD. It employed a number of strategies, such as deep baseline models (OC-NN, OC-LSTM, and DCAE) and short baseline models (OC-SVM/SVDD). With average accuracy, precision, recall, and F1-scores of 96.72%, 96.72%, 96.72%, and 96.72%, respectively, the findings showed that OC-LSTM outperformed the other models [68].

Y. Niu et al. (2022) The suggested method utilizes datasets such as KDD99, NSL-KDD, UNSW-NB15, and CSE-CIC- IDS2018 to generate multi-granularity features for intrusion detection. A number of classifiers were employed, such as XGBoost, Random Forests, Bagging, and decision trees. The KDD99 and NSL-KDD dataset's two and five-class classification results show 100/ accuracy, precision, recall, and F1-score. Additionally, using the CSE-CIC-IDS2018, multiclass investigations achieved 100% detection accuracy [103].

Al Abassi et al. (2020) The proposed study employs two real-world datasets, gas pipelines and secure water treatment, to identify cyberattacks in Industrial Control Systems (ICS). Using stacked autoencoders (SAE), deep neural networks (DNN), and decision tree (DT) classifiers, it employs an ensemble deep learning approach. With accuracy of 95.86% for Gas Pipeline and 99.67% for SWaT, precision of 98.21% for Gas Pipeline and 99.95% for SWaT, recall of 97.54% for Gas Pipeline and 99.53% for SWaT, and F1-scores of 97.87% for Gas Pipeline and 99.74% for SWaT, the findings show a significant performance improvement [104].

Choi et al. (2020) provide multi-agent-based cyber-attack detection and mitigation methods for Distribution Automation Systems (DAS) using power system domain analysis and authentication techniques. The techniques, which have been tested on a distribution system like KEPCO's, are intended to enhance cybersecurity by detecting and reducing risks, including denial-of-service attacks, configuration change, remote switch control, and man-in-the-middle attacks [105]. In order to develop standards and criteria for evaluating critical infrastructure resilience, Osri-Kyei et al. (2023) conducted a comprehensive review of the literature from 1990 to 2020. 44 articles were chosen after a three-step review process. Six categories were created from the defined standards and criteria: technical, social, management/organizational, safety, sustainability, and financial. The study's conclusions show that the proposed paradigm may impact CI resilience assessment research and practice in the future [71].

With the help of the above literature, the researcher can understand traditional ML methodology to DL approaches and their impacts on our digital era world by enhancing the performance and techniques from time to time from 2005 to 2023. With this, researchers also know about the current trends and new attacks on networks from intruders and solutions for them. Literature review plays a vital role in understanding the workflow of intrusion detection systems for critical information infrastructure.

Table 1. Comparison of studies in the literature

References	Datasets	Method	Models	Result (%)
Naseer et al. (2018)	NSL-KDD	ML/DL	KNN, DT,CNN, LSTM	
R. Vinaya Kumar et al. (2019) [47]	Virtual box, cuckoo sand-box	DL	Hybrid	Achieved detection Zero-day malware
Liu Lan et al. (2020)	NSL-KDD, CSE-CIC- IDS2018	DL	DSSTE Alexet, DSSE Mini-VGGet	Acc. 82, Pre. 83, Recall 82, F1 score 81 Acc. 96, Pre. 97, Recall 96, F1 score 97
Derhab et al. (2020)	BoT-IoT	DL	TCNN, SMOTE-C	Acc. 99, Pre. 99, Recall 97, F1 score 98
Al Abassi et al. (2020)	Real – GP, Real – SwaT	ML/DL	SAE, DNN, DT	Acc. 95.86, Pre. 98.21, Recall 97.54, F1 score 97.87 Acc. 96.67, Pre. 99.95, Recall 99.53, F1 score 99.74
Ahmad et al. (2020)	KDDCup99, NSL-KDD	ML/DL		Challenge – Imbalanced datasets and real-world performance
Susilo Et al. (2020)	Real time	ML/DL		Improve accuracy
Guangming Xian et al. (2020)	KDDCup99, NSL-KDD	DL	DBN	Improve performance and time Decreases as the number of hidden layer increases
Derhab et al. (2020)	BoT-IoT	DL	Hybrid (CNN+LSTM)	Accuracy 99.99
Abbas et al. (2021)	CICIDS2017	ML	Ensemble	Accuracy for multi-class 93 and Binary class 88.92
Ullah et al. (2021)	BoT-IoT, MQTT- IoT-IDS20, IoT-23	DL	CNN (1D,2D, 3D)	Achieved improved accuracy, recall, f1-score, and precision
Alkahtani et al. [48] (2021)	IoTIDS20	DL	CNN, LSTM, CNN+LSTM	Accuracy 96.60, 99.82, 98.80
Fatani et al. (2021)	KDDCup99, NSL-KDD, BoT-IoT, CICIDS2017	DL	CNN, TSODE	Improve accuracy
Idrissi et al. (2021)	BoT-IoT	DL	CNN	Accuracy 99.94, execution time less than 0.34ms
Disha and Waheed	UNSW- NB 15(20),	ML	GIWRF-DT	Acc. 93.01, Recall 94.76, F1 score 93.72 Acc. 99.90, Recall 99.87, F1 score 99.85
Gehlot et al. (2022)	Real-Time	DL	RNN, LSTM	Improvement of response time and accuracy
He Yanjie and Li Wei (2022)	Shadow-Socks, ISCX VPN- non-VPN	DL	CNN	Acc. 99, Pre. 99, Recall 100, F1 score 99
Dwairi et al (2022)	500 .odt file	SH-VARR	Zip, gzip, bzip2	
Li yanmiao et al. (2022)	NSL-KDD, CICIDS2017, MAWILab	ML/DL	OC-LSTM	Acc. 96.86, Pre. 96.72, Recall 96.86, F1 score 96.72
Y. Niu et al. (2022)	KDD99, NSL-KDD, UNSW NB15, CSE-CIC-IDS2018	ML	Multi- Granularity Feature generation, DT, RF, XGBoost, Bagging	Acc. 100, Pre. 100, Recall 100, F1 score 100
Al Taleb and Saqib Nazar (2022)	TON-IoT, BoT-IoT	Hybrid	CNN+QRNN	Acc. 99, Pre. 100, Recall 100, F1 score 100

Wu Zihan et al. (2022)	CICIDS2017, CICIDS2019	AI/DL	Fuzzy LSTM	F1-Score 99.17, 98.48
Bar Rotem et al. (2022)	USTC-TFC2016, ISCXVPN2016	NLP	SimCSE	Accuracy 99
Vitorino et al. (2022)	CICIDS2017, IoT-23		JSMA, OnePixel	High accuracy
Sogut et al. (2023)	Their own dataset (Testbed 15)	DL/ML	CNN, LSTM, CNN+LSTM, DT, RF	Acc. 93.54, Pre. 94., Recall 93.53, F1 score 93.57 Acc. 84.60, Pre. 86.03, Recall 84.60, F1 score 83.73 Acc. 94.73, Pre. 94.90, Recall 94.73, F1 score 94.74 Acc. 98.77, Pre. 98.77, Recall 98.77, F1 score 98.77 Acc. 95.84, Pre. 97.21, Recall 95.84, F1 score 96.51
Sadhwani et al. (2023)	BoT-IoT (15), TON-IoT (15)	ML	NB, RF	Acc. 77, Pre. 100, Recall 100, F1 score 100 Acc. 100, Pre. 100, Recall 100, F1 score 100
Adejimi et al. (2023)	CICIDS2017	ML	PCA, m-SVM	Acc. 97.64, Pre. 99, Recall 98, F1 score 98
Alqudhaibi et al. (2023)		ML	Linear, LR, DT classifier, Poly	Acc. 66.25, Pre. 18.5, Recall 36, F1 score 13 Acc. 58.5, Pre. 15.6, Recall 12.4, F1 score 10.4 Acc. 60, Pre. 46.5, Recall 41.1, F1 score 41.7 Acc. 65, Pre. 7.2, Recall 11.1, F1 score 8.7
Park et al. (2023) [95]	CAR- HACKING-Attack and Defense, Real- Datasets	G-IDCS, ML	GAN	Acc. 98, Pre. 98, Recall 96, F1 score 98 Acc. 90, Recall 86, F1 score 88
Villegas et al. (2023) [64]	Real-Datasets	ML	Autoencoder VAE	Acc. 88, Recall 84, F1 score 86
Salma et al. (2023)	CICIDS2017, KDDCup199	DL	Transformer Model, CNN, RNN	Acc. 96, Pre. 94, Recall 94, F1 score 94 Acc. 94, Pre. 92, Recall 91, F1 score 92 Acc. 95, Pre. 93, Recall 92, F1 score 93
I. Sharafaldin et al. (2023)	CICIDS2019	ML	ID3, RF, Naïve Bayes, Logistic Regression	Pre. 78, Recall 65, F1 score 69 Pre. 77, Recall 56, F1 score 62 Pre. 41, Recall 11, F1 score 5 Pre. 25, Recall 2, F1 score 4 Acc. 95, Pre. 93, Recall 92, F1 score 93
Chen Ling and Lai Lin (2023)	CDX (Simu-lation)		Poisson Distribution model	Accuracy 96.13
Alghamdi et al. (2023)	Publicly available	Ensemble	Lambdatecture LSTM, ANN Archi-with CNN,	Achieved high accuracy and less processing time
Wang et al. (2023)	Publicly available	DL	Transfer Learning	False Positive reduced
Alnifie et al. (2023)		Meta-Analysis		Huge impact on cybersecurity

Henry et al. (2023)	Publicly available	DL	CNN+GRU	Accuracy 98.73 but low FP rate

Table 2. Datasets and deep learning techniques used

Dataset	Deep Learning Technique Used	Advantages	Disadvantages
NSL-KDD [106]	CNN, LSTM RNN,	Comprehensive and Used widely	Contains outdated and unrealistic data
CICIDS2017	CNN, LSTM, DNN	Represents real-world network traffic	Large and complex, it may require significant resources
UNSW-NB15	DNN, GAN	Contains diverse attack scenarios network	Limited labeled data for certain attack types
KDD Cup 1999	RNN, LSTM	Historical benchmark dataset	Focuses on older network environment
DARPA	CNN, RNN	The early dataset used for intrusion detection research	Limited in representing contemporary threats
AWID	CNN, LSTM	Wireless network dataset	Limited coverage of wired network scenarios
Kyoto2006+	LSTM, RNN	Focuses on HTTP traffic logs	Limited types diversity in attack
ADFA-LD	CNN, DNN	Specifically designed for anomaly detection	Limited scale, not suitable for large-scale testing
CICIDS2018	RF, LSTM, AlexNet, mini-VGGNet	For real attacks, improvement of CICIDS2017	Imbalance dataset
CICIDS2019	ID3, RF, Multinomial LR	It is for DDoS attacks in real environment	

Table 3. Comparison of various datasets used in previous work (RQ4)

Datasets (Features)	Attacks	Benign	Malware	Techniques	Training Time (MS)	Predict Time (MS)
Testbed (25)	DDoS	3391	19377	DT, RF, LSTM, CNN, Hybrid		
CICIDS2017+KDDCup1999 (49+79)	Web-based					
CICIDS2017(83) [A.O. adejimi]	attacks	2359087	471688	m-SVM		
CICIDS2018()						
CICIDS2019()						
BOT-IOT (15)	All attack	733,603	101	NB	0.2776	0.03385
2-7				RF	56.404	0.795
2-7	With DDoS	385,326	95	NB	0.3554	0.0334
2-7				RF	14.3046	0.3173
TON-IOT (15)	All attack	59,925	32,284	NB	2.9326	0.3043
2-7				RF	351.873	2.9125
2-7	With DDoS	60,055	3945	NB	1.5244	0.1685
2-7				RF	194.65	1.6478

3. Research Methodology

To locate, assess, and understand diverse research relevant to specific study topics, authors incorporate methods from earlier articles into our Systematic Literature Review and divide the review into three levels, as depicted in figure 5.

Review the Planning: This level's objectives are:

1. To determine the necessity.
2. Create standards and processes. Assess the standards and processes in relation to this SLR.

Reporting the Analysis: Researchers report the results in an appropriate format to the intended audience and distribution strategy once authors have finished reviewing every study.

3.1. Research Questions RQs

Selecting research questions is the starting phase in determining a given study's general goal and anticipated results. Because a carefully designed question increases trust in a topic, and formulate our research questions with researchers in mind. In order to acknowledge the latest work in the field of Critical Information Infrastructure, and outlined four key questions (RQ1-4) in addition to a few supplemental questions (SRQs), which are displayed in Table 1.

In order to follow internal details, explore even further by giving some more questions (SRQ-2.1 to SRQ-2.4). These details include Outlining the datasets that were used in the research. Algorithms or models for CII IDS detection. Measuring measures that evaluate how well different methods identify thesekinds of attacks. Next, the overall effectiveness of several approaches in RQ-5 will be compared usinga range of measuring criteria. Lastly, the same dataset and measurement metric will be used to compare models regarding efficiency. All research questions and their objectives are discussed in Table 4 as follows:

- Outlining the datasets that were used in the research.
- Algorithms or models for CII IDS detection.

Measuring measures that evaluate howwell different methods identify thesekinds of attacks. Next, the overall effectiveness of several approaches in RQ-5 using a range of measuring criteria is compared. Lastly, the authors use the same dataset and measurement metric to compare models in terms of efficiency. All research questions and their objectives are discussed in Table 4.

3.1.1. Search Strategy (SS)

The author's goal was to collect as numerous related works as possible for study questions. To avoid bias, the author tried to contain every probable blend of relevant search terms or phrases when gathering data for CII IDS research. To guarantee an accurate search, examined multiple repositories rather than depending just on one or two. On the other hand, finding research publications was possible through various digital sources. From them, we chose ten well-known sources based on accessibility and relevance, which are stated below: Web of Science, IEEE Xplore Digital Library, ACM Digital Library, ScienceDirect (ELSEVIER), SpringerLink, Google Scholar, Semantic Scholar, Cornell University, Computing Research Repository, Database Systems and Logic Programming (DBLP), Google Scholar Archives, conferences, and periodicals are among the repositories. Authors restrict the amount of time researchers spend in searching [54].

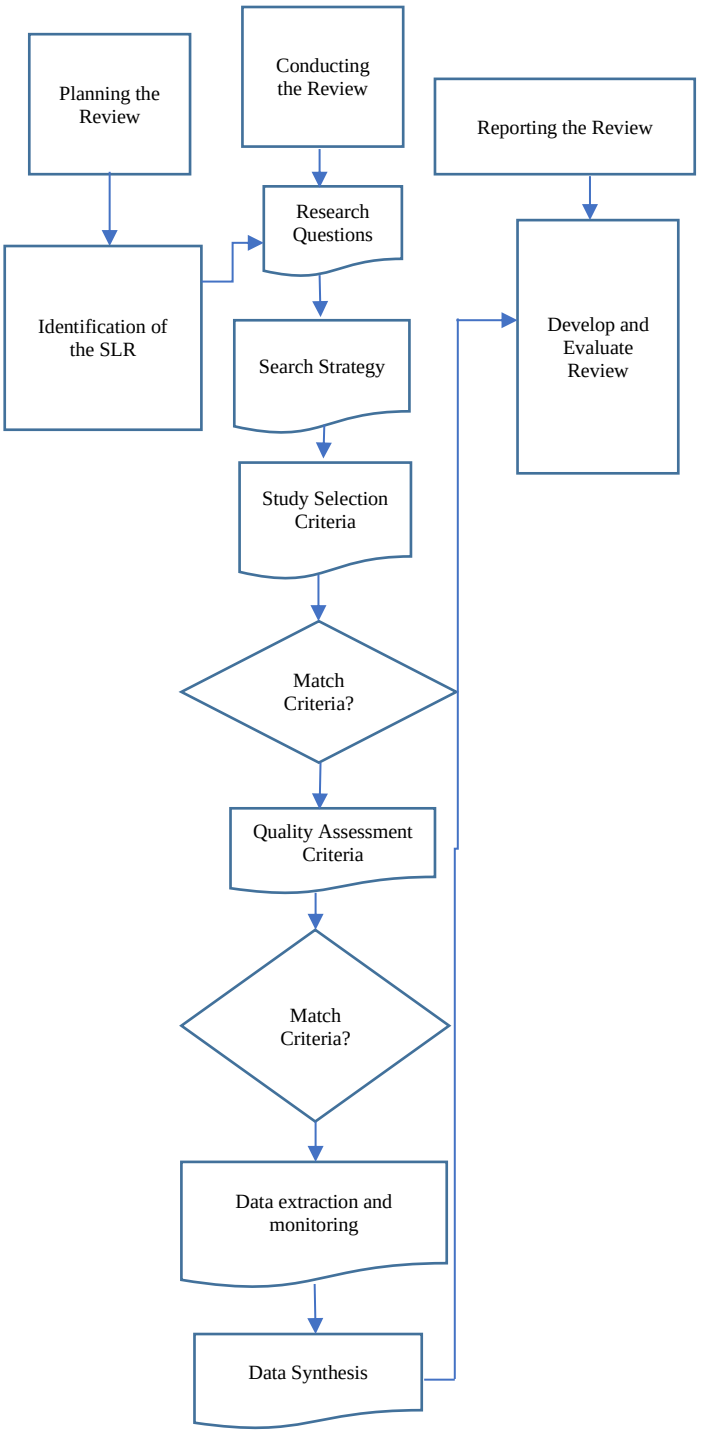

Fig. 5 The process of the SLR

3.1.2. Search Process Study Selection Criteria

To choose the right articles from among these ten digital repositories, and set up three inclusion criteria in our search process. The search terms are included in the keywords in the abstract title. The author explores different parts of the literature to determine the main points that need to be considered. If the author goes through such works, then include them.

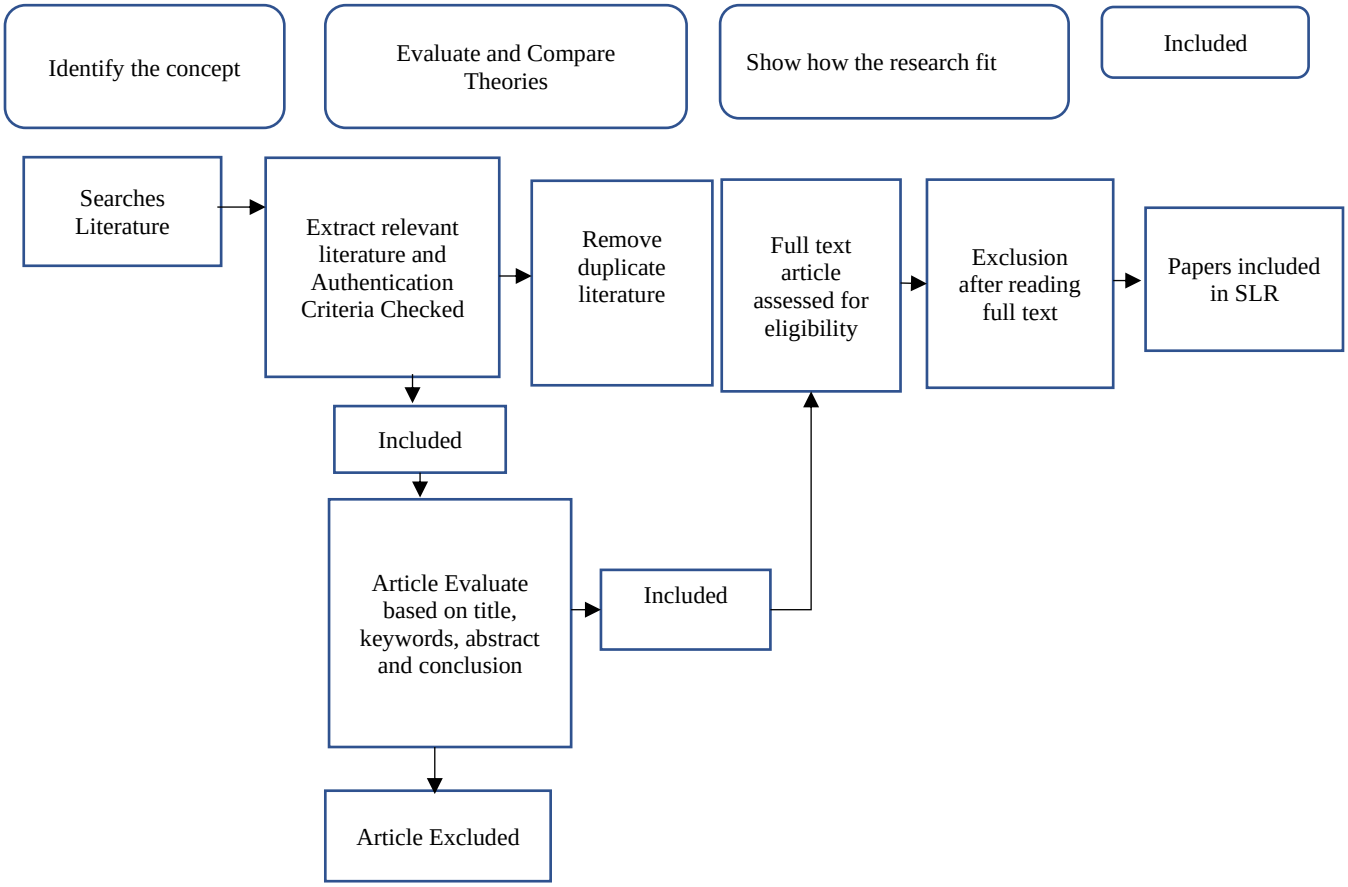

Fig. 6 Flowchart for search process

Table 4. Research questions and objectives

QID	Research Questions	Aims/Objectives
RQ1	What kinds of critical infrastructures are currently in place?	Describe the importance of critical infrastructures and their different sec-tors.
RQ2	What methods were used to find IDS for CII?	Describe various techniques of IDS for CII.
RQ3	Which preprocessing and feature engineering methods are applied to IDS?	Describe different techniques and processes of Feature engineering and pre-processing steps.
SRQ-2.1	Algorithms detection. Or models for CII IDS	Detailed models used for CII IDS in previous studies.
SRQ-2.2	Measuring measures that evaluate how well, different methods identify these kinds of attacks.	Recognize the measurement metrics that are generally used to assess performances.
SRQ-2.3	Outlining the datasets that were used in the research.	Describe the strengths and limitations of various datasets.
RQ4	What features of the dataset were used by IDS?	Describe various datasets that were used in research.
RQ5	Overall effectiveness of several approaches using a range of measuring criteria?	Evaluate the efficiency of many IDS methods for CII.

Table 5. Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
Indexed papers of Web of Science or Scopus	Research papers that are not related to our research problems.
Research papers from journals and Conferences	Research papers that are not included in good or related journals
Paper updated in English	Papers in which the absence of DOI (Digital Object Identifier)

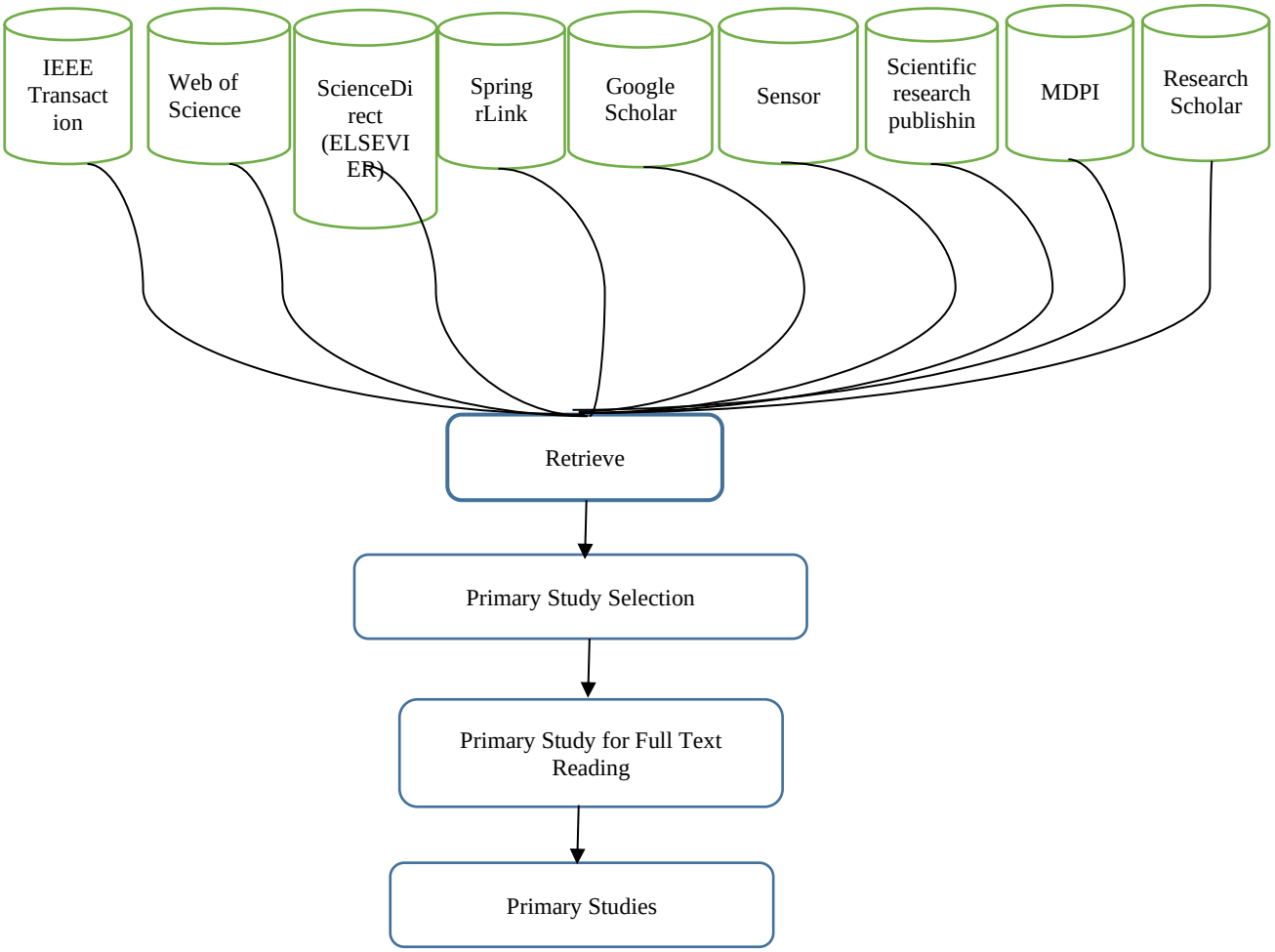

Fig. 7 Study selection process

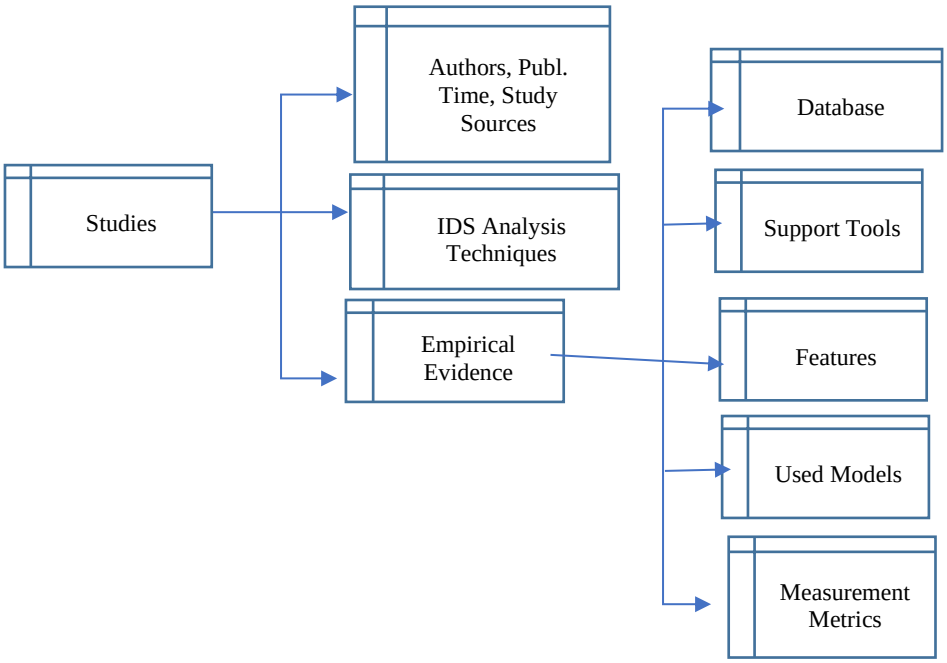

Fig. 8 The information of the extracted data

3.1.3. Quality Assessment Criteria

It is just as crucial to assess the superiority of the evidence as it is to analyze the facts in an SLR. Interpreting the results of a poorly executed study should be done so with caution because preconceptions in the research method may disturb the outcomes. These revisions should be either lost from the organized review or recognized as such. It's also critical to select accurate standards to evaluate the superiority of the suggestion and any ingrained preconceptions in individual studies. Based on the specified criterion, apply the requirements to these studies and validate the chosen ones based on these criteria. Following this round of quality assessment, and examined 90 research papers and 19 more reviews.

3.1.4. Data Monitoring and Extraction

This step involves creating mechanisms to extract data from research. To identify potentially related studies, and extensively searched nine popular libraries (Figure 7).

- *We selected articles that met the following necessities:* The approaches or outcomes segment specified the articles to be extracted.
- *Researchers publishing sources and time of publication:* This section includes author information, publication period, and publication type (conference, journal or workshop).
- *Analysis Methods:* This study recognized feature-based algorithms for detecting attacks.
- *Empirical Evidence:* This section focuses on four components: The study authors utilize datasets, analyze characteristics, use models or techniques, and assess findings using measurement measures.

3.1.5. Data Synthesis (DS)

During the data synthesis step, findings from the extraction process are reviewed and compared to support detailed replies to RQs. After collecting data, we analyze and visualize it using tools like histograms, pie charts, and tables.

3.2. Background

ML/DL, a powerful technology, has made significant contributions across various fields, revolutionizing how authors approach problems. One notable impact is in modern security systems, which is crucial for organizations of all sizes [55]. Organizations face constant threats from millions of new malware and viruses in today's digital landscape. Even large enterprises, like banks and government agencies, are not immune to attacks that target technological flaws. Although security solutions already exist, the field is constantly evolving [56]. Deep learning [57] has helped to improve cybersecurity operations. It accomplishes this by detecting and stopping network assaults, removing dangerous software, highlighting system flaws, and generally maintaining the security of digital settings. This novel technique has opened up new paths for cybersecurity research, addressing businesses' continuous issues in protecting their digital infrastructure [58]. Among the strategies presented, the deep learning method is the most popular for malware detection, closely followed by Recurrent Neural Networks (RNNs), which are used for malware detection and identifying information security concerns. While machine learning is famed for its simplicity, its use in security research remains limited [59]. However, determining the absolute effectiveness of these strategies is difficult due to differences in datasets and metrics across safekeeping domains. The information security area has a large range of data from many sources, which presents obstacles to thorough Deep Learning testing. Existing research studies have limitations, notably in terms of dataset availability, which is typically tiny and out of date [19]. For the development of meaningful security solutions, thorough testing on large, up-to-date, and reliable datasets is crucial. Results obtained from these methods should undergo comparisons in real-time scenarios to better understand their effectiveness, particularly in the context of critical information infrastructure.

3.3. Processing Datasets for IDS(RQ3)

Table 6 compares several data pre-processing approaches used in different research articles. Each study provides strategies for dealing with missing values, categorical data, data standardization or normalization, and feature selection or extraction [60]. For example, Sadhwani et al. from 2023 recommend utilizing the mean approach for missing values, cross-entropy for categorical data, the pinging method for standardization, and SMOTE for feature selection.

Meanwhile, from 2023 recommends using RMSProp for missing values, one-hot encoding for categorical data, a standard scaler or min-max scaler for normalization, and the Extra-Tree Classifier for feature selection. Each work takes a distinct strategy to address duplicate entries, positional encoding, or feature selection using techniques such as PCA or mutual information criteria. Data pre-processing techniques are important for the good performance of models to extract the relevant features [62] that are important for our work or experiments [62]. With the help of these techniques, we can remove the non-relevant and duplicate features or fill in the missing values, eliminate other issues related to the data, and apply suitable techniques according to the demand of our work to achieve desirable results.

4. Deep Learning Model (SRQ 2.1)

Figure 10 discuss DL techniques from previous studies; DL approaches [76, 77] such as CNN, Generative Adversarial Networks, RNN, GRU, DNN [78] and LSTM focus on generating a hybrid model for detecting web-based attacks. Every model of DL has its own advantages in enhancing intrusion detection systems for critical information infrastructures in various sectors and networks. It also combines these advantages to improve the accuracy, performance, and robustness of the framework.

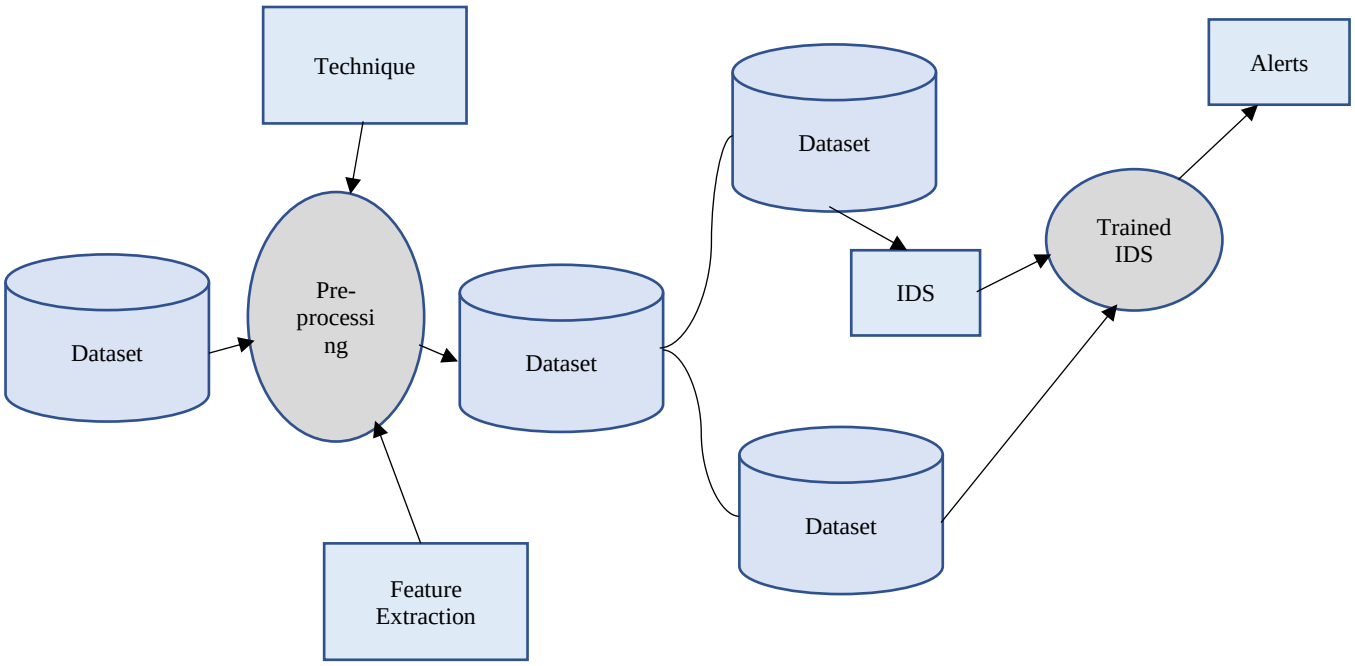

Fig. 9 Process of Pre-processing of datasets

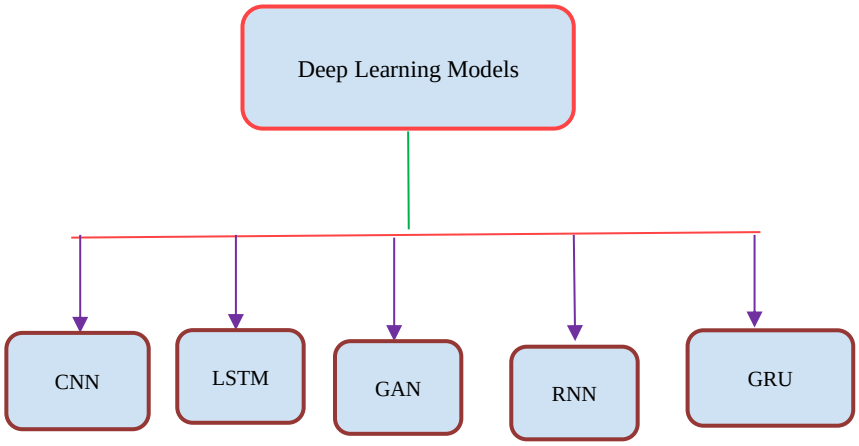

Fig. 10 Deep Learning Model

CNN has proven effective in detecting intrusion patterns within network traffic. By investing in their ability to automatically learn related features and classify data, CNN can detect and categorize shady activities [79]. RNN and LSTM are both analyses of sequential data, but the difference is that RNN is for a short time and is simple to implement and on the other hand, LSTM is for a long period of time and is complex in nature. The Generative Adversarial Network produces man-made data to mimic network behavior [77] [80, 81]. DL techniques supply possible resolutions to overcome these threats and challenges. These techniques are capable of extracting useful and meaningful features from big-scale data. CNN, LSTM, RNN [82], GRU models and ensemble techniques [83] are already used for effective performance and detecting threats to mitigate them [25]. That's why further research is important to improve their real-world performance. However, further research is essential to refine these methods and improve their real-world performance. By using deep learning, IDS can accommodate and guarantee the security of system integrity and data confidentiality [2].

The effectiveness of each method can vary based on the precise features of the data and the nature of the intrusion detection task. The table provides a general overview of the advantages and disadvantages of each method in the context of IDS for Critical Information Infrastructure [31][84]. In Table 7, we discussed about the few major pros and limitations of the ML and DL techniques, with some mostly used methods or techniques of ML and DL. With the help of this table, researchers know about the weaknesses and strengths of each machine technique and deep learning, which helps them choose methods for achieving objectives.

Table 6. Comparison of existing data pre-processing techniques

Paper	Year	Missing Values	Categorical Data	Standardization/ Normalization	Feature Selection/Extraction
[56]	2023	Mean Method, Cross-entropy	Pinging Method, SMOTE, Chi-Square Test		
[58]	2023	RMSProp, One Hot	Label Encoding, Standard Scaler, Min-Max Scalar	Extra-TreeClassifier (15)	
[63]	2020	Cross-entropy	Standardization and batch normalization, dropout	SMOTE-NC	
[64]	2023	Mean		PCA	
[65]	2023	Removing duplicate entries	Positional encoding, scaled dotted product	Min-Max Scaling	PCA/mutual Information Criterion
[59]	2023	Random Oversampling	Regularize (overfit-ting)		PCA
[66]	2022	Label Encoding	Normalization-Min-Max scaling		GIWRF
[67]	2023	Filter method (Pearson's correlation coefficient)	Cross Entropy		
[68]	2022	Neuralization	One-hot encoding	Min-Max Scaling	DCAE and OC-LSTM
[69]	2022	Filter method	Label encoding	Scaling	CNN and QRNN
[70]	2020	Entropy	Normalization	Information gain	
[71]	2023	Correlation and Classifier subset evaluator	One Rule and REP-Tree	PCA and Supervised Dis-cartelization	
[72]	2022	Imputation, Removal and KNN	One hot, label and target encoding and embedding	PCA, AE, LDA, DFF, CNN and RNN	
[73]	2020	Filter based		Normalization, Min-Max Scaling and Z-score Stan-dardization	XGBoost
[76]	2023	SMOTE for imbalance dataset		Normalization	VGG-16 Transfer learning
[75]	2022	PSO		Normalization	ELM, GWA

Table 7. Countermeasure method, advantages, limitations

Countermeasure Method	Advantages	Limitations
CNN (Convolutional Neural Network)	- Effective in image-based intrusion detection. - Automatic feature extraction.	May require large amounts of labeled data for training. Computationally intensive.
RNN (Recurrent Neural Network)	- Suitable for sequence data capturing temporal dependencies. - Handles variable-length input.	- Can be prone to vanishing/exploding gradient problems. - Computationally demanding.
LSTM (Long Short-Term Memory)	-Overcomes vanishing/exploding gradient issues in RNNs. - Effective in capturing long-term dependencies.	Still computationally intensive. Requires careful tuning of hyperparameters.
RF (Random Forest)	- Non-parametric, handles non-linear relationships well. - Robust to overfit- ting.	- Lack of interpretability for individual decision trees. - Might require careful tuning of hyperparameters.

DT (Decision Tree)	- Simple to understand and interpret. - Does not require extensive data pre-processing.	- Prone to over-fitting. - Sensitive to small changes in the data.
Hybrid Models	- Combines strengths of different models. - Improved performance due to diversity.	- Complexity in integrating multiple models. - Requires careful selection of component models.
Ensemble Methods	- Model-independent, good generalization. - Reduces overfit-ting and increases robustness.	- May have increased computation overhead. - Selection and combination of base models require attention.
GAN	Can generate realistic data for training	Training can be unstable and may not always converge

In standard IDS workflow for ML [107] and DL, firstly collect the relevant data from publicly available datasets or real data with the help of simulation on real network traffic; after that, data preprocessing techniques are used to clean, remove missing values, transform and then prepare the data for training, followed by feature selection and feature extraction techniques for extracting the necessary information to achieve desired result according to individual objectives and then spilt the data into training as well as test data after that train the model with training data and, adjusting its parameters iteratively to minimize error. Once the model is trained, the model is prepared for testing on unseen data to analyze its performance [24].

F1-Score are calculated to assess the effectiveness of the model in making predictions. All these steps ensure that ML/DL models learn from data effectively and produce accurate results for the given task. How well a classifier recognizes all real positives when authors measure recall while precision detects the accuracy of its positive predictions. However, these measures are balanced by the F1-Score, which offers a single value to the device and determines the holistic effectiveness of the classifiers. Although accuracy provides a holistic overview, it may omit some performance components, and both the measures, i.e., precision and recall, help to provide an overall evaluation of the effectiveness of a classifier (SRQ 2.2).

$$Recall = \frac{TP}{TP+FN} \quad (1)$$

Correctly identified the proportion of real positives and evaluated by the recall while accuracy is assessed by the precision(P) for positive predictions among all the cases classified as positive, and these metrices provide an understanding of the model's performance on positive classes.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

After that, F1-Score combines the performance of accuracy and recall into one unit and then reflects a balancing of the classifier performance measure because it helps to recall and precision both are considered equally, providing an overall assessment of the model's accuracy on the test data.

$$F1 - Score = \frac{2*(P*R)}{P+R} \quad (3)$$

Symbols are indicated as follows:

TP- True Positive, TN- True Negative, FP- False Positive, FN- False Negative, P for precision, and R for recall. Deep Learning has become an effective tool in the field of security, outperforming conventional methods and conventional Machine Learning algorithms. The survey presented in Table 2 highlights various Deep Learning [108] papers that focus on solving security problems. Notably, the majority of researchers have concentrated their efforts on malware detection and intrusion detection. This survey also points to promising applications in health security and vehicle security, expanding the horizons of Deep Learning.

The BOT-IOT dataset, containing 15 features, was utilized for various attack scenarios, including all attacks and those with DDoS incidents. Different machine learning techniques, such as Naive Bayes (NB) and Random Forests (RF), were applied, with corresponding training and prediction times measured [92]. Similarly, the TON-IOT dataset, which also had 15 features, was employed for all attacks and those involving DDoS. The dataset was evaluated using NB and RF techniques, with associated training and prediction times documented [93]. In summary, these studies assessed the performance of machine learning techniques on diverse datasets, emphasizing the importance of understanding the intricacies of each dataset and selecting appropriate techniques based on the characteristics of the data and the nature of the attacks. The provided training and prediction times offer insights into the computational efficiency of the applied methods.

4.1. Dataset Used (SRQ 2.3)

Datasets commonly used for different Deep Learning techniques in Intrusion Detection Systems (IDS) for Critical Information Infrastructure, along with their advantages and disadvantages [27]. The pros and cons stated here are overall all characteristics and may differ based on definite use cases and research goals. The suitability of a dataset depends on the context, research objectives, and the nature of the Deep Learning techniques being employed. In Table 3, the comparison of various datasets used in previous works for intrusion detection, several datasets [109] were evaluated based on their features, types of attacks, and the performance of different ML techniques. One such testbed with 25 features focused on DDoS attacks showcasing the application of Decision Trees (DT), Random Forests (RF), LSTM, CNN, and a hybrid method. The training and prediction times for

each technique were measured in milliseconds. The CICIDS2017[110] and KDDCup1999 datasets, both with a combination of 49 and 79 features, were used for evaluating web-based attacks. Additionally, CICIDS2018, a highly imbalanced dataset with normal traffic data distribution [65] and CICIDS2019 extract 80 features from the dataset using the CIC-Flow Meter and then choose the best detection feature set for each DDoS attack [67].

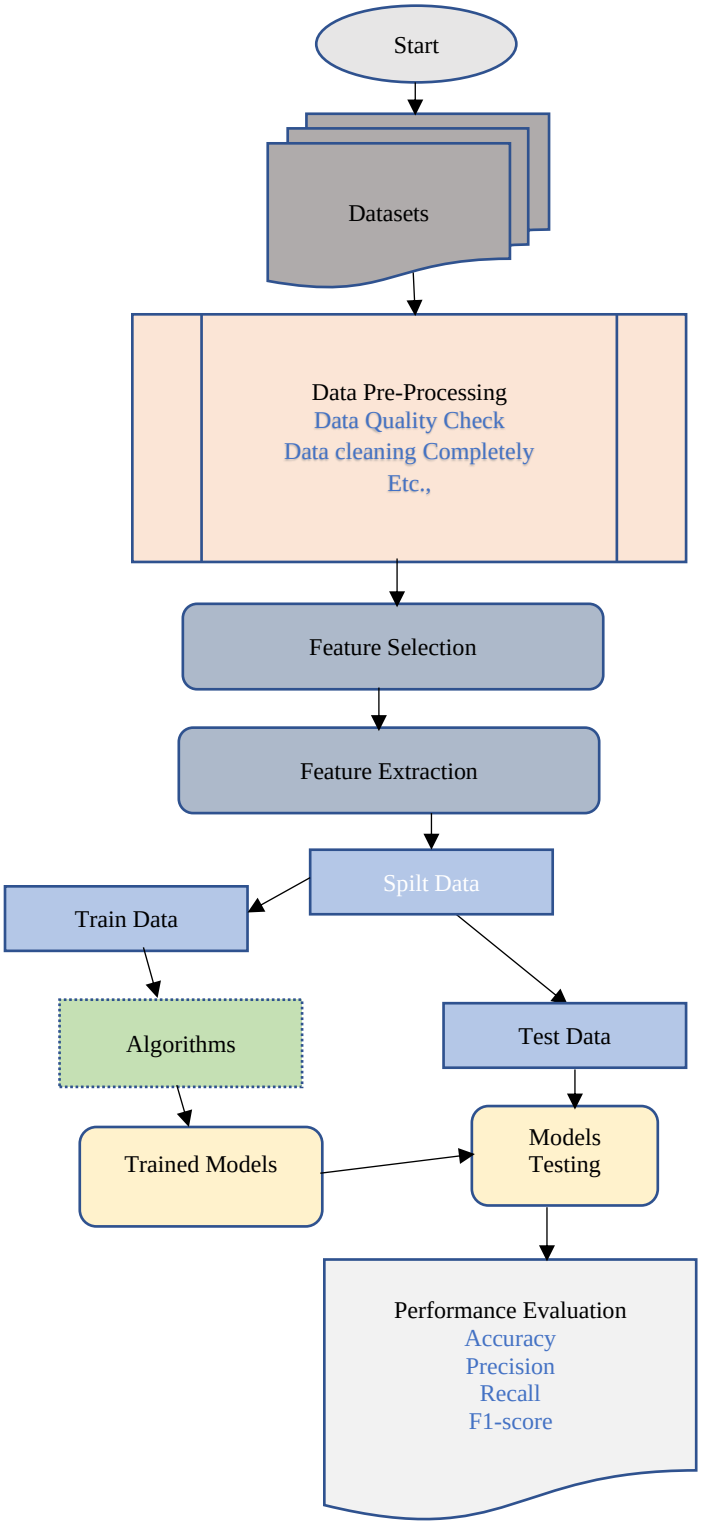

Fig. 11 Standard workflow of IDS

5. Demonstration of the Most Efficient ML and DL-based

5.1. Techniques on a Synthetic Dataset

According to the different research, works analyzed earlier during the review of existing literature in the relevant domain, and demonstrated the synthetic data which was integration of a portion of the DDoS, SQL Injection dataset along with ransomware attack data, merged and processed these datasets using six of the best Machine Learning (ML) and Deep Learning (DL) techniques figure out most efficient methods among existing previous studies reviewed so far.

These six techniques are DT, RF, SVM, CNN, LSTM and GRU. Using these techniques, the dataset was trained and tested, and their performances were compared. Our findings discovered that the performance of each model depends on the size and nature of the data and which type of data we were using. When the dataset was in tabular form, where DT, RF, and SVM performed well, on the other hand, CNN is suited for image-based datasets, while GRU (less complex than LSTM) and LSTM are better appropriate for sequential or time-series data.

For ransomware, data is also synthetically generated with DDoS and SQL Injection for training and testing the data and finding the performance of every model.

In our demonstration, when using small data sample (n=10K) in Figure 1,

Generate a synthetic dataset with valid parameters (0: DDoS, 1: Ransomware, 2: SQL Injection)

X, y = make_classification(n_samples=10000, n_features=20, n_classes=3, n_clusters_per_class=1, n_informative=10)

In Figure 12, if the dataset is small, then machine learning techniques outperform like in Figure1 and also the deep learning approaches that are used here. However, increasing the dataset size to n=1-lakh in Figures 13 and 14 gradually increases deep learning performance.

Generate a synthetic dataset with valid parameters (0: DDoS, 1: Ransomware, 2: SQL Injection)

X, y = make_classification(n_samples=100000, n_features=20, n_classes=3, n_clusters_per_class=1, n_informative=10)

After that, in Figure 13, the DL techniques in progress create results nearly equal to the ML techniques used here. This indicates that deep learning models perform better with big/large datasets. But, due to the limitations of the hardware that we used, we were incapable of implementing larger datasets on our system. However, our examination proves that DL models improve their performance as we increase the size of the dataset, while ML models tend to perform better with smaller datasets. And also use a balanced (random oversampling) technique to balance the unbalanced data to accurate the result. For all techniques applied, the author gets results in graph format for accuracies and time taken by each technique.

Finally, in Figure 14, take n= 10-lakhs samples of the dataset, and deep learning models outperformed than machine learning, but execution time will increase with the increase in sample size; the result is shown below:

Generate a synthetic dataset with valid parameters (0: DDoS, 1: Ransomware, 2: SQL Injection)

X, y = make_classification(n_samples=1000000, n_features=20, n_classes=3, n_clusters_per_class=1, n_informative=10)

Table 8. Models' accuracy & time

N=1000000		
Models	**Accuracy**	**Time (s)**
0 Decision Tree	0.934979	4.029444
1 Random Forest	0.971388	43.840735
2 SVM	0.980256	7008.604629
3 CNN	0.988903	609.953117
4 LSTM	0.985454	1667.459045
5 GRU	0.985364	2854.611508

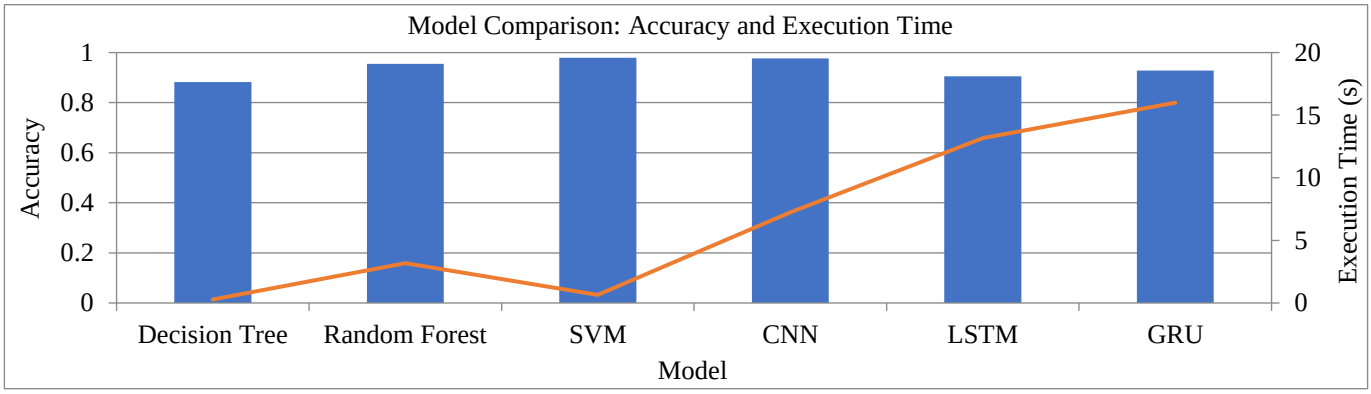

Fig. 12 Sample size, n= 10K

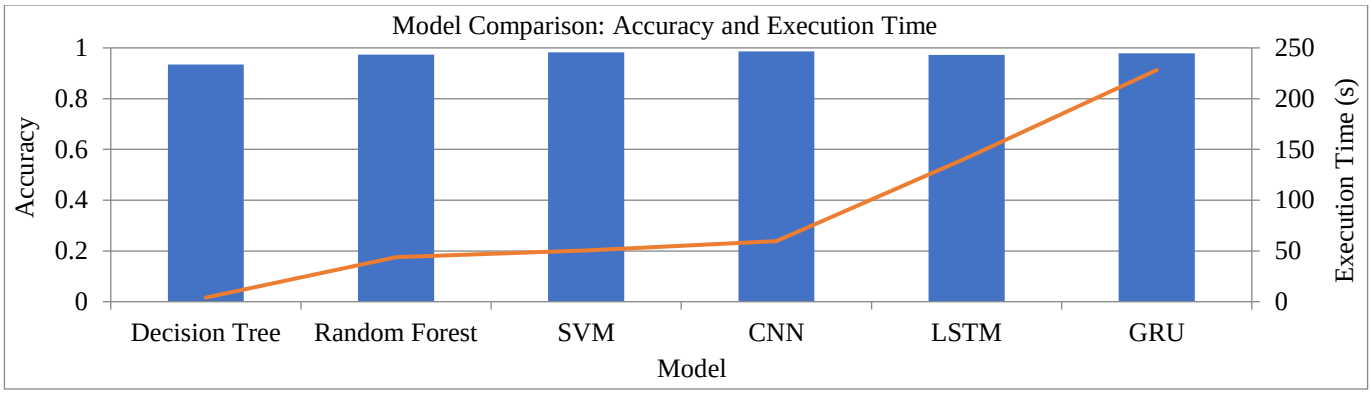

Fig. 13 Sample size, n=1-lakh

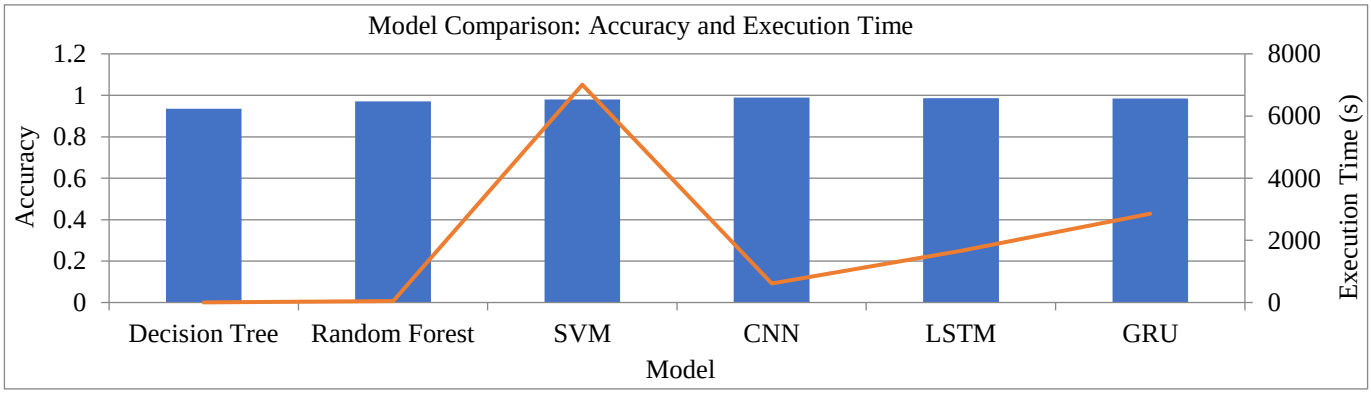

Fig. 14 Sample size, n=10-lakhs

Results Summary in Tables 1 and 2: Table Model, Accuracy, and Execution Time (s)

Table 9. Model accuracy & time with n=10-lakhs

N=100000			N=10000		
Models	**Accuracy**	**Time (s)**	**Models**	**Accuracy**	**Time**
0 Decision Tree	0.934659	4.029444	0 Decision Tree	0.882793	0.278580
1 Random Forest	0.974073	43.840735	1 Random Forest	0.954613	3.183135
2 SVM	0.981916	50.613571	2 SVM	0.979052	0.640612
3 CNN	0.986512	59.550375	3 CNN	0.976559	7.234769
4 LSTM	0.972225	141.539877	4 LSTM	0905237	13.160636
5 GRU	0.978669	228.168497	5 GRU	0928678	15.997882

6. Observations and Recommendations (RQ5)

In the realm of security, DL has appeared as a powerful tool, surpassing traditional techniques and classical Machine Learning [107] algorithms. The survey presented in Table 2 highlights various deep-learning papers that focus on solving security problems. Notably, most of the authors have concentrated their efforts on intrusion detection and malware detection. This review also points to promising applications in health security and vehicle security, expanding the horizons of Deep Learning. Among the techniques discussed, the Deep learning method stands out as a favorite for malware detection, followed closely by CNN and Recurrent Neural Networks (RNNs), which are applied not only for malware detection but also for identifying information security threats. While Machine learning is known for its simplicity, its application in security research appears less extensive.However, defining the absolute performance of these techniques proves challenging due to variations in datasets and metrics across different security areas. The information security area includes a wide variety of data from diverse sources, posing challenges for comprehensive Deep Learning testing. Existing research studies face limitations, particularly dataset availability, often being small and outdated.

For the development of meaningful security solutions, thorough testing on large, up-to-date, and reliable datasets is crucial. Results obtained from these methods should undergo comparisons in real-time scenarios to better understand their effectiveness, particularly in the context of critical information infrastructure.

7. Conclusion

With the fast growth of internet users in current times results from more instances of cyberthreats to critical information infrastructures. This SLR presented an in-depth analysis of several methods developed for the detection of attacks by intrusion detection systems on critical information infrastructures that were published in 108 papers during the period from the year 2005 to the end of 2024. This study also analyzes some widely used datasets that are publicly available and used in previous IDS models proposed by the researchers to test these models to enhance performance after applying optimized feature engineering. The dataset analysis investigates the performance dependence against the number of features used in feature engineering, followed by feature selection and feature extraction techniques. In this study, it is concluded that CICIDS2017, CICIDS2018, NSL-KDD, KDDCup99, BOT-IoT datasets are the most frequently used datasets for developing IDS for CIIs based on machine and deep learning techniques detecting attacks.

The broadly used metrics are accuracy, precision, recall and F1-score for performance measures of IDS for CIIs. The experimental results of previous studies demonstrate that deep learning-based approaches are effective for detecting attacks for IDS on critical infrastructures. Further, after analysing previous studies, this study concluded that CNN, LSTM, DT, RF, SVM and hybrid models hold significant percentages. LSTM reflects the most prominent results in terms of various performance parameters. To analyze the performance and demonstrate implementation of these leading ML and DL techniques on a synthetic dataset that contains DDoS, SQL Injection and ransomware attack instances that include 20 attributes, and compare the result in terms of accuracy and time taken by each technique. After analysis, we found that DL techniques (CNN, LSTM and GRU achieved accuracy of 98.8, 98.54 and 98.53, respectively; this result averaged from varying instances in a range of 10K to 10-lakhs) CNN outperformed with an increase in data samples, while on other hand, ML techniques (DT, RF and SVM achieved accuracy of 88, 95 and 97 respectively on 10K number of data sample) perform quite well on small dataset comparatively. Reviewing, demonstrating, and trying to provide a valuable roadmap for the researchers in developing effective intrusion detection methods and counter-measures to assist them in future research works.

7.1. Future Trends

Researchers can use advanced techniques like advanced autoencoder-based feature selection with a transformer model for IDS and GAN-based class balancing techniques to address imbalanced dataset challenges. Further, researchers can include explain-ability tools like SHAP and LIME to enhance the model transparency and more focused approaches for deeper insights.

Ethical Implications of Using Synthetic Data in Intrusion Detection

The author has done experiments on the artificial syntactic dataset that includes three attack classes: DDoS, Ransomware and SQL Injection by using SMOTE, control over class balancing that minimizes biases in the dataset, unlike real-world data that may be imbalanced. Hence, our code is ethically acceptable because, in this research, it is clearly stated that the dataset is completely synthetic and apply pre-processing or SMOTE for balancing to overcome the limitations of the dataset and to validate models on publicly available datasets of IDS for CII in the future work.

Author Contributions

Please state each author's contribution to this work; it can be up to several sentences long and should briefly describe the tasks of individual authors. For example, AB conducted the research, CD analyzed the data, AB wrote the paper, and all authors approved the final version.

References

[1] Fauziyah Fauziyah, Zhaosun Wang, and Gabriel Joy, "Knowledge Management Strategy for Handling Cyber-Attacks in E- Commerce with Computer Security Incident Response Team CSIRT," *Jounal of Information Security*, vol. 13, no. 4, pp. 294-311, 2022. [Google Scholar] [Publisher Link]

[2] Mario Aragonés Lozano, Israel Pérez Llopis, and Manuel Esteve Domingo, "Threat Hunting System for Protecting Critical Infrastructures Using a Machine Learning Approach," *Mathematics*, vol. 11, no. 16, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Zubair Baig, "Multi-Agent Systems for Protecting Critical Infrastructures: A Survey," *Journal of Network and Computer Applications*, vol. 35, no. 3, pp. 1151-1161, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[4] Suhail Qadir, and Uzair Bashir, "Measuring the Impact of DoS Attack on Availability: Empirical Study Based on Accessibility," *Journal of Information Security*, vol. 13, no. 2, pp. 66-75, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] William Grant Hatcher, and Wei Yu, "A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends," *IEEE Access*, vol. 6, pp. 24411-24432, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] Weihong Ren et al., "Technical Framework Research on Critical Information Infrastructure Cybersecurity Classified Protection," *Proceedings of the 4th International Conference on Machinery, Materials and Information Technology Applications*, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[7] Hugo Riggs et al., "Impact, Vulnerabilities, and Mitigation Strategies for Cyber-Secure Critical Infrastructure," *Sensors*, vol. 23, no. 8, pp. 1-26, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Mitchell Kirshner, "Model-Based Systems Engineering Cybersecurity for Space Systems," *Aerospace*, vol. 10, no. 2, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Sheraz Naseer et al., "Enhanced Net Work Anomaly Detection Based on Deep Neural Networks," *IEEE Access*, vol. 6, pp. 48231-48246, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[10] Rosslin John Robles et al., "Common Threats and Vulnerabilities of Critical Infrastructures," *International Journal of Control and Automation*, vol. 1, no. 1, pp. 17-22, 2008. [Google Scholar] [Publisher Link]

[11] Rui Filipe Silva, Raul Barbosa, and Jorge Bernardino, "Intrusion Detection Systems for Mitigating SQL Injection Attacks: Review and State-Of- Practice," *International Journal of Information Security and Privacy*, vol. 14, no. 2, pp. 1-21, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Tianqing Zhu et al., "More Than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 6, pp. 2824-2843, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Javier Lopez, Roberto Setola, and Stephen D. Wolthusen, *Overview of Critical Information Infrastructure Protection*, Critical Infrastructure Protection: Advances in Critical Infrastructure Protection: Information Infrastructure Models, Analysis, and Defense, Springer, Berlin, Heidelberg, pp. 1-14, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[14] Eugene Nickolov, "Critical Information Infrastructure Protection: Analysis, Evaluation and Expectations," *Information and Security*, vol. 17, pp. 105-116, 2006. [Google Scholar]

[15] Leandros Maglaras, Helge Janicke, and Mohamed Amine Ferrag, "Cybersecurity of Critical Infrastructures: Challenges and Solutions," *Sensors*, vol. 22, no. 14, pp. 1-4, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Hussein Ali, Omar M. Elzeki, and Samir Elmougy, "Smart Attacks Learning Machine Advisor System for Protecting Smart Cities from Smart Threats," *Applied Sciences*, vol. 12, no. 13, pp. 1-24, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Manuel Domínguez et al., "Design of Platforms for Experimentation in Industrial Cybersecurity," *Applied Sciences*, vol. 12, no. 13, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Leandros Maglaras et al., "Threats, Countermeasures and Attribution of Cyber-Attacks on Critical Infrastructures," *EAI Endorsed Transactions on Security and Safety*, vol. 5, no. 16, pp. 1-9, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[19] Mazen Gazzan, and Frederick T. Sheldon, "Opportunities for Early Detection and Prediction of Ransomware Attacks Against Industrial Control Systems," *Future Internet*, vol. 15, no. 4, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Bambang Susilo, and Riri Fitri Sari, "Intrusion Detection in IOT Networks Using Deep Learning Algorithm," *Information*, vol. 11, no. 5, pp. 1-11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Yang Xin et al., "Machine Learning and Deep Learning Methods for Cybersecurity," *IEEE Access*, vol. 6, pp. 35365-35381, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[22] Yadigar N. Imamverdiyev, and Fargana J. Abdullayeva, "Deep Learning in Cybersecurity: Challenges and Approaches," *International Journal of Cyber Warfare and Terrorism*, vol. 10, no. 2, pp. 82-105, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[23] Fotios Gioulekas et al., "A Cybersecurity Culture Survey Targeting Healthcare Critical Infrastructures," *Healthcare*, vol. 10, no. 2, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Kitty Kioskli et al., "The Importance of Conceptualizing the Human-Centric Approach in Maintaining and Promoting Cybersecurity-Hygiene in Healthcare 4.0," *Applied Sciences*, vol. 13, no. 6, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Namhla Mtukushe et al., "Review of Cyberattack Implementation, Detection, and Mitigation Methods in Cyber-Physical Systems," *Energies*, vol. 16, no. 13, pp. 1-25, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Daria Gaskova, and Elena Galperova, "Decision Support in the Analysis of Cyber Situational Awareness of Energy Facilities," *Engineering Proceedings*, vol. 33, no. 1, pp. 1-9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[27] Nikolaos Nikolaou et al., "Vulnerability Identification and Assessment for Critical Infrastructures in the Energy Sector," *Electronics*, vol. 12, no. 14, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] Ricardo Severino et al., "Performance Assessment and Mitigation of Timing Covert Channels Over the IEEE 802.15.4," *Journal of Sensor and Actuator Networks*, vol. 12, no. 4, pp. 1-25, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[29] Andrea Pinto et al., "Survey on Intrusion Detection Systems Based on Machine Learning Techniques for the Protection of Critical Infrastructure," *Sensors*, vol. 23, no. 5, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[30] Mengmeng Ge et al., "Towards a Deep Learning-Driven Intrusion Detection Approach for Internet of Things," *Computer Networks*, vol. 186, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[31] Rotem Bar, and Chen Hajaj, "Simcse for Encrypted Traffic Detection and Zero-Day Attack Detection," *IEEE Access*, vol. 10, pp. 56952-56960, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[32] Asmaa Halbouni et al., "Machine Learning and Deep Learning Approaches for Cybersecurity: A Review," *IEEE Access*, vol. 10, pp. 19572-19585, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[33] Elochukwu Ukwandu et al., "Cyber-Security Challenges in Aviation Industry: A Review of Current and Future Trends," *Information*, vol. 13, no. 3, pp. 1-22, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[34] Daniel Fährmann et al., "Lightweight Long Short-Term Memory Variational Auto-Encoder for Multivariate Time Series Anomaly Detection in Industrial Control Systems," *Sensors*, vol. 22, no. 8, pp. 1-23, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[35] M.R. Gauthama Raman, Chuadhry Mujeeb Ahmed, and Aditya Mathur, "Machine Learning for Intrusion Detection in Industrial Control Systems: Challenges and Lessons from Experimental Evaluation," *Cybersecurity*, vol. 4, no. 1, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[36] Dimitrios Papamartzivanos, Félix Gómez Mármol, and Georgios Kambourakis, "Introducing Deep Learning Self- Adaptive Misuse Network Intrusion Detection Systems, *IEEE Access*, vol. 7, pp. 13546-13560, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[37] Dipanjan Das Roy, and Dongwan Shin, "Network Intrusion Detection in Smart Grids for Imbalanced Attack Types Using Machine Learning Models," *IEEE International Conference on Information and Communication Technology Convergence*, Jeju, Korea (South), pp. 576-581, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[38] Gbenga Ikuomenisan, and Yasser Morgan, "Meta-Review of Recent and Landmark Honeypot Research and Surveys," *Journal of Information Security*, vol. 13, no. 4, pp. 181-209, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[39] Iman Sharafaldin et al., "Developing Realistic Distributed Denial of Service (DDOS) Attack Dataset and Taxonomy," *IEEE International Carnahan Conference on Security Technology*, Chennai, India, pp. 1-8, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[40] Patrick Vanin et al., "A Study of Network Intrusion Detection Systems Using Artificial Intelligence/Machine Learning," *Applied Sciences*, vol. 12, no. 22, pp. 1-27, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[41] Magdy M. Fadel et al., "HDLIDP: A Hybrid Deep Learning Intrusion Detection and Prevention Framework," *Computers, Materials and Continua*, vol. 73, no. 2, pp. 2293-2312, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[42] Auwal Sani Iliyasu, Usman Alhaji Abdurrahman, and Lirong Zheng, "Few-Shot Network Intrusion Detection using Discriminative Representation Learning with Supervised Autoencoder," *Applied Sciences*, vol. 12, no. 5, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[43] Mohd Nor Akmal Khalid, Amjed Ahmed Al-Kadhimi, and Manmeet Mahinderjit Singh, "Recent Developments in Game-Theory Approaches for the Detection and Defense against Advanced Persistent Threats (APTS): A Systematic Review," *Mathematics*, vol. 11, no. 6, pp. 1-34, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[44] Joakim Kävrestad et al., "Evaluation of Contextual and Game-Based Training for Phishing Detection," *Future Internet*, vol. 14, no. 4, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[45] Ruipeng Yang et al., "Subspace Clustering Via Graph Auto-Encoder Network for Unknown Encrypted Traffic Recognition," *Cybersecurity*, vol. 5, no. 1, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[46] Yakub Kayode Saheed et al., "A Machine Learning-Based Intrusion Detection for Detecting Internet of Things Network Attacks," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9395-9409, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[47] R. Vinayakumar et al., "Robust Intelligent Malware Detection using Deep Learning," *IEEE Access*, vol. 7, pp. 46717-46738, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[48] Hasan Alkahtani, and Theyazn H. H. Aldhyani, "Intrusion Detection System to Advance Internet of Things Infrastructure-Based Deep-Learning Algorithms," *Complexity*, vol. 2021, pp. 1-18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[49] Kun Jia et al., "A Lightweight DDOS Detection Scheme under SDN Context," *Cybersecurity*, vol. 5, no. 1, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[50] Zihan Wu et al., "RTIDS: A Robust Transformer- Based Approach for Intrusion Detection System," *IEEE Access*, vol. 10, pp. 64375-64387, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[51] Hashida Haidros Rahima Manzil, and S. Manohar Naik, "Android Malware Category Detection Using a Novel Feature Vector-Based Machine Learning Model," *Cybersecurity*, vol. 6, no. 1, pp. 1-11, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[52] R. Vinayakumar et al., "Deep Learning Approach for Intelligent Intrusion Detection System," *IEEE Access*, vol. 7, pp. 41525-41550, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[53] Mohamed Amine Ferrag et al., "Edge- IIOTSET: A New Comprehensive Realistic Cyber Security Dataset of IOT and IIOT Applications for Central Ized and Federated Learning," *IEEE Access*, vol. 10, pp. 40281-40306, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[54] Jokha Ali, "Intrusion Detection Systems Trends to Counteract Growing Cyber- Attacks on Cyber-Physical Systems," *22nd International Arab Conference on Information Technology*, Muscat, Oman, pp. 1-6, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[55] Masoud Mehrabi Koushki et al., "On Building Machine Learning Pipelines for Android Malware Detection: A Procedural Survey of Practices, Challenges and Opportunities," *Cybersecurity*, vol. 5, no. 1, pp. 1-37, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[56] Esra Söğüt, and O. Ayhan Erdem, "A Multi- Model Proposal for Classificationm and Detection of DDOS Attacks on Scada Systems," *Applied Sciences*, vol. 13, no. 10, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[57] Muhammad Imran Tariq et al., "A Review of Deep Learning Security and Privacy Defensive Techniques," *Mobile Information Systems*, vol. 2020, pp. 1-8, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[58] Sapna Sadhwani et al., "A Lightweight Model for DDOS Attack Detection using Machine Learning Techniques," *Applied Sciences*, vol. 13, no. 17, pp. 1-31, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[59] A.O. Adejimi et al., "A Dynamic Intrusion Detection System for Critical Information Infrastructure," *Scientific African*, vol. 21, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[60] F. Zare, and P. Mahmoudi-Nasr, "Feature Engineering Methods in Intrusion Detection System: A Performance Evaluation," *International Journal of Engineering, Transactions B: Applications*, vol. 36, no. 7, pp. 1343-1353, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[61] Chen Hajaj, Nitay Hason, and Nitay Hason, "Less is More: Robust and Novel Features for Malicious Domain Detection," *Electronics*, vol. 11, no. 6, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[62] Ondrej Linda, Todd Vollmer, and Milos Manic, "Neural Network-Based Intrusion Detection System for Critical Infrastructures," *International Joint Conference on Neural Networks*, Atlanta, GA, USA, pp. 1827-1834, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[63] Abdelouahid Derhab et al., "Intrusion Detection System for Internet of Things Based on Temporal Convolution Neural Network and Efficient Feature Engineering," *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1-16, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[64] William Villegas-Ch, Jaime Govea, and Angel Jaramillo-Alcazar, "IoT Anomaly Detection to Strengthen Cybersecurity in the Critical Infrastructure of Smart Cities," *Applied Sciences*, vol. 13, no. 19, pp. 1-20, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[65] Abdu Salam et al., "Deep Learning Techniques for Web-Based Attack Detection in Industry 5.0: A Novel Approach," *Technologies*, vol. 11, no. 4, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[66] Raisa Abedin Disha, and Sajjad Waheed, "Performance Analysis of Machine Learning Models for Intrusion Detection System using GINI Impurity-Based Weighted Random Forest (GIWRF) Feature Selection Technique," *Cybersecurity*, vol. 5, no. 1, pp. 1-22, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[67] Azriel Henry et al., "Composition of Hybrid Deep Learning Model and Feature Optimization for Intrusion Detection System," *Sensors*, vol. 23, no. 2, pp. 1-22, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[68] Yanmiao Li et al., "One-Class LSTM Network for Anomalous Network Traffic Detection," *Applied Sciences*, vol. 12, no. 10, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[69] Najla Al-Taleb, and Nazar Abbas Saqib, "Towards a Hybrid Machine Learning Model for Intelligent Cyber Threat Identification in Smart City Environments," *Applied Sciences*, vol. 12, no. 4, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[70] Bedine Kerim, "Securing IoT Net- Work Against DDOS Attacks Using Multi-Agent IDS," *Journal of Physics: Conference Series*, vol. 1898, no. 1, pp. 1-7, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[71] Robert Osei-Kyei et al., "Systematic Review of Critical Infrastructure Resilience Indicators," *Construction Innovation*, vol. 23, no. 5, pp. 1210-1231, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[72] Mohanad Sarhan et al., "Feature Extraction for Machine Learning-Based Intrusion Detection in Iot Networks," *Digital Communications and Networks*, vol. 10, no. 1, pp. 205-216, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[73] Sydney M. Kasongo, and Yanxia Sun, "Performance Analysis of Intrusion Detection Systems using a Feature Selection Method on the Unsw-Nb15 Dataset," *Journal of Big Data*, vol. 7, no. 1, pp. 1-20, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[74] Dhiaa Musleh et al., "Intrusion Detection System using Feature Extraction with Machine Learning Algorithms in IoT," *Journal of Sensor and Actuator Net-Works*, vol. 12, no. 2, pp. 1-19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[75] Anita Gehlot, and Ankita Joshi, "Neural Network Based Intrusion Detection System for Critical Infrastructure," *2nd Mysore Sub Section International Conference*, Mysuru, India, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[76] Ahmad Javaid et al., "A Deep Learning Approach for Network Intrusion Detection Systems," *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies*, pp. 21-26, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[77] Dimitris Deyannis et al., "The Diversification and Enhancement of an IDs Scheme for the Cybersecurity Needs of Modern Supply Chains," *Electronics*, vol. 11, no. 13, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[78] Opeyemi Lateef Usman et al., "Advance Machine Learning Methods for Dyslexia Biomarker Detection: A Review of Implementation Details and Challenges," *IEEE Access*, vol. 9, pp. 36879-36897, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[79] Tao Peng et al., "A Lightweight Multi-Source Fast Android Malware Detection Model," *Applied Sciences*, vol. 12, no. 11, pp. 1-25, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[80] Jorge Hochstetter-Diez et al., "A Prioritization Strategy for Public Institutions to Improve Information Security Maturity," *Applied Sciences*, vol. 13, no. 14, pp. 1-20, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[81] Damilola Adesina et al., "Adversarial Machine Learning in Wireless Communications using Rf Data: A Review," *IEEE Communications Surveys and Tutorials*, vol. 25, no. 1, pp. 77-100, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[82] Ioannis Karamitsos et al., "Malware Detection for Forensic Memory using Deep Recurrent Neural Networks," *Journal of Information Security*, vol. 11, no. 2, pp. 103-120, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[83] Adeel Abbas et al., "A New Ensemble-Based Intrusion Detection System for Internet of Things," *Arabian Journal for Science and Engineering*, vol. 47, no. 2, pp. 1805-1819, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[84] Volodymyr Tkach et al., "Non-Pattern-Based Anomaly Detection in Time-Series," *Electronics*, vol. 12, no. 3, pp. 1-25, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[85] Guangming Xian, "Cyber Intrusion Prevention for Large-Scale Semi-Supervised Deep Learning Based on Local and Non-Local Regularization," *IEEE Access*, vol. 8, pp. 55526-55539, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[86] Oyeniyi Akeem Alimi, Khmaies Ouahada, and Adnan M. Abu-Mahfouz, "A Review of Machine Learning Approaches to Power System Security and Stability," *IEEE Access*, vol. 8, pp. 113512-113531, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[87] Lan Liu et al., "Intrusion Detection of Imbalanced Network Traffic Based on Machine Learning and Deep Learning," *IEEE Access*, vol. 9, pp. 7550-7563, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[88] Idriss Idriss et al., "Toward a Deep Learning-Based Intrusion Detection System for IoT against Botnet Attacks," *International Journal of Artificial Intelligence*, vol. 10, no. 1, pp. 110-120, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[89] Abdulaziz Fatani et al., "IoT Intrusion Detection System using Deep Learning and Enhanced Transient Search Optimizationm," *IEEE Access*, vol. 9, pp. 123448-12346, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[90] Imtiaz Ullah, and Qusay H. Mahmoud, "Design and Development of a Deep Learning-Based Model for Anomaly Detection in IoT Networks, *IEEE Access*, vol. 9, pp. 103906-103926, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[91] Adel Alqudhaibi et al., "Predicting Cybersecurity Threats in Critical Infrastructure for Industry 4.0: A Proactive Approach Based on Attacker Motivations," *Sensors*, vol. 23, no. 9, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[92] Khaled M. Alnifie, and Charles Kim, "Appraising the Manifestation of Optimism Bias and its Impact on Human Perception of Cyber Security: A Meta-Analysis," *Journal of Information Security*, vol. 14, no. 2, pp. 93-110, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[93] Haizhou Wang, Anoop Singhal, and Peng Liu, "Tackling Imbalanced Data in Cybersecurity with Transfer Learning: A Case with Rop Payload Detection," *Cybersecurity*, vol. 6, no. 1, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[94] Rubayyi Alghamdi, and Martine Bellaiche, "An Ensemble Deep Learning Based IDs for IoT using Lambda Architecture," *Cybersecurity*, vol. 6, no. 1, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[95] Sung Bum Park, Hyo Jin Jo, and Dong Hoon Lee, "G-IDCS: Graph-Based Intrusion Detection and Classification System for can Protocol," *IEEE Access*, vol. 11, pp. 39213-39227, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[96] Zeeshan Ahmad et al., "Network Intrusion Detection System: A Systematic Study of Machine Learning and Deep Learning Approaches," *Transactions on Emerging Telecommunications Technologies*, vol. 32. no. 1, pp. 1-29, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[97] Béla Genge, Piroska Haller, and Adrian-Silviu Roman, "E-Apt Detect: Early Advanced Persistent Threat Detection in Critical Infrastructures with Dynamic Attestation," *Applied Sciences*, vol. 13, no. 6, pp. 1-22, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[98] Konstantinos Ntafloukas, Daniel P. McCrum, and Daniel P. McCrum, "A Cyber-Physical Risk Assessment Approach for Internet of Things Enabled Transportation Infrastructure," *Applied Sciences*, vol. 12, no. 18, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[99] Yanjie He, and Wei Li, "A Novel Lightweight Anonymous Proxy Traffic Detection Method Based on Spatial-Temporal Features," *Sensors*, vol. 22, no. 11, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[100] Chin-Ling Chen, and Jian Lin Lai, "An Experimental Detection of Distributed Denial of Service Attack in CDX 3 Platform Based on Snort," *Sensors*, vol. 23, no. 13, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[101] Mahmoud Al-Dwairi et al., "Ransomware-Resilient Self- Healing Xml Documents," *Future Internet*, vol. 14, no. 4, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[102] João Vitorino, Nuno Oliveira, and Isabel Praça, "Adaptative Perturbation Patterns: Realistic Adversarial Learning for Robust Intrusion Detection," *Future Internet*, vol. 14, no. 4, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[103] Yingchun Niu et al., "Application of a New Feature Generation Algorithm in Intrusion Detection System," *Wire-Less Communications and Mobile Computing*, vol. 2020, pp. 1-7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[104] Abdulrahman Al-Abassi et al., "An Ensemble Deep Learning-Based Cyber-Attack Detection in Industrial Control System," *IEEE Access*, vol. 8, pp. 83965-83973, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[105] In-Sun Choi, Junho Hong, and Tae-Wan Kim, "Multi-Agent Based Cyber-Attack Detection and Mitigation for Distribution Automation System," *IEEE Access*, vol. 8, pp. 183495-183504, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[106] Yakub Kayode Saheed, Oluwadamilare Harazeem Abdulganiyu, and Taha Ait Tchakoucht, "A Novel Hybrid Ensemble Learning for Anomaly Detection in Industrial Sensor Networks and Scada Systems for Smart City Infrastructures," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 5, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[107] Sydney Mambwe Kasongo, and Yanxia Sun, "A Deep Learning Method with Filter-Based Feature Engineering for Wireless Intrusion Detection System," *IEEE Access*, vol. 7, pp. 38597-38607, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[108] Tomás Sureda Riera et al., "A New Multi-Label Dataset for Web Attacks Capec Classification using Machine Learning Techniques," *Computers and Security*, vol. 120, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[109] Shailesh Singh Panwar, Y.P. Raiwani, and Lokesh Singh Panwar, "Evaluation of Network Intrusion Detection with Features Selection and Machine Learning Algorithms on Cicids-2017 Dataset," *International Conference on Advances in Engineering Science Management & Technology*, Uttaranchal University, Dehradun, India, pp. 1-10, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[110] Ansam Khraisat et al., "Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges," *Cybersecurity*, vol. 2, no.1, pp. 1-22, 2019. [CrossRef] [Google Scholar] [Publisher Link]