Review Article

Yahya Layth Khaleel, Mustafa Abdulfattah Habeeb, A. S. Albahri*, Tahsien Al-Quraishi, O. S. Albahri, and A. H. Alamoodi

Network and cybersecurity applications of defense in adversarial attacks: A state-of-the-art using machine learning and deep learning methods

https://doi.org/10.1515/jisys-2024-0153
received March 08, 2024; accepted May 20, 2024

Abstract: This study aims to perform a thorough systematic review investigating and synthesizing existing research on defense strategies and methodologies in adversarial attacks using machine learning (ML) and deep learning methods. A methodology was conducted to guarantee a thorough literature analysis of the studies using sources such as ScienceDirect, Scopus, IEEE Xplore, and Web of Science. A question was shaped to retrieve articles published from 2019 to April 2024, which ultimately produced a total of 704 papers. A rigorous screening, deduplication, and matching of the inclusion and exclusion criteria were followed, and hence 42 studies were included in the quantitative synthesis. The considered papers were categorized into a coherent and systematic classification including three categories: security enhancement techniques, adversarial attack strategies and defense mechanisms, and innovative security mechanisms and solutions. In this article, we have presented a systematic and comprehensive analysis of earlier studies and opened the door to potential future studies by discussing in depth four challenges and motivations of adversarial attacks, while three recommendations have been discussed. A systematic science mapping analysis was also performed to reorganize and summarize the results of studies to address the issues of trustworthiness. Moreover, this research covers a large variety of network and cybersecurity applications of defense in adversarial attack subjects, including intrusion detection systems, anomaly detection, ML-based defenses, and cryptographic techniques. The relevant conclusions well demonstrate what have achieved in defense mechanisms against adversarial attacks. In addition, the analysis revealed a few emerging tendencies and deficiencies in the area to be remedied through better and more dependable mitigation methods against advanced persistent threats. The findings of this review have crucial implications for the community of researchers, practitioners, and policy makers in network and cybersecurity using artificial intelligence applications.

Keywords: adversarial attack, defense, network, cybersecurity, methods, artificial intelligence

* Corresponding author: A. S. Albahri, Technical College, Imam Ja’afar Al-Sadiq University, Baghdad 10001, Iraq; University of Information Technology and Communications (UoITC), Baghdad, 10013, Iraq, e-mail: ahmed.albahri@ijsu.edu.iq
Yahya Layth Khaleel: Department of Computer Science, College of Computer Science and Mathematics, Tikrit University, Salahuddin 3400, Iraq, e-mail: yahya@tu.edu.iq
Mustafa Abdulfattah Habeeb: Department of Computer Science, College of Computer Science and Mathematics, Tikrit University, Salahuddin 3400, Iraq, e-mail: mustafa@tu.edu.iq
Tahsien Al-Quraishi: Information Technology and Systems, Victorian Institute of Technology (VIT), Melbourne 3000, Australia, e-mail: tahsien.a@vit.edu.au
O. S. Albahri: Australian Technical and Management College, Melbourne 3000, Australia; Computer Techniques Engineering Department, Mazaya University College, Nasiriyah 64001, Iraq, e-mail: osamah.shihab@atmc.edu.au
A. H. Alamoodi: Applied Science Research Center, Applied Science Private University, Amman, 11931, Jordan; MEU Research Unit, Middle East University, Amman, 11831, Jordan, e-mail:alamoodi.abdullah@asu.edu.jo, alamoodi.adjunct.researcher@meu.edu.jo

Open Access. © 2024 the author(s), published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.
1 Introduction

With the proliferation of network gear and technologies, virtually every individual now accesses the Internet on a daily basis. According to projection studies, the number of Internet users will surpass 7.5 billion by 2030 [1]. These individuals regularly engage with Internet services such as online banking, healthcare transactions, marketing, entertainment, and education. Just as in the physical realm, malicious entities online, known as cybercriminals, seek to deceive and exploit genuine users for their gain. They utilize various cyberattacks, such as malware attacks, as tools to achieve their objectives. Cyberattacks represent the swiftest expanding form of crime worldwide, resulting in financial losses that exceed the global trade in all illicit drugs [2]. According to projections, the global cost of cybercrime is expected to exceed more than ten trillion US dollars annually by 2025, which is a significant increase from the three trillion US dollars recorded in 2015 [1]. Information to be succeeded by new technology and methods of detection and identification of immediate threats and attacks is much needed to fill in the cybersecurity gap due to a rapid change in the cyberthreat environment.

Cybersecurity experts are looking into the application of deep learning (DL), a machine learning (ML) technique. This capability has partly become a reason why this particular domain is so sprung and expanded [3–5]. DL-integrated cybersecurity solutions are implied to spontaneously expect and detect malicious threats and later on, automatically updating themselves by expanding associated capabilities [6]. Not surprisingly, DL models are increasingly hailed as a key tool for combating various cyberattacks that have been seen in recent years. For example, the range of security data refers to many data sources including network sensors and logs, so become more numerous types. Moreover, they have a rich depth and diversity of the data as well as several connections [7]. Traditional ML algorithms cannot make use of the high dimensional information and do not benefit from larger volumes of data either. Therefore, for analysts, many problems have become too difficult or advanced [3].

On the other hand, data hierarchy is a key concept that enables DL to work without specific domain knowledge for feature extraction [8,9]. To sum up, DL actually substitutes feature engineering by using multilayered, nonlinear hidden layers for feature representation learning. This minimizes the cost and time taken in hiring staff to re-engineer the functionality at any time a new change in cyberattacks comes in, say zero-day malware. The satisfaction from these productive uses will spotlight the developing trust of both the public and private sectors in the usage of DL in the cybersecurity arena [10,11].

Both the nature and complexity of the task of protecting digital environments in the face of the special threats are illustrated in the provided study [12]. This methodical review aims to enlighten readers on the cybersecurity as well as network defense strategies that are being used through an in-depth analysis of current tactics, pitfalls, and what is new in cyberattack methods [13]. Aversive assaults cover techniques, such as attack libraries, homomorphic encryption, and side channel attacks to break through the systems’ notoriety, compromise the integrity of data and hamper essential services [14]. The variety of the threat landscape is wide and dynamic, including distributed denial of service (DDoS) aimed to spoil the Internet infrastructures to ransomware cases that target confidential data. To establish comprehensive defenses and prevent possible damage from these threats, it is vital to explore in depth the features of these measures [15].

The first line of the defense against hostile invasions is provided by cyber and network security measures, which are the spectrum of the technologies, mechanisms, and practices that should be adopted either individually or together [16]. The target of this evidence-based study is the evaluation of both positive and negative aspects of current cyber defense mechanisms to have a deeper understanding of their readiness and ability to detect, respond, and counteract hostile intrusions. The integration of validated information, theoretical models, and practical aspects could be considered the principal purpose of this review, which aims to provide the subject of the current state of security in computer networks from different points of view. The article also makes it obvious how attackers and defenders play the cat-and-mouse game by trying to resolve problems between offensive strategies and the security systems they apply. Adversaries mutate their strategies by applying new methods as cybersecurity techniques become more advanced, and hence, this points out the flaws and detects vulnerabilities to overcome traditional defenses. This piece of mystery becomes a cornerstone in the issue of what new threats may arise and how we can make our network security more secure.
The following inquiries help clarify the study goals aligned with these reasons:

Q1: What is the appropriate taxonomy for incorporating protection techniques and strategies against defense from adversarial attacks in networks and cybersecurity?

Q2: What are the motives, challenges, recommendations, and limitations related to the integration of defense methods addressed in research?

Q3: What are the most significant gaps in the literature currently concerning network and cybersecurity applications of defense in adversarial attacks security strategies against adversarial attacks?

The study’s primary contributions include the following:

(1) A comprehensive investigation of adversarial attacks in defense technique invention and optimization, adversarial attack generation, defensive strategies, defense robustness, and applications in malware, intrusion, and anomaly detection is performed.

(2) A thorough literature analysis is offered that delves into current research trends, obstacles, drivers, constraints, and suggestions in the area of network and cybersecurity applications of defense in adversarial attacks and tactics against adversarial attacks.

(3) Identifying research gaps and suggesting future lines of inquiry, providing a roadmap for improving defense tactics in a variety of adversarial attack scenarios.

The planned outline of the research article contributes structure to the entire paper that can be followed step by step by the readers and starts with presenting the objectives, the applied methodology, the results, and implications. It is the beginning section of the article that gives a concise introduction to the research, thus setting the canvas to be further developed. The second part gives an explanation of why the research topic was selected starting with a general introduction to adversarial attacks and followed by the best existing papers. Then a finer analysis was performed that resulted in the research project. As a result, the research methodology in Section 3 explicates the rational pattern used to focus and pick research for examination purposes, which incontrovertibly augments the transparency and reliability of the research approach. Section 4 highlights the critical steps of this study, including the analysis of research articles, the extraction of key terms, and the division of results. These elements provide a solid basis for the study’s findings. Section 5 consolidates the main points in previous studies and presents them through motivations, challenges, and recommendations. In Section 6, which highlights areas for future research, gaps in this systematic field are mapped, and some points are reviewed to locate their deficiencies and to reveal the new exploratory field. Finally, the conclusion part gives a well-grounded statement on a major finding of the study and its implications in a gist.

2 Adversarial attack: Overview and analysis

This section commences with an introduction to adversarial attacks, including frameworks of their functioning, methods of implementation, and strategies used. The phenomenon of adversarial attacks explains the much deeper level of how such attacks are executed and the techniques used to destabilize ML systems. The second section provides a critical analysis of the most important and recent papers in this field. It introduces the global indicator (citations) for the papers within this field, showing their importance and connection to the debate on choosing the research project topic, and explains the logic behind this choice.

2.1 Overview

Adversarial attacks are deliberate modifications to inputs that cause ML models to make incorrect predictions, posing serious risks in applications such as autonomous vehicles (AV) and medical diagnosis [17]. Many researchers have worked in the field of adversarial attacks. First, Chakraborty et al. [18] described the goals of adversarial learning, as shown in Figure 1.
The process of determining the circumstances that may lead to undesirable results or anomalies in the ML and DL models is called an adversarial attack. Such occurrences are security concerns, especially when the model includes private or sensitive data [19]. Across diverse domains, extensive research endeavours are ongoing to explore adversarial ML attacks, which pose a substantial threat to the broad adoption of ML and DL solutions in contexts crucial for security [20, 21].

Since Szegedy et al. [22] asserted that neural networks are susceptible to adversarial attacks, there has been an increasing focus on examining adversarial technologies within AI. Scholars continually devise novel methods to counter hostile attacks [23].

During the life cycle of ML and DL systems, five types of security threats can be distinguished: (1) poisoning attacks; (2) backdoor attacks; (3) adversarial example attacks; (4) model theft; and (5) recovery of critical training data. The initial three attacks occur during the testing phase, whereas the first two attacks transpire during the training phase (Figure 2) [24].
For example, in a poisoning attack targeting an automated vehicle, the attacker interferes with the laser signal and visual input during the ML model’s training. This alteration aims to cause the automated vehicle system to inaccurately classify traffic signs and misinterpret object distances during the testing phase [25]. Backdoor attacks implant concealed associations or triggers within ML and DL models to supersede accurate inference, such as classification, compelling the system to operate maliciously based on the attacker’s designated target. In the absence of the trigger, the system behaves normally [26].

Conversely, white-box attacks and black-box attacks represent two categories of adversarial attacks. In a white-box attack scenario, the attacker possesses information regarding the system’s architecture, modelling, weights, training set, and sample data. Within this context, the classification function becomes vulnerable to adversarial attacks, which can jeopardize the system due to the attacker’s comprehensive knowledge [27]. In black-box scenarios, adversaries lack access to information about the target model. However, they can employ techniques such as model inversion, exploit the transferability of adversarial samples, or query the target model to create a local substitute model [23].

Conversely, ML and DL are gaining traction in the field of malware detection because they cannot only detect known malware but also uncover new and covert malicious software [28]. When using ML/DL for malware analysis, there are two primary steps [29]. In the training phase, ML/DL algorithms process a set of features derived from both malicious and non-malicious data to create a predictive model [19]. In the following phase, the testing phase, the predictive model developed in the training phase is utilized to forecast the benign behaviour of the malware. ML-based adversarial models are often incorporated into various solutions to assess the impact of an attacker’s manipulation of a classifier during the training phase on the testing phase [19,20,30,31].

On the other hand, the latest research about adversarial attacks has been rich in its content and diversity, and this study elaborates on these concepts. Zhu et al. [32] propose an approach for image-to-image translation without paired training data, employing adversarial loss to learn a mapping from a source domain to a target domain. Coupling this with a cycle consistency loss ensures high-quality results across various tasks. Isola et al. [33] explore conditional adversarial networks, offering a versatile solution for diverse image translation tasks with effective results demonstrated and popular adoption evident from community engagement. Despite the significant progress completed in enhancing the accuracy and speed of single-image super-resolution through the utilization of faster and deeper convolutional neural networks, a fundamental challenge persists: How can we effectively restore the intricate texture details when performing super-resolution at considerable upscaling ratios? [34]. The introduction of new technologies like big data [35], cloud computing, and Internet of things (IoT) is the main driving force in technology development of network attacks, and also this triggers network attack detection techniques to always evolve. Three main problems are associated with these technologies: complex traffic from networks automatic representation, skew of attacks by samples in networks, and trade-off between anomaly detection model accuracy and repeated notice evolution [36]. In addition, in the modern world, web applications are a vital means of facilitating the provision of services. In addition, web app usage has experienced phenomenal growth, resulting in a greater number of cyberattacks. Cross-site scripting (XSS) is one of the most prevalent attack vectors in cybersecurity and affects both end users and service providers to an equal degree. In the past few years, there has been a clear increase in the adoption of ML/DL techniques for XSS attack detection. Therefore, it was aimed at highlighting ML and DL techniques [37]. Aldhaheri et al. [38] discuss recent advancements in IDS for the IoT, focusing on embedded DL algorithms and associated datasets, the types of attacks encountered by each model defense model, and the evaluation of the detection metrics. This case study broadly describes a number of challenges involved. A network-based intrusion detection system (IDS) sets up the initial defense against network attacks that breach the integrity of data, systems, and networks [38]. In recent years, deep neural networks (DNNs) have been widely applied in this domain to identify malicious traffic because they can efficiently detect malicious traffic and achieve high detection accuracy. Nevertheless, He et al. [39] established that this NIDS attack launching and detection problem has been a considerable research challenge.
2.2 Analysis

The topic of adversarial attack is important, and researchers began researching and developing it nearly 20 years ago. According to the IEEE website, when searching for “Adversarial Attack” and filtering the results according to years, this topic is considered to be at the height of its spread and work on it, as shown in Figure 3 with the number of contributions in adversarial last 5 years.

![Graph showing number of contributions in adversarial last 5 years.](image)

Figure 3: The adversarial goals [18].

On the other hand, the topic of adversarial attacks resonates deeply within the research community and beyond. This resonance is evidenced by the substantial number of citations garnered by studies exploring this phenomenon. When conducting a search in IEEE “Adversarial Attack,” the 10 highest number of research contributions in terms of the number of citations from the papers [40–49], most of them in the last 10 years and on various topics related to the adversarial attack (Figure 4).

Furthermore, studies that analyse the previous literature on adversarial attacks, whether it be reviews, surveys, or state-of-the-art, by simple search query (“Adversarial Attacks”) AND (“Review”) OR (“Survey”) OR (“State-of-the-Art”)), come into sight the top ten research papers that received the highest number of citations related to adversarial attacks over the last decade [44, 48, 50–57], gathered more than 5,000 citations. This great deal of citations is an indication that the subject’s needs are paramount and that it deserves more emphasis and focus. In addition, it stresses the importance of the adversarial attack issues for the technological advancements in the ML as a key part of the security and reliability of these systems. The aforementioned research findings on the phenomenon of adversarial attacks clearly show the attention given and also the huge visibility of such research. The knowledge across this area should continue to be pursued.

On the other hand, an examination of ScienceDirect using the query (“Defense”) AND (“Adversarial Attack”) and filtering for the years from 2014 to 2023 reveals a notable trend: the research papers related to the mix of defense mechanisms and adversarial attacks are in the stage of strong growth. Figure 5 demonstrates this trend using the strategies of combating the adversarial attacks on the ML and DL models which have been depicted in it. The increasing trend of research in this area signifies the growing consciousness of the need for robust defense measures against the adversary’s manipulation, on the part of the researchers who
are on the constant quest for novel solutions to strengthen the ML and DL systems. This pattern is a sign that the research community puts a great emphasis on addressing the constantly changing nature of security risks caused by aggressive actions and the creation of the most advanced methods of counteraction. Not only so, it underscores the significance of taking the initiative to make ML and DL models secure against possible risks through embracing pro-active approaches, hence enhancing trust and reliability in this technology being used in a broad spectrum of the domains.
In the same context, upon searching the ScienceDirect website using the query (“Network”) OR (“Cybersecurity”) AND (“Adversarial Attack”) and filtering for the years from 2014 to 2023, a discernible trend emerges: the vast number of research papers is now getting involved in research on network security, cybersecurity, and hacker attacks. Figure 6 illustrates the ascending growth pattern, which portrays the intellectual involvement of researchers in the understanding and management of digital risks characterized by adversarial manipulation in technological domains and networks.

![Figure 6: Number of papers related to network and cybersecurity in adversarial attack.](image)

The boosting of the output of research resembles the growing recognition of the multidimensional problems that cyberattacks pose to digital spaces and stimulates the conjoined design of holistic defense and counter-measures. While attackers are constantly inventing new attacks and using vulnerabilities among networked systems and their cyber security frameworks, students and researchers are trying to increase their capacities in detecting, preventing, and responding to threats. This increase is an outcome of collaborative efforts to strengthen networked systems against cyberattacks, which gives protection to critical infrastructure, sensitive information, and digital assets in spite of unlawful and unethical behaviours (Figure 6).

Moreover, according to the aforementioned analysis, the intersection of defense mechanisms in adversarial network attacks and cybersecurity is a vital and enthralling research topic, artistically presented in academic research. This convergence brings together three fundamental areas of study: network security, cybersecurity, and adverse attacks. This research has great significance from the point of view of actually tackling the growing threats of networked malicious automatic identification system. By exposing the scenarios and methods of adversarial attacks and the invention of defense plans, researchers improve cyber security practices. In addition, the interdisciplinary aspect of this topic facilitates closer cooperation, which enhances academic discussion conclusively leading to meaningful advancements in the field (Figure 7).

Figure 7 explains why there is a need for defense against intrusion in networks and cybersecurity in academic research, which demonstrates its importance. To put this in another way, the network security, cybersecurity, and threat points of origin are in constant evolution processes to match the increase of scholars in these areas. The rationale established reveals the significance of this subject, referring to its multifaceted aspect, in conjunction with the practical factors liberating digital security and the possibility of interdisciplinary cooperativity. Summing up, through the aforementioned analysis and my proposed reasons, the
selected issue which is highlighted can be noticed as one of the key issues that require further study in academic research.

3 Methodology

This research followed the approach already adopted in earlier studies [58,59], where a systematic literature review was performed with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement. The analysis was subdivided as stated in the guidelines (Figure 5) [58,60]. The recommended reporting guidelines for systematic reviews and meta-analyses filing of tabular sections outlined earlier were followed during various bibliographic citation databases drawn from a wide range of inclusion criteria that covered scientific and social science journals across different disciplines. The searching for relevant papers was conducted by using four widely recognized and reliable digital databases: SD, Scopus, IEEE, and Web of Science (WoS) [61,62]. These databases are invaluable to researchers because they provide thorough documentation of science and technology research and early indicators for further review and probing [63].

3.1 Search strategy

The search strategy process included all the scientific publications recorded in 2019 – source years up until April 2024. To search, a Boolean query was developed that utilized the AND operator for combining two sets of keywords, “adversarial attack” and (“ML” or “DL”) (consulting Figure 5, to read the most detailed original version). The choice of these keywords was adopted to guarantee an effective and efficient search strategy for relevant literature.

3.2 Inclusion and exclusion criteria

The following criteria were applied for choosing papers:
- The articles are written in English and published through credible scholarly journals or conference papers.
- The papers should encompass the adversarial that incorporates ML and DL.
As discussed earlier, the chosen articles had to address networks and cybersecurity under adversarial attacks.

The following exclusion criteria were applied:

• We omitted papers that discussed adversarial attacks in areas not related to ML and DL, and vice versa.
• We omitted studies that involved adversarial attacks in ML and DL but did not capture anything relevant to network or cybersecurity and vice versa.

3.3 Study selection

This approach involves a sequence of structured components first by identification as well as the removal of papers in duplicate. To analyse the titles and abstracts of the selected articles, the use of Mendeley software was used. This preliminary sieving technique led to the elimination of many unconnected research papers that included only relevant literature. Where some variations or findings did not agree with the authors’ appraisals, a key role was played by the corresponding author in bringing about uniformity. The following step was a comprehensive text review of the entire article after matching very carefully against Section 3.2 inclusion criteria, for which the purpose of this step was to increase the precision of the selection scenario by denying access to those materials that did not correspond with what criteria were set in advance. Figure 8 reflects the process and its results, whereby it becomes possible not only to outline the stages of filtering searched for articles but also to single out those that require proper analysis.

This research focused on identifying and selecting those articles that met a set of specified criteria. First, a comprehensive search revealed 704 entries comprising the articles from the SD, totalling 659; furthermore, only Scopus 34, only one article in IEEE, contributed 10 to the WoS. Two duplicate records were found and removed solely to eliminate redundancy, leaving no remaining number of papers at this number (702). Therefore, detailed scrutiny of titles and abstracts revealed that 442 articles were excluded because they did not comply with the predefined criteria. A comprehensive analysis was then performed for the subsequent 260 contributions. A total of 218 studies were excluded due to failure to meet other inclusion criteria. In the end, 42 of these studies were included in the final collection of articles.

4 Findings analysis

Through a systematized effort to classify and analyse attacks, this study endeavours to offer deep knowledge about adversarial attacks in networks and cybersecurity using ML and DL methods. The results of the final set of articles were obtained through two main analyses: bibliometric analysis and taxonomy analysis. Section 4.1 presents nine statistical analyses using several figures to visualize and understand publications on networks and cybersecurity relationships. In taxonomy analysis, the final set of findings in the articles is discussed in Section 4.2, where a comprehensive analysis and segregation are conducted by dividing the findings into different categories based on their specific objectives as well as what they contribute to this perspective article.

4.1 Bibliometric analysis

An influx of contributions and project research has made the identification of essence in older studies difficult. Currently, with thousands of practical and theoretical inputs, it is quite difficult to follow everything in the literature [64]. Some academics have proposed the PRISMA paradigm as a means of reworking previous reports, highlighting problems, and identifying research gaps. Furthermore, although systematic reviews expand the body of knowledge, clarify research paradigms, and synthesize literature products, they still
face reliability and objectivity challenges. This arises from the fact that they rely on the authors’ opinions to rephrase the earlier findings. To improve transparency in summarizing past study findings, myriad research efforts have called for performing more holistic science mapping analysis through RStudio [65]. The use of a bibliometric method produces indisputable results, exposes gaps, and combines literary findings with high
clarity and trustworthiness [66]. Furthermore, the proposed tools are easy to use and free, and no advanced skills are necessary [67,68]. Therefore, this article employs the more comprehensive bibliometric approach explained below. The shapes produced are based on default settings regarding the number of articles using the R Studio program (Figures 9–21).

4.1.1 Most relevant words

Common words that are often repeated in research papers in this field are indicated in Figure 9. These words are the key descriptors of popular themes and topics in the literature, providing an idea of the major areas of concentration as well as research trends. Through the identification of the recurring terms, researchers can identify the focused and central issues of the scholarly discourse in the field of network and cybersecurity applications of defense against adversarial attacks.

**Figure 9:** Most relevant words.

**Figure 10:** Most relevant sources.

4.1.1 Most relevant words

Common words that are often repeated in research papers in this field are indicated in Figure 9. These words are the key descriptors of popular themes and topics in the literature, providing an idea of the major areas of concentration as well as research trends. Through the identification of the recurring terms, researchers can identify the focused and central issues of the scholarly discourse in the field of network and cybersecurity applications of defense against adversarial attacks.
We selected the top essential words from 42 papers. For this purpose, we identified the main concepts and significant terms of each article. With a wide variety of topics presented in these papers, the most useful words will change based on which article they are associated with. There are many shared subjects in these studies, such as ML, cybersecurity, IDSs, adversarial attacks, DL, and other similar ones.

4.1.2 Most relevant sources

Figure 10 shows the most significant journals as sources of publications based on the number of times they were cited in the papers. The figure demonstrates the most influential and frequently cited journals as sources for papers.

**Figure 11:** Word frequency over time.

**Figure 12:** Source frequency over time.

We selected the top essential words from 42 papers. For this purpose, we identified the main concepts and significant terms of each article. With a wide variety of topics presented in these papers, the most useful words will change based on which article they are associated with. There are many shared subjects in these studies, such as ML, cybersecurity, IDSs, adversarial attacks, DL, and other similar ones.

4.1.2 Most relevant sources

Figure 10 shows the most significant journals as sources of publications based on the number of times they were cited in the papers. The figure demonstrates the most influential and frequently cited journals as sources for papers.
The chosen relevant sources among the 42 papers involve respectable academic publications, which include Computers & Security, the Internet of Things, the Journal of Information Security and Applications, Computer Communications, Computers and Security, Expert Systems with Applications, the Journal of King Saud University-Computer and Information Sciences, Ad Hoc Networks, Applied Soft Computing, and Computer Networks. These journals are also commonly distinguished because they make significant innovative contributions to cybersecurity, network security, artificial intelligence, and related areas. High-quality research papers, reviews, and theoretical concepts are published by the magazine with works that make profound contributions to societies. Researchers frequently use papers published in these journals because they are recognized to be accurate, the methodology is rigorous, and such research corresponds to areas of interest in cybersecurity and networking.

Figure 13: Trend topics over time.

Figure 14: Word cloud of the titles of the papers.
4.1.3 Words, sources frequency, and trend topic over time

The rarest words of the most frequent words found in the titles or abstracts of scientific papers, sorted by publication year, are represented in Figure 11. Through such visualization, one can deduce which themes are gaining momentum within the field over time and which terms have been most appealing for researchers to
engage with in each particular year. In contrast, Figure 12 represents various sources that served as a research base over time. This visualization provides insight into the research flow and represents the journals and platforms through which research works in different years were disseminated. Figure 13 shows the trend topics analyzed over time, providing an interpretation of the changing subject niches and study focus within the field. Through this visualization, these choices can be revealed, such as recurrent themes, emerging trends,
and evolving research continuing lines, which together provide a holistic overview of the advancement of networks and cybersecurity based on defense against adversarial attacks.

The compilation of 42 papers covers a broad spectrum of issues concerning cybersecurity, ML, and network security. In-depth papers have been written on many topics, including intrusion detection, adversarial attacks, anomaly detection in IoT devices, malware detection, federated learning, and predictive maintenance in smart systems. Technologies include but are not limited to DL, reinforcement learning (RL), adversarial learning, and blockchain technologies. The conclusions of those articles contribute to the development of reliable security strategies for the IoT. The incorporation of ML along with cutting-edge methods, such as
federated learning and blockchain, gives rise to interesting avenues for preventing and protecting against emerging cyber threats.

4.1.4 Word cloud

The word cloud delivers a holistic view of certain topics [69] through the most frequent terms picked from Document titles (Figure 14), Keywords (Figure 15), and Abstracts (Figure 16). The expression “adversarial attack” takes the lead in statistics. Other notable terms that have been used by researchers most frequently include “cybersecurity,” “IoTs,” “defense,” “security,” and “network.” These outcomes represent the main areas and the most relevant concepts that are focused on cyber and network security system protection against hostile attacks and unknown elements, opening more possibilities for further studies in this area.

The 42 papers covered a vast field covering all the cybersecurity viewpoints within the framework of network defense. Some professional fields, such as “computer and network security,” indicate the importance
of keeping IT structures secure. In addition, the two most significant trending topics were “ML” and “adversarial attacks,” which indicated that there was a rapid increase in the number of studies on the use of ML techniques to detect and handle adversarial attacks. In addition to “network,” “IoT” also appears, which can be assumed to focus on ensuring the security of IoT devices and the overall networked infrastructure. “Intrusion detection” or “anomaly detection” is paramount in this context because surveillance systems are fortified to address and address deviant behaviour products of abnormalities. These keywords outline what is typically done by the teams that carry out research to bolster Internet safety and shield against cyber threats that are ever-changing.

4.1.5 Tree map

The traditional approach of depicting hierarchical information, which mostly utilizes structured trees, has some flaws when portraying large and multilayered data since it becomes difficult to visualize such information in the given space constraints. The Tree-Map algorithm had to be designed to solve this problem, and this algorithm renders large trees for thousands of nodes [70].

For the application of this empirical research, Figure 17 shows the tree map generated by using the algorithm. In the higher stratum of the hierarchy, the definitions of words such as “adversarial,” “attack,” “network,” and “detection” are visualized. The words probably depict the broadest subject or anchor point for the research subdomain. The next step involves a visualization that shows further terms such as “IoT” at the lower level of the hierarchy, depicted in a forest plot that demonstrates the hierarchical relationship to the first-level terms.

Owing to the Tree-Map technique, the map covers a wide range of research subjects without causing overcomplication, which is essential for researchers to determine the correlation between various terms and concepts faster and more effectively.

The term “security” is clearly evident in the tree map visualization, symbolizing the central attention given to diverse cyber-security issues. It can refer to various issues such as adversarial, attack, network, and detection. It acts as the main hub in the tree map, which adds attention to the overall importance of the research area in the papers.

4.1.6 Co-occurrence network

Bibliometric studies incorporate co-occurrence networks as one of the main instruments that they investigate. The broad network of key concepts entails the previous terms that have been connected with the analysis, and then the professionals are provided policy-makers and experts with the conceptual structure of such a given area of study [61]. Figures 18 and 19 consist of co-occurrence networks created from titles and abstract articles in the literature.

The relevant terms and concepts in the research cybersecurity papers are linked in the co-occurrence network, and 42 papers reveal the cohesiveness of these terms and concepts. Along with these significant categories, the main categories displayed in the titles and abstracts are “learning,” “detection,” “adversarial,” and “attack.” It can be concluded that these particular terms frequently appear in the same context among the subject matter of the artworks. This implies that most of the period when one is looking for “learning,” they are doing so in the context of “detection,” which shows the wide application of ML methods in building IDSs. In addition, “adversarial” and “networks” are especially connected, which points to the continuous actions taken to develop cybersecurity defenses for fast-changing threats. As a result, the network of terms offers extremely useful data on the main ideas and passages in the topic study of cybersecurity research represented by the 42 articles.

4.1.7 Thematic map

A map of thematic map was also created based on the density and centrality indices to be divided into four topological regions (Figure 20). This conclusion was derived by surveying the papers analyzed in this study with additional relevant keywords.
The cybersecurity thematic map, generated from the 42 papers, illustrates the wide spectrum of topics covered by cybersecurity research. Thematic consists of associated concepts that are spread across abstracts. For instance, one topic may be “ML in Cybersecurity,” which comprises various elements, including anomaly detection, adversarial attacks, and IDSs. There can also be a theme called “IoT Security,” which might cover topics such as vulnerability analysis, anomaly detection in IoT networks, and blockchain-based security solutions for IoT devices. Each group characterizes the complex interrelations among research subjects within the larger framework of cybersecurity themes in the study, thereby clarifying the overall research domain.

4.1.8 Factorial analysis

Factorial analysis assesses similarity by enabling users to standardize bibliographic coupling, co-occurrence, and cocitation measures. It is utilized to chart the conceptual framework of a discipline by analysing word frequency within specific bibliographic clusters [71] (Figure 21).

Factorial analysis of the 42 articles allowed us to study the interrelation and trend among distinguished issues such as research methods and findings. Through examination of the factors listed earlier, discerning trends as well as the interrelations they have with each other can be accomplished. Suppose we realize that texts about “adversarial attacks,” in most cases, take “ML” techniques into account and evaluate “robustness evaluation” as major issues of their analysis. Factorial analysis, in turn, purposefully reveals the backstage structure and tendencies of cybersecurity research in general.

4.2 Network and cybersecurity applications of defense in adversarial attacks: Taxonomy

To create the systematization of the 42 chosen articles, 3 groups were defined to perform a systematic analysis based on evidence derived from studies that satisfied the specified criteria. These research findings can act as major classifications and have been provided with various subcategories for better organization and elucidation during presentation (Figure 22). Under these subdividing processes, the papers that used networks and cybersecurity within adversarial attacks are deeply explored, allowing more thorough discussion about developments and quasiness in this field. Furthermore, in the selected articles, the subcategories “delve to other terms and techniques against adversarial attacks in networks” and “cybersecurity” refer to the subcategories “more subcategories” in this article. The established types consist of \( n = 42 \) contributions, as described below:

1. **Security enhancement techniques**: This category included 16 of 42 contributions (38.1%).
2. **Adversarial attack strategies and defense mechanisms**: This category included 7 of 42 contributions (16.7%).
3. **Innovative security mechanisms and solutions**: This category included 19 of 42 contributions (45.2%).

4.2.1 Security enhancement techniques

In the realm of this category, a notable subset comprises 16 articles out of the 42 chosen articles. Two studies delve into neural network-based malware detectors. Shaukat et al. [72] developed 10 malware classifiers based on neural networks. Among this group of systems, nine were trained to face a specific type of assault, whereas the tenth was not. The approach involved not only the creation of but also the introduction to the defending system of new techniques. Such a mode of training a neural network requires the application of different adversarial strategies to the network. A similar work by Falana et al. [73] used an ensemble of deep convolutional neural network (CNNs) and GANS (known as Mal-Detect) to perform malware analysis, classification, and detection. Two studies delve into this field, but more research is needed because it is a critical area in cybersecurity. Enhancing these detectors can bolster digital defense strategies effectively.

In the field of IDSs, seven studies have been conducted. Sharma et al. [74] proposed a novel anomaly-based IDS system for IoT networks using the DL technique. In particular, a filter-based feature selection DNN model
in which highly correlated features are removed has been presented. In the study by Sethi et al. [75], a multiagent IDS model in which deep Q-networks are engaged by several agents and core attention mechanisms are applied to effectively perceive and categorize sophisticated network intrusions was introduced. Mishra et al. [76] presented a weighted stacked ensemble model that combined bidirectional long short-term memory networks with deep convolutional generative adversarial networks. The main objective of Khan et al. [77] was to examine various forms of CAN-bus traffic data for intrusion attack detection while also taking into account time complexity. In addition, Saheed et al. [78] proposed an IoT defender framework that employs a modified genetic algorithm (MGA) and an LSTM model to detect cyberattacks in IoT networks. The implementation is pioneering because the MGA is used for feature selection, while the GA for the refinement of the LSTM parameters is incorporated within the evolutionary computation framework. Furthermore, Rookard and Khojandi [79] defined RRIoT, which is based on a RL algorithm that works in an adversarial environment. Along with these factors, an LSTM layer is also applied. Here, the performance of the aforementioned approach is evaluated against that of both novel methods and state-of-the-art ML/RL algorithms. Finally, the research conducted by van Geest et al. [80] brought to light a new experiment of simulating bypass attacks on models whose principle of operation is limited to a single analysis. Then, the potential of hybrid methods was explored, and the advantages of these techniques with multiple models are that they compromise confidentiality and effectiveness. These studies further revealed the drawbacks of hybrid models, which largely favour their effectiveness while prioritizing their universality. These studies could improve clarity by condensing lengthy sentences and emphasizing the practical implications of the research findings for cybersecurity practices.

In addition, seven papers on cybersecurity tactics and procedure identification were published. Imran et al. [81] focused on identifying and detecting cybersecurity tactics, techniques, and procedures based on an ML approach. For I-IoT systems, Gungor et al. [82] showed that cybersecurity is a great challenge because of inadequate standardization and the lack of skills required to implement such systems. The goal of Alzahrani and Asghar [83] was to ascertain how IoT frameworks centred around logistics can categorize different threats. Consequently, this study sought to explore the optimal approach for deploying a system to detect cyber vulnerabilities within the data exchange of logistics-based IoT, utilizing historical data. Alshaikh et al. [84] aimed to understand decision makers’ and specialists’ perceptions of ML capabilities for defensive cybersecurity applications. The contributions are made in the following three areas: (1) MLCS capabilities, (2) MLCS implementation, and (3) MLCS evaluation and communication. Aurangzeb et al. [85] aimed to commence a study on deep black box adversarial attacks against smart power grids, demonstrating that statistically significant impacts on a national smart power grid can be achieved with absolute security. In addition, they investigated the detection of cybersecurity attacks on smart power grids. In addition, Nkoro et al.’s [86] methodology endeavours to identify various cyber threat categories that could impact virtual reality (VR) learning platforms by employing a straightforward DNN that offers explanations for its detection. In addition, Aygul et al. [87] investigated the capabilities and possible weaknesses of ML-driven transient stability prediction (TSP) models when confronted with such cyber threats. By tackling the hurdles associated with integrating renewable energy and modifying grids, their objective is to offer valuable insights that can bolster grid security and maintain a dependable power distribution system. From these studies, we infer insufficient practical implementation insights in the study of IoT frameworks, potential oversight of technical feasibility in perception-based analysis in defensive cybersecurity, lack of scalability considerations in the study of adversarial attacks, incomplete coverage of potential threats in the methodology employed for threat identification in VR platforms, and a failure to address broader system vulnerabilities beyond cyber threats in the investigation of ML-driven TSP models.

4.2.2 Adversarial attack strategies and defense mechanisms

This section includes adversarial attack strategies and defense mechanisms, as this section consists of seven contributions out of 42.
Two studies of secure physical systems and interconnection architecture were conducted. Jia et al. [88] devised an attack that bypasses the anomaly detectors and rules checkers of cyber-physical systems (CPSs). Due to the popularity of gradient-based methods, adversarial attacks produce noise in sensor action values, and a genetic algorithm is used to enhance this noise. Ahmed et al. [89] presented a secure mm-wave wireless interconnection architecture for mobile communication management centre systems. Even though wireless interconnects can provide an advantage to off-chip communication in MCMC systems by decreasing single-hop link energy consumption, when they are attacked, for example, by jamming-based DoS attacks, they will offer little protection. Sports the ML algorithm-based firewall and protection schemes our architecture believed to counter both external and internal persistent DoS attacks based on jamming.

Four studies on generating adversarial attacks and model robustness. Dai et al. [90] present negative sampling-based network embedding adversarial training with refinement for subtraction and for textured models that evolve in a specific manner. The central aspect in Pawlicki et al. [91] is to provide a way to counter the attacks on ML due to cyberattacks with a suggestion of the adversarial ML solution for such detection, and after the evaluation of the possibility of reducing the performance of the well-optimized intrusion detection, four methods for attack detection are provided. Furthermore, Duy et al. [92] explored the method of generating adversarial attack samples via the GAN model to develop an IDS. They proposed DIGFuPAS, a framework that can create attack samples that can bypass ML-based IDSs in software-defined networks in a black-box manner. Research on the weaknesses of wireless communication toward remote connections as a means of control in mini-electric AVs was performed [93]. Moreover, the field of disablement of IDSs consisted of only one study. Chen et al. [94] designed an attack model characteristic of IDS evasion – the Anti-Intrusion Detection AutoEncoder – to generate features that can bypass an IDS. The proposed structure works out with an encoding feature, and in the opposite direction, different decoders reconstruct continuous and discrete features. In addition, a GAN is utilized to gather the expansive prior distribution of the latent space. These studies provide insights into adversarial attack strategies and defense mechanisms with many contributions. This article details studies on secure physical systems, wireless interconnection architecture, and adversarial attacks on ML models. However, it lacks clarity due to dense information and could benefit from clearer organization and summarization of key findings.

4.2.3 Innovative security mechanisms and solutions

Within the innovative security mechanisms and solutions category, 19 of 42 papers focused on six subcategories. There were two papers in the field of blockchain and semantic computing networks (SCN). In the study by Mirsky et al. [95], (1) a novel approach for anomaly detection and (2) a lightweight framework that utilizes the blockchain to assemble an anomaly detection model in a distributed environment were proposed. Shi et al. [96] proposed a knowledge-guided SCN constructed with a primary knowledge-guided semantic tree module and an auxiliary data-driven lightweight neural network module. Blockchain and SCNs are crucial areas requiring extensive research. Two studies alone cannot adequately address the complexities and potential of these fields. Further investigation is necessary for comprehensive understanding and development. Blockchain and SCNs are crucial areas requiring extensive research. Two studies alone cannot adequately address the complexities and potential of these fields. Further investigation is necessary for comprehensive understanding and development.

Four contributions to privacy preservation and adversarial robustness. Bai et al. [97] proposed detecting evidence of LM using ML and Windows RDP event logs. They explored different feature sets extracted from these logs and evaluated various supervised ML techniques for classifying RDP sessions with high precision and recall. Chen et al. [98] demonstrated that the multiexit network can reduce the impact of adversarial perturbations by outputting easily identified samples at early exits. Therefore, it can improve the adversarial robustness. Furthermore, the multiexit network can prevent catastrophic overfitting in single-step adversarial training. In the study by Roshan and Zafar [99], the objective was to introduce a dual-phase defense strategy
against the potent optimization-based adversarial attack known as Carlini & Wagner (C&W). These two defense phases consisted of training and testing phases. Through the training phase, a modified adversarial training approach employing Gaussian data augmentation is utilized. In the testing phase, the feature squeezing technique is applied to the generated list of adversarial samples before they are fed into the resilient NIDS model for the ultimate classification. In the study by Sharma et al. [100], a GAN architecture called MIGAN was proposed for the generation of malware images. This system has the advantage of being able to generate malware images of very high quality and sorting out the malware samples according to families. One potential downside of these studies is their focus on specific techniques or approaches, which do not fully address the diverse range of adversarial threats and privacy concerns present in complex real-world scenarios. In addition, the effectiveness of these methods in practical applications outside controlled environments may vary.

In addition, there are five studies in the field of federated learning security optimization. Wan et al. [101] proposed integrating blockchain-enabled FL with Wasserstein generative adversarial network (WGAN)-enabled differential privacy (DP) to protect the model parameters of edge devices in 5G networks. QQFL is a newly introduced federated learning scheme that optimizes hyperparameters by employing various adversarial attacks in AV settings [102]. In the study by Ahmad and Shah [103], the goal was to identify attacks while safeguarding the privacy of IoT users. To achieve this goal, they pursued a lightweight mini-batch federated learning approach that is computationally efficient and demands only a minimal number of federation rounds to detect malicious activity in an IoT network. FedGenID is a novel and highly valuable federated generative IDS, for the protection of industrial Internet of thing (IIoT) networks that was proposed by [104]. Compared to the existing attacks, the FedGenID assessment based on the sophisticated industrial cybersecurity dataset reveals that it is capable of detecting class imbalances and multiclass cyberattacks. In addition, its ability to remain stable against adversarial attacks is proven. Bukhari et al. [105] proposed a new SCNN-Bi-LSTM model for intrusion detection in wireless sensor networks (WSNs). This model is based on FL, which enables better intrusion detection performance and privacy. The FL-based SCNN-Bi-LSTM model applies a novel strategy in which multiple sensor nodes coordinate when training a global model without disclosing private data; as a result, privacy issues are resolved. While these studies have made progress in federated learning security optimization, they exhibit limitations. They primarily focus on specific aspects of security without considering the broader spectrum of potential threats. In addition, their efficacy in real-world scenarios with diverse network architectures and data distributions needs further validation.

The category of improving the detection of unknown attacks consists of four papers. Roshan et al. [106] studied important aspects related to NIDS, adversarial attacks, and defense mechanisms to increase the robustness of ML- and DL-based NIDS. In the study by Nguyen and Le [107], the research aimed to address limitations in detecting unknown attacks and provide better protection for IoT networks against DoS/DDoS attacks. The study by Xie and Chen [108] strives to offer a streamlined yet potent approach to intrusion detection, designed to operate efficiently even in settings with restricted computational and energy capacities. Liu et al. [109] focused on SeMalBERT, which is an adaptive malware detection model used for detecting malicious software in Windows-based systems. It trains the features on the utilization of API function sequences as learned features. For word representation, BERT is being used, as is semantic chaining. Moreover, CNNs and LSTMs can be used to explore chaining relationships. Moreover, an outgoing attention mechanism makes a model stay focused on the relevant information in the text. While these studies aim to enhance the detection of unknown attacks, they have some shortcomings. These methods may lack comprehensive validation in real-world environments, potentially limiting their applicability. In addition, their focus on specific aspects of intrusion detection may overlook broader security concerns and fail to address evolving threats effectively. While these studies aim to enhance the detection of unknown attacks, they have some shortcomings. These methods may lack comprehensive validation in real-world environments, potentially limiting their applicability. In addition, their focus on specific aspects of intrusion detection may overlook broader security concerns and fail to address evolving threats effectively.

Furthermore, there are three papers on advancements in security. In the study by Ardito et al. [110], the study aimed to encourage the implementation of ML security models in the context of smart grids. In the study by Duy et al. [111], the authors explored the ability of the WGAN with gradient penalty (WGAN–GP) by
generating perturbed attack samples to bypass attack detectors. This method can be used for regular assessment of the robustness of ML-based IDS in software-defined networks (SDN) to achieve the objective of upgrading it as a service in SDN. Albahri et al. [112] presented a fuzzy MCDM structure incorporating multiple perspectives of data fusion. Its purpose is to evaluate diverse ML models to quickly detect adversarial attacks in VANETs. This implies reckoning dedicated short-range communication systems as well as developing multi-ML models by applying standard and feature fusion preprocessing to the data and finally evaluating and benchmarking these models with fuzzy decision-making logic by using FDOSM (fuzzy decision by opinion score). In this field, authors can have the flexibility to explore various DL approaches beyond WGAN and MCDM, allowing for a diverse range of methodologies to address security challenges.

However, in the field of automatic generation of imperceptible adversarial examples, there was one contribution. Marchisio et al. [113] suggested a “greedy” algorithm for the automatization of imperceptible adversarial samples when an attacker has limited feedback. The limited scope of research on the automatic generation of imperceptible adversarial examples poses a critical drawback. Relying solely on one paper may hinder a comprehensive understanding and development of effective defense mechanisms against evolving adversarial threats.

5 Discussion

This section focuses on three key aspects related to networks and cybersecurity in adversarial attacks: motivations, challenges, and recommendations.

5.1 Motivations

This section addresses four main topics related to the motivation for networks and cybersecurity in adversarial attacks: (1) Improving IoT security, (2) enhancement of IDSs, (3) cyber-physical system security, and (4) general ML security and defense strategies (Figure 23).

Figure 23: Motivations of networks and cybersecurity in adversarial attacks.
5.1.1 Improving IoT security

The IoT is the aggregation of the numerous individual physical devices currently called things that form a network. The connected products in the network are constrained by limited processing power and memory storage resources. As the number of IoT heterogeneous physical devices, through which the internet is accessed, continues to increase, the amount of data generated also becomes enormous; therefore, IoT networks have become more lucrative targets for attackers [74]. The study by Gungor et al. [82] showed that computational systems in the IIoT are usually not designed with security in mind. Their limited computational power creates security vulnerabilities that attackers can exploit to prevent asset availability, sabotage communication, and corrupt system data. In addition, Alzahrani and Asghar [83] mentioned the need for a cyber vulnerability detection system in logistics-based IoT data exchange. The sharing of IoT data with the cloud data centre may affect the privacy of the user’s sensitive data [103]. In the study by Nkoro et al. [86], the swift incorporation of the Internet of artificial intelligence and IoTs (AI-IoT) technologies has brought forth a crucial aspect of the forthcoming digital age: the Metaverse. Furthermore, with the rapid development of technology, cyber threats and privacy issues always arise before the IIoT. To solve this problem, there is a great demand for an advanced sequence detection system for the protection of IIoT networks [104]. The articles seem to assert the necessity for cyber vulnerability detection systems, protective measures, and the integration of novel technologies without thoroughly exploring the intricacies of these subjects. Such broad generalizations could undermine the credibility of the arguments and fail to sufficiently address the complexities inherent in the discussed issues.

5.1.2 Enhancement of IDSs

Regarding the enhancement of IDSs [75] reported that there was a need to design IDSs for networks. The implementation of a sophisticated adversary assault against an IPS may lead to a failure of detection; hence, this can be viewed as a direct problem in the application of ML models in IDS [91]. To evaluate and improve the weaknesses of IDSs against the latest attack methods, adding traffic is necessary [92]. In addition, the primary goal of Mishra et al. [76] is to create the best possible arrangement for the detection of security breaches. The optimal selection of the activation function, optimization process, epoch count, and batch size are the main topics of this study. In the realm of intrusion detection [108], there is a research void concerning the effectiveness and flexibility of such models in environments with limited resources. On the other hand, despite the immense growth of the digital sphere nudged by innovative technologies, cybersecurity vulnerabilities have reached a high level of importance. Accordingly, IDSs are the cardinal parts of network security systems that help in identifying abnormal network traffic, which may be indicative of malicious activity [105]. In addition, automated antiphishing detection systems have become a non-negotiable necessity, given that cybercriminals are continuing to innovate their methods [80].

5.1.3 Cyberscale physical system security

CPSs of critical infrastructure are exposed to a series of threats, and this has prompted investigations into the different detection techniques that can be deployed in such environments, including the use of neural network-based anomaly detection systems [88]. Research has shown that intelligent transportation systems, namely, AVs, are gaining stronger influence and that their power to transform the modern world is enormous. However, the security and privacy of these systems should be protected [102]. The controller area network protocol serves as a vital communication mechanism in vehicular systems. Nevertheless, its extensive utilization has led to vulnerabilities in in-vehicle communication channels, rendering them prone to diverse security risks such as denial-of-service, fuzzy, and impersonation attacks [77]. The rise in successful cyberattacks raises doubts about the efficacy of ML in cybersecurity applications [84]. There have been numerous concerns raised about the vulnerability of smart grid technologies to hacking [85]. In addition, cyberattacks have the potential
to induce inaccurate forecasts, resulting in power outages, as shown by Aygul et al. [87]. TSP models must be engineered with robustness to withstand such cyberattacks. As a result of a global increase in cyberattacks occurring through computer systems, specifically IoT devices, the need for robust and automated methods to discover and mitigate these attacks in real time is urgent and sharp, and these techniques are fundamental for that purpose [79].

5.1.4 General ML security and defense strategies

Supervised learning models, even though they are being used to identify threats and address them, are being challenged in regard to detecting unfamiliar types of attacks, which could be very serious [107]. Mirsky et al. [95] showed that anomaly detection models need to be trained to consider all legitimate behaviours and actions. Moreover, the models are vulnerable to adversarial attacks because they presume that all observations during training are harmless [98]. It is observed that compared to normal training, adversarial training takes less time to achieve resilience against attacks, but it is limited by the number of epochs, resulting in suboptimal performance. In addition, ML has advanced at a great pace, and it will be necessary to study adversarial attacks and defense mechanisms to mitigate multiple cybersecurity problems [106]. The omission of the intuitive consideration of both normal and adversarial attack perspectives during the ML model generation stage, the absence of various preprocessing methods for VANET communication data, the insufficiency of the selection criteria for real-time adversarial attack detection models, and the minimum focus on explainability in adversarial attack detection procedures inspired the research carried out by Albahri et al. [112].

5.2 Challenges

This section discusses the difficulties in four groups (Figure 24).

5.2.1 Adversarial attacks and security vulnerabilities

DL malware detectors are being implemented as one of the solutions to malware detection challenges. Although no system is completely secure, these detectors are at risk of being compromised by an adversary’s assault [72]. In regard to the challenges associated with the IoT, resource-constrained peripheral devices (i.e.,
those with low computational and storage capacity) and security (i.e., those for which it is more difficult to implement necessary security measures) both make implementation more difficult [74]. Although well adapted in disciplines such as images and audio, adversarial attacks turning on CPSs prove to be difficult to implement due to the presence of other built-in defense mechanisms such as rule checkers (or invariant checkers) [88]. As DNNs are widely used in various challenging ML tasks in real life, the threat of adversarial examples is receiving particular attention in the DL community [98]. In Khan et al. [77], two notable challenges must be addressed. First, the evaluation lacks exploration of diverse real-world scenarios, limiting the understanding of the robustness and generalizability of DivaCAN beyond the experimental setup. Second, practical implementation in real vehicular systems requires the consideration of hardware constraints to ensure efficient real-time operation without compromising vehicle performance.

Currently, the problem of securing IoT technologies has become very significant because of the increased complexity, imprecise settings, and conflicts between old and new systems. Such problems heavily affect the availability and reliability of existing essential infrastructure [78]. In addition, Nkoro et al. [86] reported that the increasing dependence on the Metaverse has led to the recognition of the critical importance of implementing strong cybersecurity measures. This is essential for identifying and addressing cyber threats effectively, thus safeguarding user safety. In addition, the challenge addressed in the study by Aygul et al. [87] is accurate online transient stability prediction in modern power systems that are increasingly dependent on smart grid technology and are susceptible to cyberattacks. Furthermore, automatic driving systems, whose main goal is to enhance the safety and comfort of passengers, simultaneously become prone to attack by hackers or other malicious actors because they are dependent on network technologies and require remote connections [93]. The systems of IDS can have either a hardware or software basis. Nevertheless, classic IDS schemes are not always able to efficiently perform accurate information security tasks and identify sophisticated, rare types of cyberattacks, particularly in WSN environments [105]. These articles highlight challenges in securing IoT, metaverse power systems, and autonomous driving due to the complexity, cybersecurity risks, and limitations of existing IDSs, necessitating innovative solutions.

### 5.2.2 Cybersecurity datasets

By far, the greatest hurdle for creating advanced ML-based fault and security defensive systems is the current scarcity of significant datasets on smart electrical grids, which provide both a wide category of such grids and their characteristics [110]. Labelling such a large dataset is very challenging since it can be performed only by an expert trained for at least 5 years [96]. Nguyen and Le [107] outlined three existing security challenges with IoT networks in terms of DoS/DDoS attacks. The first problem is related to the fact that the state-of-the-art popular datasets for IDS assessment develop certain restrictions. The second challenge is that the detection of new attacks should not require any training data. The third point is adversarial attacks that can allow the attacker to use vulnerabilities of the ML/DL-based NIDS to bypass the security mechanism. A significant obstacle in the study by Mishra et al. [76] lies in the absence of balance within real-time data, posing difficulties in adapting DL techniques to function effectively as real-time classifiers for detecting attacks when dealing with imbalanced data. According to Ahmad and Shah [103], increased training time and computational expenses are challenges in an IoT network. The articles highlight significant data-related challenges, so they underscore the scarcity of large, labelled datasets for ML in smart grids and IoT networks, hindering the development of effective security systems.

### 5.2.3 Complexity and evolution of attacks

The design of an efficient network intrusion system (IDS) is a serious problem considering the large number of new emerging attacks and complex network applications [75]. Identifying attacks can prove to be very difficult because one attack may develop in several steps, each of which is created to hinder the defense and at a separate timeline [97]. In Falana et al. [73], the challenge is that malware has become more prevalent given the
complexity of today's network and the attack landscape. Computer users and business owners find it very difficult to keep pace with cybercriminals.

5.2.4 Robustness of ML models

With the application of ML-based solutions and AI technologies in almost every sphere of present-day life, several issues regarding this pursuit have appeared as well. The immediate problem relates to adversarial attacks. In a recent study, algorithms, which are the most commonly used data-driven mechanisms in many intelligent systems, have become the targets of many attacks [91]. In the same vein, SDNs, when employed with IDSs, take advantage of the centralized control plane in an SDN to support massive-scale monitoring in a network. Nevertheless, ML-based IDSs can be overlooked and fooled by adversarial examples with the addition of natural perturbations to the original IDSs [92]. The proposed system in Alzahrani and Asghar [83] exhibits various limitations, such as (i) relying solely on one statistical method, the chi-squared measure, to identify significant features (predictors) and (ii) failing to leverage pretrained DL models. In addition, Xie and Chen [108] possess the potential vulnerability to novel or hostile attacks that were not encountered during the training phase, and the complexity of the CSNN model increases to accommodate a wider range.

Gungor et al. [82] demonstrated how cyberattacks can have a dramatic effect on the efficiency of ML methods aimed at process diagnosis and monitoring (PDM), and the results were up to 120× less efficient. Thereafter, they worked on constructing a multiple-layer combination learning environment that is unchanging against an assortment of various white-box adversarial attacks. Duy et al. [111] were exposed to two problems. First, the threat should not cause any other technologies to respond in an interlocked security system. Next, the initiator must be able to cause a given output by receiving corresponding feedback. Along these lines, it is important to mention that although federated learning is one of the most suitable mechanisms for data privacy, and it is prone to various attacks as well, especially data poisoning attacks where the adversary adds vectors in the training phase [102]. Even though hyperparameters are very useful in constructing an efficient federated learning model, they are robust against possible sideline events. The convolution layers of CNNs are not able to retain hierarchical spatial relationships such as the orientations, positions, and scaling of objects. CNNs tend to locate an object of interest via featurewise pattern recognition, not by deducing spatial relations within larger spatial structures [113].

On the other hand, [85], given the computational capabilities of quantum computers, existing encryption techniques are under significant threat, and it is only a matter of time before secure cryptography is

Figure 25: Recommendation categories for networks and cybersecurity in adversarial attacks.
compromised. In addition, the ML technique has shown potential in the detection of some types of attacks. However, despite its reasonable success in subsequent assaults, this approach does not provide an ideal solution [79]. In addition, a variety of current systems utilize a single-analysis model structure, resulting in weak points and, therefore, easier targets for hackers [80]. These models can be vulnerable to adversarial examples that are dependent on a single statistical technique and are incapable of accurately predicting complicated attacks. Although ML reportedly produces positive outcomes in the detection of attacks, its effectiveness is still inadequate, especially because of the danger of a mono-variate model system structure.

5.3 Recommendations

This section discusses the recommendations and future directions for researchers in the field of networks and cybersecurity in adversarial attacks (Figure 25).

5.3.1 Defense and protection against adversarial attacks

This category addresses the difficulties and potential of using ML in adversarial environments, such as malware detection, network intrusion detection, and failure prediction. In Shaukat et al. [72], the authors envisaged several future research directions. For instance, it would be useful to investigate the performance of DL models other than the models proposed by Shaukat et al. [72] in adversarial settings. In addition, it would also be useful to examine the robustness of malware detectors trained against other evasion attacks. The proposed model by Sharma et al. [74] can be applied to other datasets with imbalanced class labels, and GANs can be adopted to generate traffic from minority attacks in the dataset. In addition, other minority attack classes can be categorized into new DNN-based classifiers with increased accuracy and reduced loss in terms of false negative (FN) and false positive (FP) predictions. Duy et al. [92] recommended conducting the framework on more datasets with diverse modern attacks, considering other more complex detection algorithms, and using other GAN formulations for training improvement. Ardito et al. [110] highlighted that defending these systems against alternative adversarial training and detection techniques would require more nuanced and in-depth research, which they hope to pursue in the future work. Another interesting future direction is to consider the privacy of fault-prediction systems such that separate zones do not need to exchange their data with a central server.

Chen et al. [98] explored other potential causes of catastrophic overfitting and discovered more properties of adversarial examples in DNNs. Wan et al. [101] recommended simultaneously improving the learning efficiency of non-independent and identically distributed (NIID) dataset data and further improving the robustness of the adversarial training process. Roshan et al. [106] explored the implications of adopting the described method on other kinds of ML and DL architectures. In the same context, there is a possibility to examine the transferability concept in the field of adversarial ML, where the approach that works in one case can also be developed for other purposes. The method should be used in the series in the same way, which also addresses the concept of drift network streaming data-based NIDS system.

On the other hand, in the ever-changing arena of adversarial attacks, forthcoming studies need to delve into the consequences of emerging attack methodologies, encompassing white-box, black-box, and gray-box approaches [99]. Xie and Chen [108] focused on mitigating constraints by exploring methods to enhance model resilience, decrease computational intricacy, and augment detection capabilities tailored to particular attack types. Nkoro et al. [86] recommended exploring the computational complexities of the SHAP and LIME XAI methods and investigating the integration of an explainable adversarial protection mechanism to fortify the NIDS against potential DNN adversarial attacks. Albahri et al. [112] moved to a stage that demands rigorous testing and validation. This involved the setting of boundaries of operation, such as attacks and examination of frameworks regarding stability. The performance metrics of framework development should focus on issues such as the amount of data and model complexity, and the real-world benefit of the implemented framework.
should also be evaluated. Sharma et al. [100] introduce a hybrid approach for Windows Malware Classification. Despite this, an area that is still under debate is the reliability of the model against adversarial attacks. In addition, the model should be checked for the ability to identify malware created for other platforms, such as Android or Linux [100]. A study conducted by Hamouda et al. [104] suggested exploring ensemble learning methods for collective decision-making and self-supervised learning techniques to improve the capabilities of generative models. Further investigations can focus on developing complex feature-selecting methods that perceive subtle attack strategies to be detected by the model. In addition, the model can also evolve to be more efficient, especially in real-time detection scenarios, which is another area that needs to be emphasized [105]. Finally, van Geest et al. [80] proposed an extended duration trial of the framework including more models, complex bypass simulations, and the integration of existing models to refine the hybrid design. There is a need for broader investigations into ML model performance in adversarial settings, including robustness against various evasion attacks and diverse attack types.

5.3.2 Data generation and augmentation

The studies in this category are concerned with the methods and use cases of creating artificial or enhanced data employing methods such as RL and GANs. Gungor et al. [82] were the first to plan to add black-box attack methods that do not have any knowledge about the attacked models. Duy et al. [111] recommended RL by being able to learn the behaviour of modifying network traffic; hence, this approach can be used to obtain adversarial network traffic at the byte level. The suggestion put forward by Mishra et al. [76] involves integrating temporal features and improving benchmark datasets. In addition, there are plans to enhance the proposed model by integrating federated learning methodologies, which can address concerns regarding data confidentiality and privacy more effectively. Alzahrani and Asghar [83] aimed to assess its performance in the future using more datasets. (ii) In addition to chi-squared analysis, they investigated additional feature selection techniques. (iii) Currently, each attack class is categorized separately. (iv) Furthermore, researchers looking into logistics-based IoT vulnerability screening may find it helpful to use the BoT-IoT dataset in combination with hybrid DL. However, there is a need for further refinement and diversification in data generation and augmentation methods.

5.3.3 IoT security and analysis

This category includes the papers that address the security and analysis of IoT devices and systems – malware, threat case generation, framework design, and log analysis. Imran et al. [81] recommended that future work aims at developing use cases based on different log outputs and deducing exact results by using ML algorithms that will remove false positives. The proposed framework by Mirsky et al. [95] has the potential to provide IoT manufacturers with a cheap and effective solution. They hope that this framework, and its variants, will assist researchers and the IoT industry in securing the future of the IoT. Bai et al. [97] evaluate the developed approach to other session-based protocols, such as Secure Shell. In addition, Windows event logs contain a variety of event types that can be leveraged to identify different stages of an APT (advanced persistent threat) attack. Finally, in the study by Falana et al. [73], MalDetect was tested against larger datasets, and the proposed framework was integrated into an IoT-based system for assessment and precision. In the study by Ahmad and Shah [103], further assessments were conducted to evaluate the computational complexity and effectiveness of the proposed model across various IoT application scenarios. In addition, they recommend exploring the implementation of an efficient data compression technique to minimize the volume of data transmitted to a cloud data centre in a federated learning framework. These studies need more comprehensive evaluation and practical implementation of the proposed frameworks in real-world IoT environments. Recommendations include developing use-case scenarios based on different log outputs, refining frameworks to ensure affordability and effectiveness for IoT manufacturers, and conducting further assessments to evaluate computational complexity and effectiveness across various IoT application scenarios. In addition, suggestions involve
exploring efficient data compression techniques to minimize data transmission volumes in federated learning frameworks. However, there is a lack of discussion on potential challenges and limitations that may arise during implementation and operation in diverse IoT settings.

5.4 Gaps, open issues, and some innovative key solutions

Constructive criticism within a literature review serves as a crucial component of scholarly discourse, offering a balanced assessment of the strengths and weaknesses of the existing body of literature. By critically evaluating the methodologies, findings, and theoretical frameworks employed in previous studies, researchers can identify gaps and opportunities for advancement within their field of study. In this context, constructive criticism is not intended to disparage or undermine the contributions of previous researchers but rather to foster intellectual growth and promote academic rigor. It offers an opportunity to reflect on the limitations of existing research, thereby laying the groundwork for more robust and comprehensive investigations in the future. Moreover, constructive criticism can stimulate dialog and collaboration among researchers, encouraging them to exchange ideas and perspectives in the pursuit of shared research goals. By fostering an environment of constructive critique, scholars can collectively contribute to the advancement of knowledge and the development of innovative solutions to pressing challenges. Therefore, constructive criticism in a literature review plays a pivotal role in shaping the trajectory of academic inquiry, and guiding researchers toward more nuanced, inclusive, and impactful research agendas. It serves as a catalyst for intellectual growth, challenging researchers to critically assess existing paradigms and explore new avenues for exploration and discovery.

This section attempts to map the gaps within the current scholarly discussion to provide a platform for further exploration by upcoming researchers. Within each subtopic given, the description revolves around the mentioned gaps, highlighting the situation where networks and cybersecurity are interrelated, especially when they follow adversarial attacks.

As a general outline, the issue of the application of defense for networks and cybersecurity in adversarial attacks seems to be a very hot and fertile topic within the corresponding academic field. Comparing the number of published articles for the past 4 years with those of the first quarter of 2024 indicates a substantial and rather remarkable increase, as shown in Figure 26. This influx of publications is therefore a great testimony to the fact that the field is increasingly gaining significance and is expected to have a substantive impact. This manifestation reflects increased scholarly interest and important breakthroughs and achievements in this sphere.

Figure 26: Comparison of the number of published studies for the past 4 years with the first quarter of the year 2024.
### Table 1: The dataset of networks and cybersecurity in adversarial attacks

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Dataset name</th>
<th>Dataset size</th>
<th>Dataset link (if available)</th>
<th>Public or private</th>
</tr>
</thead>
<tbody>
<tr>
<td>[72]</td>
<td>VirusShare</td>
<td>19,000</td>
<td><a href="https://virusshare.com/">https://virusshare.com/</a></td>
<td>Public</td>
</tr>
<tr>
<td>[73]</td>
<td>VXHeaven</td>
<td>19,000</td>
<td><a href="http://vx.netlux.org/">http://vx.netlux.org/</a></td>
<td>Public</td>
</tr>
<tr>
<td>[77]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Private</td>
</tr>
<tr>
<td>[78]</td>
<td>SWaT</td>
<td>496,800</td>
<td><a href="https://itrust.sutd.edu.sg/dataset/">https://itrust.sutd.edu.sg/dataset/</a></td>
<td>Public</td>
</tr>
<tr>
<td>[79]</td>
<td>WADI</td>
<td>449,909</td>
<td></td>
<td>Public</td>
</tr>
<tr>
<td>[80]</td>
<td>Cora</td>
<td>2,708</td>
<td></td>
<td>Public</td>
</tr>
<tr>
<td></td>
<td>Citeseer</td>
<td>3,264</td>
<td></td>
<td>Public</td>
</tr>
<tr>
<td></td>
<td>Wiki</td>
<td>2,363</td>
<td></td>
<td>Public</td>
</tr>
<tr>
<td></td>
<td>CA-GrQc</td>
<td>5,242</td>
<td></td>
<td>Public</td>
</tr>
<tr>
<td></td>
<td>CA-HeptT</td>
<td>9,877</td>
<td></td>
<td>Public</td>
</tr>
<tr>
<td>[90]</td>
<td>CICIDS2017</td>
<td>As explained in the study of [75]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[91]</td>
<td>NSL-KDD</td>
<td>As explained in the study of [75]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CICIDS2018</td>
<td>—</td>
<td><a href="https://registry.opendata.aws/cse-cic-ids2018/">https://registry.opendata.aws/cse-cic-ids2018/</a></td>
<td>Public</td>
</tr>
<tr>
<td>[92]</td>
<td>NSL-KDD</td>
<td>—</td>
<td>10.1109/CISDA.2009.5356528</td>
<td>Private</td>
</tr>
<tr>
<td></td>
<td>UNSW-NB15</td>
<td>—</td>
<td>10.1109/MilCIS.2015.7348942</td>
<td>Private</td>
</tr>
<tr>
<td></td>
<td>CICIDS2017</td>
<td>As explained in the study of [75]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[94]</td>
<td>LANL</td>
<td>222,692</td>
<td></td>
<td>Public</td>
</tr>
<tr>
<td>[95]</td>
<td>MaleVis, Malimg, and Virushare</td>
<td>2,000, 1,744, and 9,339</td>
<td></td>
<td>Private</td>
</tr>
<tr>
<td>[97]</td>
<td>CICIDS2017</td>
<td>As explained in the study of [75]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[99]</td>
<td>MNIST and FashionMNIST</td>
<td>70,000</td>
<td>10.1109/5.726791, 10.48550/arXiv.1708.07747</td>
<td>Public</td>
</tr>
<tr>
<td>[101]</td>
<td>BoT-IoT</td>
<td>73 million samples and 80 features</td>
<td><a href="https://ieeexplore.ieee.org/abstract/document/9656911/">https://ieeexplore.ieee.org/abstract/document/9656911/</a></td>
<td>Public</td>
</tr>
<tr>
<td></td>
<td>CIC-IDS-2017</td>
<td>As explained in the study of [75]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CIC-IDS-2018</td>
<td>As explained in the study of [75]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[104]</td>
<td>Edge-IoT, CICIoT, and UNSW-NB15</td>
<td>—</td>
<td>BoT-IoT, and UNSW-NB15 was explained above in [76], 10.1109/MilCIS.2015.7348942</td>
<td>Public</td>
</tr>
<tr>
<td>[105]</td>
<td>Edge-IoTset</td>
<td>—</td>
<td>10.1109/MilCIS.2015.7348942</td>
<td>Public</td>
</tr>
<tr>
<td>[106]</td>
<td>Edge-IIoTset</td>
<td>—</td>
<td>It was explained above in [76], 10.1109/ACCESS.2022.3165809</td>
<td>Public</td>
</tr>
<tr>
<td>[107]</td>
<td>CIC-DDoS2019</td>
<td>107,764</td>
<td></td>
<td>Private</td>
</tr>
<tr>
<td>[108]</td>
<td>BOT-IoT</td>
<td>As explained in the study of [76]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[109]</td>
<td>CIC-DDS2018 CIC-DDS2019</td>
<td>As explained in the study of [92], As explained in the study of [99]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[110]</td>
<td>ToN-IoT</td>
<td>As explained in the study of [76]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[111]</td>
<td>EdgeIIOT, CICCiot, and UNSW-NB15</td>
<td>—</td>
<td><a href="http://dx.doi.org/10.1109/ACCESS.2022.3165809">http://dx.doi.org/10.1109/ACCESS.2022.3165809</a>, It was explained above in [76]</td>
<td>Public</td>
</tr>
<tr>
<td>[113]</td>
<td>BoT-IoT, UNSW-NB15, and N-BaltoT</td>
<td>—</td>
<td>(BoT-IoT, and UNSW-NB15) was explained above in [76], 10.1109/MilCIS.2015.7348942</td>
<td>Public</td>
</tr>
<tr>
<td>[114]</td>
<td>Edge-IoTset</td>
<td>—</td>
<td>10.1109/ACCESS.2022.3165809</td>
<td>Public</td>
</tr>
</tbody>
</table>

(Continued)
An overview of such recent advances in networks and cybersecurity as adversarial attacks is presented in the following subsections with significant tables and analyses.

### 5.4.1 Available datasets

This approach contributes to the basic function of training AI models in the case of adversarial attacks [114]. In particular, within the adversarial attack setting, the dataset turns out to be fundamental for identifying subtleties and for pinpointing possible dangers. This impedes the evaluation of model applicability and generalizability due to the lack of specific details on the datasets, as is evident from the literature review. The nature of the data type, size, composition, and aspects specific to the training and test datasets are pivotal parts of the data used to evaluate the robustness of adversarial attacks in AI models (Table 1).

The validity and homogeneity of the training data are further emphasized, which is what validation provides, as well as the depiction of reality and consistency for successful generalization across a variety of cases [115]. It is critical to ensure transparency in the revelation of the source of the dataset, whether it is available from public sources or has been collected in the field of research on adversarial attacks; the replicability and validity of research findings depend on it.

The total number of datasets deployed throughout all the experiments is 37. The total number of datasets that are used for the development of taxonomy-related portions of the study attests to the fact that many of them still deserve further analysis. The high number of datasets applied and the discrepancy in the quantity of the studies being conducted throughout this period were remarkable because for the period from 2020 to 2023, there were 23 studies, and during only the first quarter of 2024, 19 studies were conducted. For the sake of deriving future research options in the context of past analysis, we performed an intersection analysis taking the two aforementioned intervals (Figure 27).

**Table 1: Continued**

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Dataset name</th>
<th>Dataset size</th>
<th>Dataset link (if available)</th>
<th>Public or private</th>
</tr>
</thead>
<tbody>
<tr>
<td>[109]</td>
<td>EMBER</td>
<td>1,100,000</td>
<td>10.48550/arXiv.1804.04637</td>
<td>Public</td>
</tr>
<tr>
<td>[80]</td>
<td>—</td>
<td>—</td>
<td><a href="https://zenodo.org/records/8358925">https://zenodo.org/records/8358925</a></td>
<td>Public</td>
</tr>
</tbody>
</table>

**Figure 27:** The datasets used for the period from 2020–2023 vs the datasets used in the first quarter of 2024.
Such meticulous observations are graphically represented in Figures 28 and 29, as they are a library of datasets for future studies. Through this visualization, the first figure shows the correlation between the datasets used in the research from 2020 to 2023, while the second figure shows the datasets used for the first quarter of 2024. Therefore, these graphs highlight the pursuing trend for further research in this specific sector. This connection shows datasets that at times are scanned across multiple studies, indicating their ongoing value and appropriateness for more research in the future.

Figure 28: Datasets used vs unused for the period from 2020 to 2023.
The examination of the datasets utilized versus those left unused in comparison with the first quarter of 2024 revealed a microcosm of the evolution of the cybersecurity research landscape. Initially, researchers mainly used available datasets such as NSL-KDD, CICIDS2017, and UNSW-NB15 to investigate various branches of intrusion detection, anomaly detection in IoT networks, and attacks for ML models. On the other hand, the transition of dataset utilization to more complex and specific data is evident in 2024. This suggests that people are beginning to acknowledge that more different and particular data may have to be used in addressing emerging cybersecurity challenges. This transition implies a shift toward more advanced implementation in

*Figure 29: Datasets used vs unused datasets in the first quarter of 2024.*
the field, as researchers have reviewed and explored other advanced methods, such as federated learning, quantum-based strategies, and ensemble methods, for the purpose of boosting the reliability and resiliency of cybersecurity systems. To fill the gap between the dataset used and that unused and solve the challenges in the taxonomy, the vital solutions include the promotion of the diversity of the dataset that will be achieved by collaboration with industrial partners and cybersecurity organizations, standardizing the protocol for the sharing of the dataset that would stimulate the use, leveraging synthetic data generation techniques to supplement the existing datasets, and conducting longitudinal studies to observe the changes in cyber threats.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Methods</th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>[72]</td>
<td>NN</td>
<td>N/M</td>
</tr>
<tr>
<td>[73]</td>
<td>CNN and GAN</td>
<td>Accuracy</td>
</tr>
<tr>
<td>[74]</td>
<td>DNN, GAN</td>
<td>Accuracy</td>
</tr>
<tr>
<td>[81]</td>
<td>RF, MLP, SVM, AB, SGD, GBC, RNN, GRU, LSTM</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[82]</td>
<td>RNN, LSTM, BLSTM, GRU, BGRU, CNN, WAVE, CLSTM, CGRU, and GLSTM</td>
<td>Mean</td>
</tr>
<tr>
<td>[88]</td>
<td>RNN, SVM, KNN, DT, Thresh</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[89]</td>
<td>AdvT, DEEPWALK, LINE, and node2vec</td>
<td>Accuracy</td>
</tr>
<tr>
<td>[90]</td>
<td>ANN</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[91]</td>
<td>GAN</td>
<td>N/M</td>
</tr>
<tr>
<td>[92]</td>
<td>CNN, CapsNet, and SCN</td>
<td>Accuracy</td>
</tr>
<tr>
<td>[93]</td>
<td>CNN, LSTM, LSTM + CNN, BiLSTM, LSTM-CNN, CNN-RNN, LSTM-CNN</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[76]</td>
<td>DCGAN + Bi-LSTM</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[96]</td>
<td>Neural Network</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[97]</td>
<td>LGBM, RF, ET, CNN, DNN</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[113]</td>
<td>CNN</td>
<td>N/M</td>
</tr>
<tr>
<td>[77]</td>
<td>CNN, LSTM, LSTM + CNN, BiLSTM, LSTM-CNN, CNN-RNN, LSTM-CNN</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[108]</td>
<td>CSNN</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[103]</td>
<td>MLP</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[85]</td>
<td>SVM, KNN, NB, Quantum Hybrid Voting</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[86]</td>
<td>DNN</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[87]</td>
<td>LightGBM, LSTM, MLP, SVM, RF, and KNN.</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[112]</td>
<td>KNN, RF, GB, MLPNN, SGD, SVM, LSTM, XBOOST, DBSCAN</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[93]</td>
<td>LSTM, LSTM, LSTM + CNN, BiLSTM, LSTM-CNN, CNN-RNN, LSTM-CNN</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[100]</td>
<td>CNN</td>
<td>Accuracy</td>
</tr>
<tr>
<td>[78]</td>
<td>GA-LSTM</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[79]</td>
<td>SVM, NB, MLP, DQN, AE-RL, AE-Dueling, DQN RIoT</td>
<td>Precision, Recall, F1-score, Accuracy, G-Mean</td>
</tr>
<tr>
<td>[105]</td>
<td>SCNN-Bi-LSTM</td>
<td>Precision, Recall, F1-score, Accuracy</td>
</tr>
<tr>
<td>[109]</td>
<td>BERT, CNN, LSTM</td>
<td>F1-score, Accuracy</td>
</tr>
<tr>
<td>[80]</td>
<td>DT, RF, LR</td>
<td>Precision, Recall, F1-score, Accuracy, ROC AUC</td>
</tr>
</tbody>
</table>

N/M: not mentioned.
On the one hand, researchers can apply these solutions to circumvent the complexities of modern cybersecurity, and on the other hand, they can develop efficient measures to counter emerging cyber threats.

### 5.4.2 ML/DL techniques

ML and DL techniques provide many benefits but also pose considerable difficulties. Fundamentally, it has been proven that they can adequately perform image classification, object recognition, and natural language data processing [116], the latter of which is authenticated by progress in data representation [117]. In ML [118], transferring data from models trained with training sets to other models is an interesting topic. However, one of the primary challenges is obtaining the large amounts of qualitative data needed for training these ML algorithms [119,120]. It takes time and money to collect, label, and annotate the data [121]. Moreover, the ethical implications, possible biases, and consequences of AI-driven content call for a thorough assessment, especially in the fields of NLP, computer vision, and image analysis [122–124].

However, during ML thinking, there are many security threats that ML and DL face [8]. As illustrated, attackers have a massive incentive to alter the results of ML and DL model outputs or acquire confidential information for their benefit [25]. The analysis was carried out for networks and cybersecurity in adversarial attacks, and based on the algorithms given in Table 2, we considered this unique feature to define gaps in the use of ML and DL techniques in the literature. Consequently, one of our contributions is identifying algorithms that have not hitherto been researched as such subjects for separate studies, constituting unique classifications and inquiries.

This study incorporated 62 different methods used in all the contributions. By comparing the presented taxonomy with how frequently each of these methods was employed in papers, it is revealed that there are plenty of avenues for future study. The remarkable techniques illustrated by the green accent in Figure 30 are explained below.

![Figure 30: The methods of ML and DL used for the period from 2020 to 2023 vs the methods used in the first quarter of 2024.](image)

Figure 31 shows an overview of the interlinkages between the methods applied by researchers in the period of 2020–2023 and shows the methods employed in the quarter of 2024 (1st Q 2024). Figures 31 and 32 expand on this idea, showing a list of methods previously used in research papers and methods that are not yet used (for the future). These graphical illustrations serve more as the result of an ongoing investigation within this given sector, with an emphasis on where further research may lead.
By comparing the ML and DL methods employed in 24 papers (2020–2023) to the methods used in 42 papers (first quarter of 2024), it is clear that there have been significant shifts and advancements in cybersecurity research methodology. At the beginning of ML evolution, researchers mostly worked with classical algorithms such as decision trees, random forests, support vector machines (SVMs), and k-nearest neighbours (KNNs) to design IDSs, anomaly detection models, and malware detectors, but now, these algorithms have been revolutionized by the latest DL approaches. While traditional methods, such as supervised learning and simple...
approaches, have been in use for a long time, recent developments in the realm of DL techniques, especially DNNs, CNNs, and recurrent neural networks (RNNs), have led to the emergence of more advanced and data-driven solutions. In 2024, DL is booming, while various cutting-edge approaches, including adversarial auto-encoding, GANs, and federated learning, are increasingly adopted in cybersecurity research to move the direction toward resilient security systems. In addition, hybrids dominate, which use the possibility of both

Figure 32: Methods of ML used vs unused in the first quarter of 2024.

Network and cybersecurity applications of defense in adversarial attacks
logic and DL as the most important techniques. This change is symbolic of the overall event of capitalizing on the potential of DL to solve sophisticated cybersecurity problems through techniques such as malware detection, adversarial signals, and network nastiness detection. DL methodologies will help researchers manoeuvre dynamically and develop better approaches to fight cyber threats, which will continue to evolve in the future.

5.4.3 New insights into cybersecurity

The future applications of defense networks and cybersecurity concerning adversarial attacks span a wide range. These include benchmarking DL systems, malware detectors, and other models to study their operational variability in adversarial settings and devising new classifiers for higher precision by increasing the rationality of the decision conditional on representational evidence while generating threat cases from several log sources based on ML accuracy. Moreover, researchers also seek to widen the datasets for diverse purposes and investigate complicated detection algorithms as well as utilize several formulations of GANs that are designed to be used only during training. The scope also includes supplying cost-efficient IoT security systems, combining semantic tree neural networks and survey approaches over various protocols as well as event logs. In addition, goals are set to address overfitting in DNNs and software acceleration for non-independent and identically distributed dataset optimization of learning rate tests against larger datasets. In pursuit of realism, researchers have aimed to consider black-box attack methods and study transferability within adversarial ML and DL. The use of RL in GAN-based attack generation and walking on federated learning-based IDSs contributes to expanding the research scope. In general, the focus is on handling a broad spectrum of attacks and compromising false alarms under good restoration conditions in different adversary settings.

6 Conclusion

In this systematic study, comprehensive network and cybersecurity analysis related to attacks from opponents was conducted on adversarial attacks. The study clarified different security mechanisms, vulnerability types, and tactics to fight the ever-changing cyber threat landscape of an organization. Examination, however, showed that special attention to temporal parameters was the main feature of the analysis. The defense of network and cybersecurity systems, which is integrally related to digital sphere safety against threats, has come to light. Adversaries are prone to continuously tweaking their tactics; therefore, the need for something responsive and adaptable becomes increasingly critical.

The principle element of our analysis emphasizes the interactive dynamics between attackers and defenders, where it is a never-ending battle similar to that of cats and mice. To successfully counter the constantly increasing dangers, cybersecurity regulations should evolve at the same pace as adversary innovations and methodical enhancements. Deciphering this intricate link is crucial for shaping efficient countermeasures that will be strong enough to impede adversary penetrations.

This review emphasizes the vital role that DL and ML techniques play in cybersecurity. Organizations can directly identify and manage hazards before they occur by applying the predictive abilities of ML algorithms. However, implementing ML-based security mechanisms comes with its own set of challenges, such as adversarial attacks, through which the elements of fraud in ML models are identified.

Overall, this complete analysis presents a unique tool for understanding how networks and cybersecurity operate during hostile attacks. We strive to strengthen security skills and lessen the overall negative impact of adversary operatives in the digital domain. The state of the art and related issues are revealed, as well as the parameters of innovation are indicated. The imperative of fostering cooperation and proactiveness in response to the dynamically evolving cybersecurity landscape is emphasized in this study. In this regard, there are general and specific limitations to the study at hand. From the specific dimension, the information fails to describe the investigation of the parameter types and nature in the development of the adversarial ML/DL techniques. In contrast, the scope of the analysis is specifically narrowed down to the applications of defense in network and
cybersecurity against adversarial attacks. Under this context, the results cannot be generalized to other areas or uses.

**Funding information:** The author states no funding involved.

**Author contributions:** Yahya Layth Khaleel: data curation, writing – original draft preparation, and supervision. Mustafa Abdul fattah Habeel: visualization, investigation, and supervision. A.S. Albahri: conceptualization, methodology, supervision, and editing. Tahsien Al-Quraishi: writing – reviewing and editing. O. S. Albahri: reviewing and editing. A. H. Alamoodi: writing – reviewing and editing.

**Conflict of interest:** The authors declare no conflict of interest.

**Data availability statement:** Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

**References**


