Information Fusion 97 (2023) 101804

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Artificial intelligence for cybersecurity: Literature review and future research directions

Ramanpreet Kaur *, Dušan Gabrijelčič, Tomaž Klobučar

Laboratory for Open Systems and Networks, Jožef Stefan Institute, Ljubljana, Slovenia

ARTICLE INFO

Keywords:
Detection
Protection
Response
Recovery
Identify
Learning
Cyberattacks
Taxonomy

ABSTRACT

Artificial intelligence (AI) is a powerful technology that helps cybersecurity teams automate repetitive tasks, accelerate threat detection and response, and improve the accuracy of their actions to strengthen the security posture against various security issues and cyberattacks. This article presents a systematic literature review and a detailed analysis of AI use cases for cybersecurity provisioning. The review resulted in 2395 studies, of which 236 were identified as primary. This article classifies the identified AI use cases based on a NIST cybersecurity framework using a thematic analysis approach. This classification framework will provide readers with a comprehensive overview of the potential of AI to improve cybersecurity in different contexts. The review also identifies future research opportunities in emerging cybersecurity application areas, advanced AI methods, data representation, and the development of new infrastructures for the successful adoption of AI-based cybersecurity in today's era of digital transformation and polycrisis.

1. Introduction

The term cybersecurity refers to a set of technologies, processes and practices to protect and defend networks, devices, software and data from attack, damage or unauthorized access [1]. Cybersecurity is becoming complex because of the exponential growth of interconnected devices, systems and networks. This is exacerbated by advances in the digital economy and infrastructure, leading to a significant growth of cyberattacks with serious consequences. In addition, researchers report the continued evolution of nation-state-affiliated and criminal adversaries, as well as the increasing sophistication of cyberattacks, which are finding new and invasive ways to target even the savviest of targets [2]. This evolution is driving an increase in the number, scale and impact of cyberattacks, and necessitating the implementation of intelligence-driven cybersecurity to provide a dynamic defence against evolving cyberattacks and to manage big data. Advisory organizations, such as the National Institute of Standards and Technologies (NIST), are also encouraging the use of more proactive and adaptive approaches by shifting towards real-time assessments, continuous monitoring and data-driven analysis to identify, protect against, detect, respond to, and catalogue cyberattacks to prevent future security incidents [3].

AI is an intriguing tool that can provide analytics and intelligence to protect against ever-evolving cyberattacks by swiftly analysing millions of events and tracking a wide variety of cyber threats to anticipate and act in advance of the problem. For this reason, AI is increasingly being integrated into the cybersecurity fabric and used in a variety of use cases to automate security tasks or support the work of human security teams. The flourishing field of cybersecurity and the growing enthusiasm of researchers from both AI and cybersecurity have resulted in numerous studies to solve problems related to the identification, protection, detection, response and recovery from cyberattacks.

Several reviews on cybersecurity and AI applications were published in recent years [4–7]. However, to the best of our knowledge, there is no comprehensive review that covers state-of-the-art research to explain cybersecurity activities covered by AI techniques and the details of how they are applied. Therefore, our objective was to provide a systematic review, a comprehensive view of AI use cases in cybersecurity, and a discussion of the research challenges related to the adaptation and use of AI for cybersecurity to serve as a reference for future researchers and practitioners. Table 1 shows a comparison of the study with review articles from recent years.

We performed a systematic literature review (SLR) on the use of AI for the provision of cybersecurity, with a particular focus on practical applications within five different cybersecurity functions (Identify, Protect, Detect, Respond and Recover) defined by the NIST cybersecurity framework [3]. The specific research questions addressed by this

* Corresponding author.
E-mail address: raman@e5.ijs.si (R. Kaur).

https://doi.org/10.1016/j.inffus.2023.101804
Received 19 January 2023; Received in revised form 16 March 2023; Accepted 6 April 2023
Available online 7 April 2023

SLR are:

- RQ1: What would be the taxonomical representation of the application of AI for the provision of cybersecurity?
- RQ2: What are the specific use cases of AI for cybersecurity?
- RQ3: What are the current research trends associated with AI for cybersecurity?
- RQ4: What are the trending topics and future research directions for the adoption of AI for cybersecurity?

To answer these research questions and to provide a valuable output for the research community, 236 articles were examined prior to February 2022. Then, the selected studies were further analysed to specify the cybersecurity applications where AI was used, the selected AI domain, and the resulting impact. The SLR led to the following:

- A taxonomy of AI for cybersecurity that provides the multi-level classification of the reviewed articles based on the cybersecurity functions, solution categories, and specific use cases.
- Specific use cases of AI for cybersecurity to reveal the potential areas to harness the capabilities of AI.
- A descriptive analysis of the literature to explore the research trends of AI for cybersecurity.
- A critical analysis of the existing literature, identifying research gaps, to stimulate future research in the field.

The rest of the article is structured as follows. Section 2 discusses the relevant background to provide an introduction and conceptualization of cybersecurity and AI topics along with an explanation of the classification paradigms used in the literature related to AI for cybersecurity. Section 3 describes the research methodology adopted to conduct the SLR. Section 4 discusses the data extraction process to feed the descriptive analysis and state-of-the-art research presented in Section 5. Section 6 provides a descriptive analysis of the synthesized literature review. Section 7 identifies various research gaps that new studies can target, while Section 8 points out the limitations of our study. Finally, Section 9 presents the main conclusions and the research implications of this SLR.

2. Background

This section is dedicated to analysing the background information concerning the key concepts of this review, including the operational definition of cybersecurity using the NIST cybersecurity framework [3] and the AI taxonomy proposed by AI Watch [8] to clarify the concept of different applications of AI for cybersecurity.

2.1. Cybersecurity

Cybersecurity puts policies, procedures and technical mechanisms in place to protect, detect, correct and defend against damage, unauthorized use or modification, or exploitation of information and communication systems and the information they contain. The rapid pace of technological change and innovation, along with the rapidly evolving nature of cyber threats, further complicates the situation. In response to this unprecedented challenge, AI-based cybersecurity tools have emerged to help security teams efficiently mitigate risks and improve security. Given the heterogeneity of AI and cybersecurity, a uniformly accepted and consolidated taxonomy is needed to examine the literature on applying AI for cybersecurity. This structured taxonomy will help researchers and practitioners come to a common understanding of the technical procedures and services that need to be improved using AI for the implementation of effective cybersecurity.

For this purpose, a well-known cybersecurity framework proposed by NIST was used to understand the solution categories needed to protect, detect, react and defend against cyberattacks [3]. The NIST cybersecurity framework's core describes the practices to improve the cybersecurity of any organization. The framework's core has four elements: Functions, Categories, Subcategories and Informative references. The first two levels of the NIST framework, which consist of 5 cybersecurity functions and 23 solution categories, were used to classify the identified AI use cases. The functions provide a comprehensive view of the lifecycle for managing cybersecurity over time. The solution categories listed under each function offer a good starting point to identify the AI use cases to improve the cybersecurity. The main purpose of selecting these two levels is to provide a clear and intuitive categorization to classify the existing AI for cybersecurity literature into the appropriate solution category. The proposed taxonomy introduces a third level consistent with the first two levels by specifying AI-based use cases corresponding to each level of the cybersecurity framework, as shown in Fig. 1. A detailed description of the proposed taxonomy with a state-of-the-art review of AI for cybersecurity is provided in Section 5.

This taxonomy forms the basis for our SLR, by providing a description of the related subfields to cover the main aspects and fundamental keywords in the definition of cybersecurity solution categories. A detailed description of the keyword selection can be found in Section 3.

2.2. Artificial intelligence

Several definitions of AI systems can be found that relate to (a) the

Table 1
Comparison of this review with existing studies.

References	Taxonomy	Use case identification	Classification of defence solution based on the		SLR	Coverage	Research gaps	Purpose
			Function	AI domain				
Wiafe et al. [4]	No	No	No	Yes	Yes	IEEE & ACM digital library	No	Provides an overview of existing research on AI for cybersecurity
Zhang et al. [5]	No	No	No	No	Yes	Google Scholar, SpringerLink, ScienceDirect, IEEE & ACM digital library	No	Provides a review constrained to the application of AI in the areas of user authentication, dangerous behaviour monitoring, network situation awareness, & identification of abnormal traffic
Torres et al. [6]	No	No	No	No	No	Nil	No	Reviews the application of machine learning techniques in spam, malware, and phishing detection
Truong et al. [7]	No	No	No	No	No	Nil	No	Reviews the application of AI techniques in intrusion, malware, APT and phishing detection
Proposed Study	Yes	Yes	Yes	Yes	Yes	Scopus Database	Yes	Explores research from 2010 to February 2022 related to AI applications for cybersecurity from a descriptive point of view, and a detailed state-of-the-art analysis

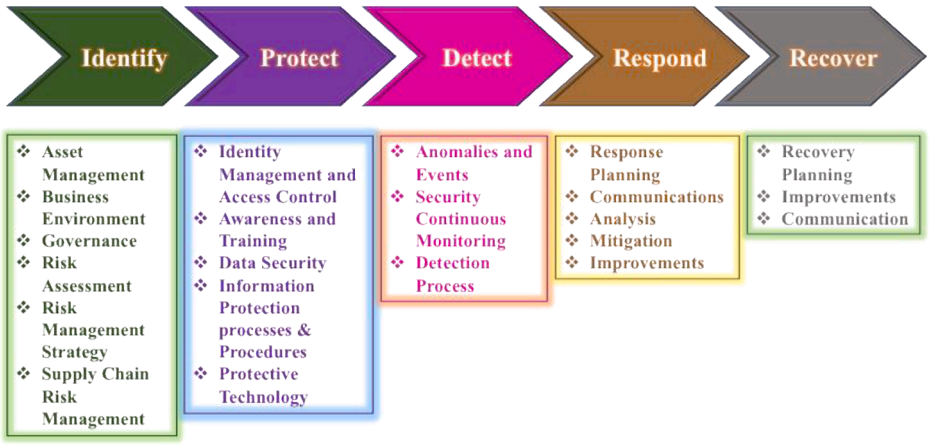


Fig. 1. NIST cybersecurity framework.

fields in which they are used and (b) the stages of an AI system's lifecycle, such as research, design, development, deployment and use. Since the focus of this paper is on AI applications for cybersecurity, a prevailing, but simplified, definition of AI is adopted: "systems that exhibit intelligent behaviour by analysing their environment and with some degree of autonomy take actions to achieve specific goals" [9]. In practical terms, AI refers to a number of different technologies and applications that are used in a variety of ways. AI use cases in cybersecurity describe which environmental situations are desirable and undesirable, and assign actions to sequences.

For this SLR, the AI taxonomy proposed by Samoili et al. [8], which defines the core and transversal AI domains and subdomains, is used. The core AI domains, i.e., reasoning, planning, learning, communications and perception, were found to be useful as they encompass the main scientific areas of AI. Reasoning deals with knowledge representation and different ways of reasoning, while planning also covers searching and optimisation. Learning includes machine learning; communication is related to natural language processing; and perception is about computer vision and audio processing [8]. The approaches and technologies that make up these AI domains include, but are not limited to, fuzzy logic, case-based reasoning, genetic algorithm, Bayesian optimization, evolutionary algorithm, planning graph, artificial neural network, deep learning, support vector machine, natural language processing, text mining, sentiment analysis, image processing, sensor networks, object recognition and speech processing.

AI is a large, multidisciplinary research area, with a large body of literature addressing its applications and consequences from a variety of perspectives, e.g., technical, operational, practical and philosophical. This study focuses on the literature's thread that discusses the implications of the aforementioned methods and AI applications in cybersecurity scenarios. It analyses in detail how AI methods can be used for the identification, protection, detection, response and recovery in the domain of cybersecurity.

3. Research methodology

The SLR aims to identify, evaluate and interpret all the available research in the area of interest to identify potential research gaps and highlight the frontiers of knowledge. It provides a high-quality, transparent and replicable review to summarize the large number of research studies. This study follows an SLR methodology for the following reasons: (i) AI for cybersecurity is a diverse field with a large quantity of literature; (ii) this study aims to answer specific research questions; (iii) the rigour and replicability it provides leads to an unbiased scientific study. The procedure for the SLR is described in detail below.

3.1. Selection of bibliometric database

Scopus and Web of Science (WoS) are the two most popular bibliometric databases. The Scopus database was chosen for this study because its coverage is almost 60% larger than that of the WoS [10]. In addition, Scopus offers better data management due to its wider coverage, advanced search filters and data analysis grids.

3.2. Search strategy

Between November 2021 and February 2022, a comprehensive search for terms related to AI and cybersecurity was conducted for the purpose of a thorough literature review of the impact of AI on cybersecurity. The search was performed using the well-specified search terms for the AI and cybersecurity fields, as shown in Table 2. The keywords of the AI and cybersecurity fields were combined using the logical AND operator. The logical OR operator within the different keywords was used to find studies that are related to any of the terms in each field. Specifically, the AI keywords correspond to the AI taxonomy proposed by AI Watch [8], and the cybersecurity keywords were taken from the NIST cybersecurity framework [3].

3.3. Inclusion and exclusion criteria

Following the search stage, the studies identified were screened to eliminate irrelevant work. To find the pertinent papers that address the research questions, the studies gathered in the earlier stage were subject to inclusion and exclusion criteria. A significant, yet manageable, selection of studies must be ensured at this point. The search conducted was not limited to a specific period and also considered early publications to avoid overlooking any important studies. The inclusion criteria were as follows:

- The article is written in English.
- The article is a full research paper (i.e., not a presentation or supplement to a poster).
- The article should make it apparent that AI is its primary emphasis or include AI as a large part of the methodology. For example, publications that explicitly include machine learning as a core component of their methodology/research.
- One or more of the research questions posed in this research are directly answered by the article.
- For studies that have appeared in multiple journals or conferences, the most recent version is considered.

The following publications were excluded from further review:

Table 2
Search string.

AI Keywords	Cybersecurity Keywords
(("reasoning" OR "optimization" OR "machine learning" OR "artificial intelligence" OR "Natural language processing" OR "text mining" OR "classification" OR "feature extraction" OR "data mining" OR "sentiment analysis" OR "computer vision" OR "recognition" OR "genetic" OR "filtering" OR "GAN" OR "deep learning" OR "reinforcement learning" OR "data driven" OR "topic modelling")	("cybersecurity") AND (("asset management" OR "inventory" OR "configuration" OR "security control validation" OR "assessment" OR "asset" OR "security control testing" OR "security posture" OR "business impact" OR "governance" OR "risk management" OR "team" OR "risk indicators" OR "risk assessment" OR "automated vulnerability" OR "vulnerability" OR "fuzzing" OR "penetration" OR "vulnerability severity" OR "vulnerability management" OR "threat hunting" OR "automated penetration" OR "attack graph" OR ("risk" AND "investment") OR "risk quantification" OR ("risk" AND "supply chain") OR "role mining" OR "role maintenance" OR "Multi-Factor authentication" OR "authentication" OR "identity" OR ("contextual" AND "authentication") OR "access control" OR "unauthorized access" OR "VPN" OR ("attribute based access") OR "Role based access" OR "segregation" OR "isolation" OR "isolate" OR "network segmentation" OR "data loss" OR "data leakage" OR "SQL injection attack" OR "APT" OR "email" OR "malicious domain" OR ("integrity" AND "files") OR ("integrity" AND ("monitoring" OR "auditing")) OR ("automated" AND "configuration") OR "fake news" OR "backup" OR (("backup") AND ("data" OR "code")) OR "plan" OR (("business continuity" OR "disaster Recovery" OR "incident response") AND ("automated")) OR "risk scoring" OR "risk prioritization" OR "vulnerability exploitation" OR (("Risk") AND ("remediation")) OR (("log" OR "audit") AND ("analysis")) OR "SIEM" OR "VPN" OR "firewall" OR "IPS" OR "antivirus" OR "antimalware" OR "immune system" OR ((("anomaly") OR ("event") OR ("intrusion") OR ("fraud")) AND ("detection") OR (("event correlation") OR ("security intelligence") OR ("event analysis") OR ("correlation") OR ("SIEM") OR ("SOC") OR ("monitoring") OR ("behavior Based") OR ("social network") OR ("threat intelligence")) OR ("dark web") OR ("chatter noise") OR ("translate") OR ("topic modelling") OR ("sentiment analysis") OR ("cyber trap") OR ("threat intelligence") OR ("darknet") OR ("deepnet") OR ("social network") OR ("sentiment") OR ("honeypot")) OR (("incident") AND (("detection") OR ("response") OR ("playbook") OR ("case based") OR ("case") OR ("identification") OR ("assessment") OR ("classification") OR ("categorisation") OR ("categorising"))) OR (("cybersecurity") AND ("triage")) OR (("forensic") AND (("intelligent") OR ("incident")) OR "isolation" OR "remediation" OR "risk quantification" OR "recommender system" OR (("incident") AND (("analysis") OR ("report") OR ("document") OR ("information"))) OR (("recovery") AND (("planning") OR ("dynamic")) OR ("safe")) OR "recovery")

- The studies that are not written in English;
- The studies that provide a review or survey of AI in different cybersecurity domains;
- The articles that represent the same work by authors in different conferences or journals were also filtered to remove duplicates;
- The articles that provide a comparative analysis of different AI models or existing techniques for cybersecurity tasks;
- The articles that improve the security of AI techniques to make them attack resistant;
- The papers providing only recommendations, guidelines or principles for cybersecurity (non-scientific);
- Editorials, books, chapters and summaries of workshops and symposiums;
- The studies that do not provide sufficient information;
- The studies that have fewer than 5 pages;
- The studies where a full text could not be found.

3.4. Selection of primary studies

Fig. 2 shows in detail the selection process for the study. After the initial step of identifying and applying a search term, the inclusion and exclusion criteria were applied to refine the 2395 studies retrieved from the Scopus database. Based on the removal of non-English papers, posters, reviews, surveys, non-scientific publications, editorials, books, chapters, summaries of workshops and symposia, duplicates, guideline documents, and comparative studies, 366 articles were removed, leaving 2029. These 2029 studies were analysed based on the title and abstract. The title and abstract provided a clear indication of whether the study was outside the focus of the review and could therefore be excluded. If the title or abstract did not clearly indicate the application domain or contribution of the study, it was included in the review for subsequent steps where full text of the article was examined. Based on the title and abstract analysis, the 2029 studies were further narrowed down to 638. After a thorough examination of the full articles, 402 additional studies were eliminated. As a result, a total of 236 primary studies served as the basis for this SLR. The next sections present the findings and analysis of these 236 primary studies.

4. Data extraction

After the selection of the primary studies, data extraction began to feed the state-of-the-art and descriptive analysis phase. The main goal of data extraction is to break down each study into its constituent parts and describe the overall relationships and connections. The data extraction parameters (explained in Table 3) collect the qualitative and contextual data from the primary studies selected for the SLR. The qualitative data are collected to write a short summary of each primary study to present the contribution along with the demographic information. The contextual data include details about the cybersecurity function, solution category, use cases, and core AI domain, to have a clear understanding of the existing literature. These qualitative and contextual data are further examined to identify the relationships between the different studies.

5. State of the art

To identify the studies that evaluate the application of AI for cybersecurity, a taxonomy is proposed to classify the studies that address the first two research questions (RQ1 and RQ2). The first two levels of the taxonomy are adopted from the NIST cybersecurity framework. The first level organizes the cybersecurity literature into five core functions: identify, protect, detect, respond and recover. These five cybersecurity functions cover the use of AI tasks from the prevention of the security attack to the more complex mechanism of actively looking for new threats and counterattack. The functions cope with different aspects of the cybersecurity attack lifecycle for an effective defence. The second level of taxonomy uses the NIST framework's categories to

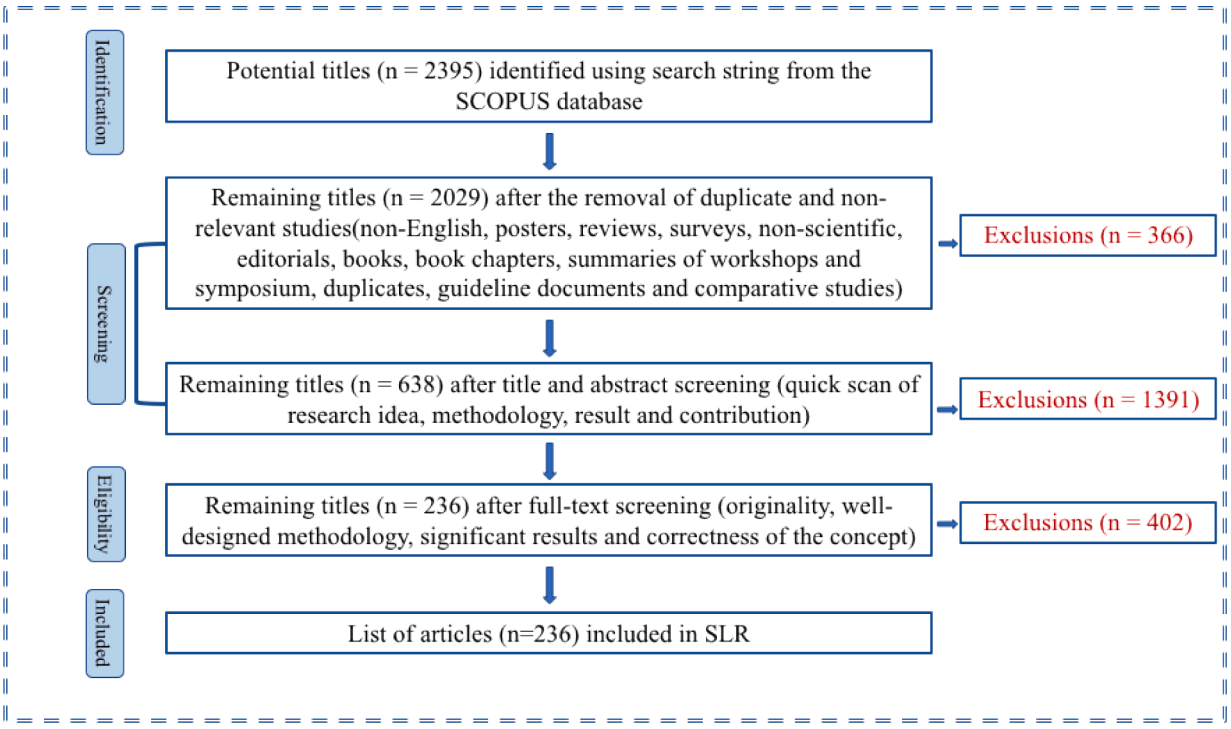


Fig. 2. Selection process and study count at each stage of the SLR protocol.

Table 3
Data extracted from each primary study.

Data Type	Data Item	Description
Qualitative Data	Title	Title of the primary study
	Author	Author of the study
	Year Published	Publication year of study
	Article Type	Publication type, i.e., conference, journal
	Source	Journal/Conference name that published the study
	Geographical Region	Geographic region of the authors of the primary study
	Summary	A summary of the paper, with major contribution.
Contextual Data	Cybersecurity Function	Type of cybersecurity activity in primary study. NIST taxonomy defines cybersecurity activities as 5 functions: Identify, Protect, Detect, Respond, Recover.
	Solution Category	Identification of the main solution category in which primary study falls. The NIST taxonomy provides a subdivision of each cybersecurity function into groups of cybersecurity solution categories, e.g., the detection function is divided into 3 categories: anomalies and events, security continuous monitoring and detection processes.
	Specific Use Case	Specific cybersecurity use case of primary study for AI application to match the function and solution category.
	Core AI Domain	Core AI domain of the AI technique used by the primary study as defined by the AI Watch [7].

expand the core functions into different cybersecurity solutions with closely tied programmatic needs and particular activities. The last level of the taxonomy presents the AI use cases associated with the upper level of taxonomy and link the SLR with each identified use case. Fig. 3 summarizes the proposed taxonomy and presents the logical progression of cybersecurity functions along with a detailed description of the different categories of cybersecurity solutions implemented using the AI technologies.

5.1. *Identify*

The **identify** function provides the foundation for the other cybersecurity functions by pinpointing the critical functions and risks associated with systems, people, assets and data. This helps develop an understanding of the current state of the cybersecurity, identify gaps, and develop an appropriate risk management strategy to achieve the desired security based on the organization's own needs, risks and budget. Table 4 summarizes the main contribution of each primary study in the identify function. The various categories of cybersecurity solutions in this function are detailed below.

5.1.1. *Asset management*

Asset management is the process of identifying and keeping track of the information, people, equipment, systems and buildings that help an organization accomplish its goals and are proportionate to the asset's relative importance to those goals and risk strategies. It includes the discovering, inventorying, managing and tracking of assets to protect them. Cybersecurity asset management is becoming increasingly complex as organizations have more platforms than ever before: from operational technology systems and Internet of Things (IoT) to on-premises and cloud-based services. This proliferation of new asset types and the ability to work remotely have resulted in highly distributed assets that are difficult to manage and inventory.

An AI-based asset management system can solve many of these challenges by feeding new levels of intelligence to the human team across the following use cases.

5.1.1.1. Asset inventory management. Asset discovery and management are critical to ensure complete visibility and control over all assets in an extended network. AI can help with the continuous and automatic discovery of all devices, applications and users, as well as their classification and criticality for operations. With an accurate and up-to-date inventory, assets can be tracked and analysed for a risk assessment

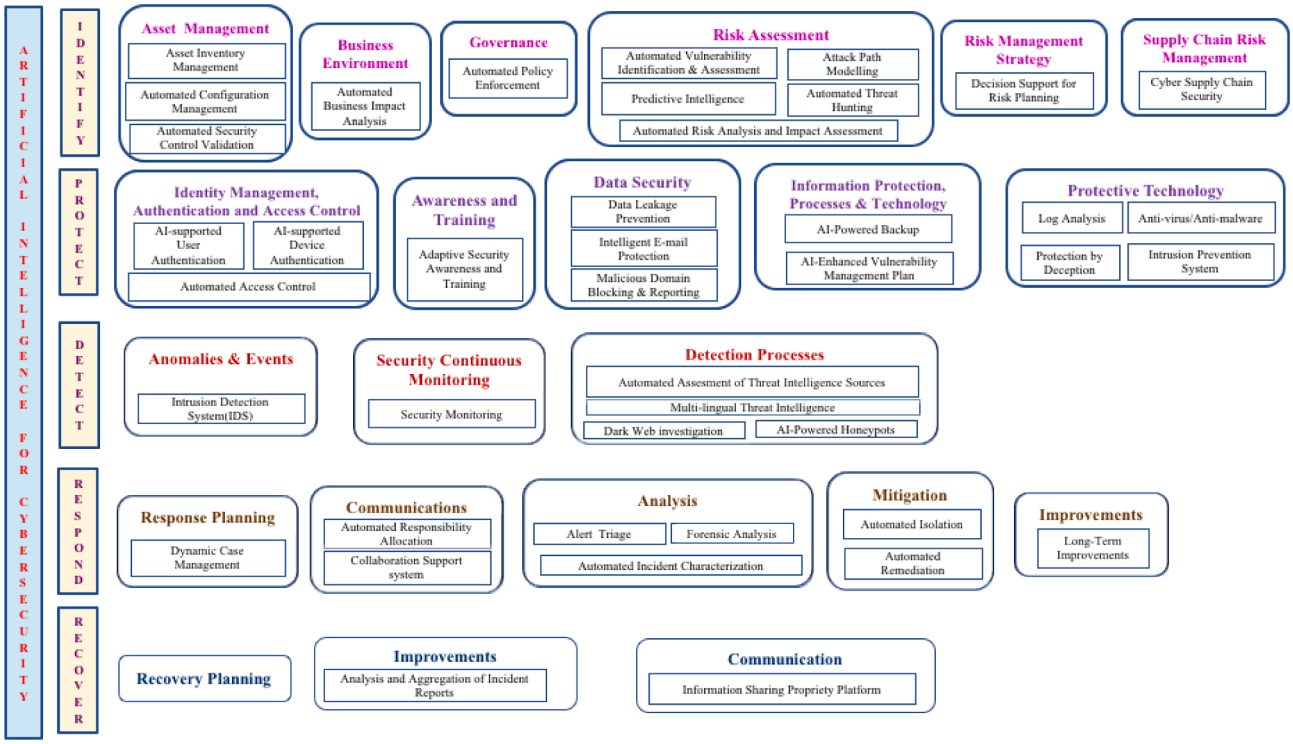


Fig. 3. Proposed taxonomy of AI techniques in the cybersecurity domain.

against known attack vectors, and compliance monitors can detect rogue assets and unauthorized use.

Researchers have developed different approaches to asset classification using machine learning algorithms. Promyslov et al. [11] used a k-means clustering to classify the assets according to their cybersecurity requirements based on their safety, functionality and integrity in a nuclear power plant. Millar et al. [12] proposed a random forest-based machine learning classifier for operating system classification and identification of the vulnerable devices on the network. Several studies [13–15] focus on identifying and classifying IoT devices based on their network-traffic characteristics. Aksoy et al. [13] and Sivanathan et al. [14] worked for single-device identification and classification using multiple and multi-stage machine learning methods, respectively, and are only suitable for a small IoT network. Cvitić et al. [15] proposed a solution to the classification problem in a rapidly evolving, heterogeneous and dynamic environment by using a supervised machine learning method capable of assigning IoT devices to predefined classes based on the values of their traffic flows.

Researchers are also working on the identification and blocking of malware infected assets [16], the determination of asset criticality [17], and the risk assessment of individual assets [18] to manage and ensure their security.

5.1.1.2. Automated configuration management. Configuration management is a governance process for defining and maintaining the desired state of a system and providing timely alerts for any misconfiguration. The automated configuration management system will consistently define the system settings and maintain the system accordingly by only allowing changes in a controlled and authorized environment.

The customization of the system's configuration to ensure the required level of performance and security is important to reduce human error due to manual or sub-optimal configuration settings. Researchers [19,20] are working on dynamic configuration systems for online file sharing systems and distributed cloud storage based on system characteristics and operating environments using multi-objective reinforcement learning and genetic algorithms, respectively. Sharifi et al. [21] and Bringhenti et al. [22] proposed a fully automated framework for the customization of security controls by observing the user's behaviour and by refining high-level security requirements expressed in a human-friendly language, respectively.

Automated configuration assessment provides the compliance team with the ability to continuously review and test the configuration to identify the vulnerable configuration in time to reduce or avoid cybersecurity incidents. Varela-vaca [23,24] proposed a method based on software product line techniques to automatically analyse the vulnerable configurations of the system. Liu et al. [25], on the other hand, demonstrated the use of a random forest model to predict cybersecurity incidents based on the misconfiguration of the DNS and BGP protocols along with the externally visible malicious activities originating from the network.

5.1.1.3. Automated security control validation. The automation of security control validation will provide the real-time monitoring of security in a changing environment and threat landscape. Researchers are working on the implementation of AI techniques for a definitive assessment of the overall security of the system using a network's telescope data [26], a building's cybersecurity framework [27], or by correlating the threat, vulnerabilities and security measures [28].

5.1.2. Business environment

The business environment category is defined to identify the critical processes and applications that ensure business continuity in adverse scenarios. This information is critical to business sustainability and serves as the basis for developing effective response and recovery strategies. AI technology can be used to automate this process via the following use case.

5.1.2.1. Automated business impact analysis. Business impact analysis is the most important technique to determine critical functions and applications in the business environment by evaluating the impact of cybersecurity incidents on the business. AI techniques can be used to automate the business impact analysis by evaluating the economic risks

Table 4
Summary of the primary studies focused on the identify function.

Solution Category	Use Case	Contribution	AI domain	Author
Asset management	Asset inventory management	Asset classification	Learning	Promyslov et al.[11]., Millar et al. [12]
		Asset classification	Planning & Learning	Aksoy et al. [13]
		Asset classification	Communication & Learning	Sivanathan et al. [14]
		Asset identification & monitoring	Learning	Cvitič et al. [15]
		Identification of malware infected assets	Learning	Cam [16]
		Asset criticality & risk prediction	Reasoning & Learning	Kure et al. [17]
		Threat & risk assessment	Reasoning	Vega-Barbas et al. [18]
	Automated configuration management	Dynamic configuration system	Learning	Tozer et al. [19]
		Dynamic configuration system	Planning	García-Hernández et al. [20]
		Customization of configuration	Reasoning	Sharifi et al. [21]
		Customized optimal allocation and network security configuration	Planning	Bringhenti et al. [22]
		Automated configuration assessment	Reasoning	Varela-Vaca et al. [23,24]
		Automated configuration assessment	Learning	Liu et al. [25]
	Automated security control validation	Characterization of cybersecurity posture	Learning	Zhan et al. [26]
		Self-assessment to determine cybersecurity posture	Reasoning	Gourisetti et al. [27]
		Automation of security analyst work	Planning	Stepanov et al. [28]
Business environment	Business impact analysis	Model and measure the economic risk	Learning	Narasimhan [29]
		Estimation of the probability of security event	Planning	Nguyen & Nicol [30]
		Feasibility of attack and impact assessment on business asset	Reasoning	Ponsard et al. [31]
Governance	Automated policy enforcement	Automated policy enforcement on network traffic	Planning	Odegbile et al. [32]
Risk assessment	Automated vulnerability identification and assessment	Vulnerability detection	Communication & Learning	Nembhard et al. [33]
		Vulnerability detection	Communication	Liu et al. [34], Jeon & Kim [35]
		Vulnerability tracking	Communication & Learning	Huff et al. [36]
		Vulnerability tracking	Communication	Iorga et al. [37]
		Vulnerability identification	Learning	Saha et al. [38]
		AI-based fuzzing	Reasoning	Wang et al. [39]
		AI-based fuzzing	Learning	Wang et al. [40], Godefroid et al. [41], Cummins et al.[42]., Xu et al. [43], She et al. [45], Liu et al. [46]
		AI-based fuzzing	Planning & Learning	Chen et al. [44]
		Automated penetration testing	Learning	Zhou et al. [47], Gangupantulu et al. [48], Neal et al. [49]
		Vulnerability classification	Communication	Russo et al. [50], Aota et al. [51], Vanamala et al. [52]
		Vulnerability exploration	Reasoning	Bakirtzis et al. [53]
		Vulnerability exploration	Communication	Kuppa et al. [54]
		Vulnerability exploration	Learning	Chatterjee & Thekdi [55]
		Vulnerability assessment & remediation	Communication	Jiang & Atif [56], Samtani et al. [57]
		Vulnerability assessment & remediation	Planning	Brown et al. [58]
	Automated threat hunting	Open-source cyber threat intelligence (OSCTI)	Communication	Gao et al. [59]
	Attack path modelling	Path modelling using intrusion alerts	Learning	Nadeem et al. [60]
		Path modelling using vulnerability description	Communication & Learning	Binyamini et al. [61]
		Path modelling using vulnerability description	Planning	Falco et al. [62]
		Simulation of attacker/defender activities and preventive measures	Learning	Cam [63]
		Simulation of attacker/defender activities and preventive measures	Planning	Wollaber et al. [64]
	Automated risk analysis and impact assessment	Automated calculation of risk score	Learning	Sancho et al. [65]
		Automated calculation of risk score	Reasoning	Tubis et al. [66]
		Inference of probability of security incident	Learning	Qin et al. [67]
		Identification of key vulnerability risk indicators	Learning	Falco et al. [68]
		Identification of key vulnerability risk indicators	Reasoning	Vega-Barbas et al. [69]
		Automated risk evaluation & decision analysis	Learning	Kalinin et al. [70], Al-Hadhrami et al. [72]

(*continued on next page*)

Table 4 (*continued*)

Solution Category	Use Case	Contribution	AI domain	Author
		Automated risk evaluation & decision analysis	Communication	Biswas et al. [71]
	Predictive intelligence	Intrusion alert prediction	Learning	Ansari et al. [73], Wang & Jones [74], Najada et al. [75], Mueller et al. [76]
		Malware prediction	Learning	Rhode et al. [77]
		Attack prediction	Communication	Perera et al. [78]
		Attack prediction	Reasoning	Marin et al. [79]
		Attack prediction	Learning	Polatidis et al. [80]
Risk management strategy	Decision support for risk planning	Decision support system (DSS) for risk planning	Planning	Rees et al. [81], Paul & Wang [82], Paul & Zhang [83]
		Attack graph modelling for risk planning	Planning	Zheng et al. [84]
Supply chain risk management	AI-based supply chain security	Threat analysis & prediction	Learning	Yeboah-Ofori et al. [85]
		Optimal cybersecurity investment	Planning	Sawik [86,87], Sawik & Sawik [88]
		Assessment of cyber resilience	Reasoning	Rahman et al. [89]

based on a known attack vector or by calculating the threat feasibility along with the probability of high-impact security events on critical business areas. Researchers are measuring the economic risk of cybersecurity in different businesses using the modelling of different known attack profiles [29], rare-event simulation [30], or by linking the business objective to the attacker's capabilities to guide a scenario analysis [31] to determine its impact on business assets.

5.1.3. Governance

Governance involves the policies, procedures and processes for understanding the environmental and operational requirements, and monitoring the regulatory requirements of the organization. This helps to identify an organization's responsibilities and provides information about cyber risks to the management. AI can be used for policy enforcement or automating the retrieval of key risk indicators. While there is some research on policy enforcement, no research articles on measuring risk indicators in real time were found in this study. Therefore, it is a temping future research direction to develop an early-warning system to indicate risk development over time due to policy violations, red flags or other symptoms. The automatic retrieval of key risk indicators, such as the mean time between failures, the presence of unpatched systems, risk appetite or the number of attempted breaches, and converting them into knowledge, will be beneficial to prevent a cybersecurity breach by rapidly remediating the risk.

5.1.3.1. Automated policy enforcement. Policy enforcement is crucial for organizations to ensure their compliance with the regulations and appropriate risk management. AI is being used in automated policy enforcement in traditional non-SDN networks by using a controller and policy proxies [32]. The controller is a centralized management server used to manage software defined middleboxes for traditional routers and a policy proxy that will identify the traffic that is subject to polices and assists it in policy enforcement.

5.1.4. Risk assessment

Risk assessment is the process of identifying, estimating and prioritizing cybersecurity risks associated with operations, operational assets and individuals currently or in the near future. It requires a careful analysis of threat, vulnerability and attack information to determine the extent to which cybersecurity events could adversely impact on the organization and the likelihood that such events will occur. The manual risk assessment process is complex, costly and time consuming due to the large number of risk factors, and it requires active human involvement at every stage. The AI-based risk assessment process addresses these challenges by supporting the risk management team in the following use cases.

5.1.4.1. Automated vulnerability identification & assessment. An automated vulnerability assessment is the process of systematically reviewing security weaknesses in a system using automated tools for vulnerability identification, classification, exploration and prioritization These automated tools rely on vulnerability repositories, vendor vulnerability announcements, asset management systems, and threat intelligence feeds to identify, classify and assess the severity, and make recommendations for the remediation.

- **Automated vulnerability detection:** It is an important step to identify the vulnerabilities of applications, servers or other systems and assets of an organization. Researchers have worked on the software vulnerability detection by checking the source code using deep learning and transfer learning [33–35]. These studies employ text-mining techniques to feed the machine learning based vulnerability detection models in unison with a recommendation system to help programmers to write secure code.

  Researchers have also worked on detecting the emerging vulnerabilities of software [36] and cybernetic infrastructure [37] using vulnerability repositories or social networks, respectively. Saha et al. [38] proposed a new scheme for the identification of vulnerabilities across the system and network levels by modelling the behaviour of cyber-physical systems (CPS)/IoT under attack at the system and network levels and then use machine learning to discover any potential attack space.

  Researchers use AI-based fuzzing to discover vulnerabilities in software and hardware interfaces and applications. This is done by injecting erroneous, unexpected or randomly generated data into a program or interface and then monitoring for events like crashes, failed code assertions, undocumented jumps or debug routines, and possible memory leaks. AI techniques are leveraged to develop an automated system for identifying potential attack options, input generation, the generation of probable test cases and analysing crashes, as shown in Fig. 4. Researchers [39,40] used reasoning and natural language processing for seed generation techniques to increase the code coverage with more unique execution paths as one of the basic steps of the smart fuzzing system. Test case generation is one of the most widely studied fields of AI-based fuzzing for web browsers [41], compilers [42,43], cyber-physical systems (CPS) [44], software libraries [45] and simple computer programs [46].
  Automated penetration testing is an efficient and intelligent attempt to penetrate an attack surface by exploiting known or zero-day vulnerabilities to identify what the attacker can gain from current environments. Researchers are working on devising autonomous penetration testing using reinforcement learning for large networks [47,48] and microgrid control algorithms [49].

- **Automated vulnerability classification:** Vulnerability classification is an important step that facilitates a deeper understanding of security related information to accelerate the vulnerability

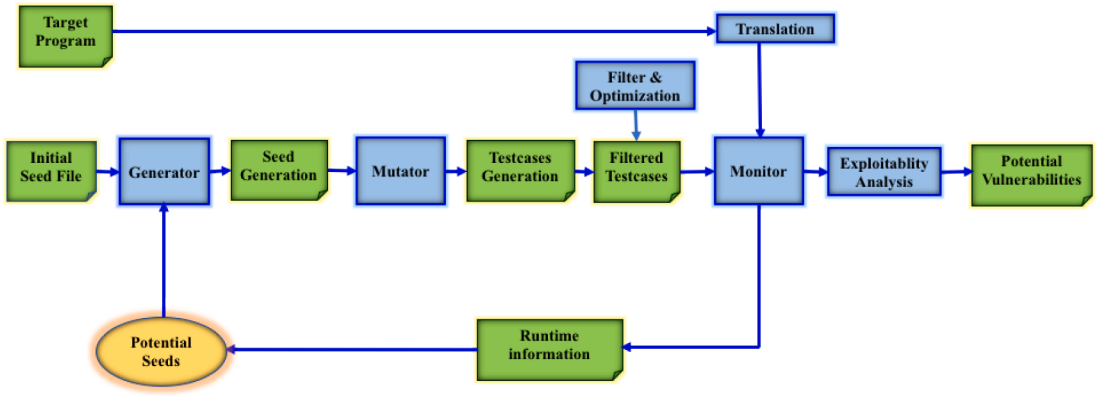


Fig. 4. Smart fuzzing process for vulnerability detection.

assessment. Researchers are working on automatic vulnerability classification systems and for labelling vulnerability descriptions in vulnerability reports. Russo et al. [50] proposed an approach to summarize the daily posted vulnerabilities and categorize them according to the defined taxonomy model for the industry. Aota et al. [51] addressed the use of text mining to classify vulnerabilities using the description provided in the Common Vulnerabilities and Exposures (CVE) list. Vanamala et al. [52] also categorize the CVE entries to the Open Web application Security Project (OWASP) top-10 risks.

- **Vulnerability exploration:** An important step in vulnerability assessment is to identify the potential attack vectors that can exploit the vulnerabilities to effectively evaluate and manage them. For this purpose, some researchers use the dataset from the MITRE corporation [53,54], while others use the CVSS provided by the Forum of incident response and security teams [55]. Bakirtzis et al. [53] and Kuppa et al. [54] used a model-driven process to automatically map adversarial tactics, techniques and common knowledge to the given system model. In contrast, Chatterjee & Thekdi used a probabilistic model to evaluate and manage the system vulnerabilities by rapidly adapting to the changing network and attack characteristics [55].
- **Vulnerability assessment and prioritization:** The main objective of this step is to prioritize vulnerabilities and to provide an assessment report based on its severity and the vulnerability exposure of the systems. AI techniques are used to assign a severity score to each vulnerability based on the system, data and business at risk, along with the ease, severity and potential damage of a resulting attack. Jiang and Atif [56] worked on the automatic assessment and reconciliation of the vulnerability severity from conflicting vulnerability reports using the machine learning pipeline based on the vulnerability severity and threat profile metrics. Samtani et al. assessed the vulnerabilities of the SCADA devices from the Shodan dataset and classified them into four main risk levels: critical, high, medium and low [57]. Brown et al. calculate the vulnerability and exploit risk scores for every IoT device in the attack graph created using the network topology specified by the network administrator [58].

5.1.4.2. Automated threat hunting. Automated threat hunting is a proactive security search across networks, endpoints and datasets to detect potentially malicious, suspicious or risky activities within an organization. It identifies and categorizes prospective threats in advance using fresh threat intelligence on data that has already been gathered. Threat hunting is a relatively new application area that is very important for early detection. However, existing approaches still work on anomaly-based threat detection and overlook the rich external knowledge about threats provided by open-source cyber threat intelligence (OSCTI) [59].

5.1.4.3. Attack path modelling. Attack path modelling is a proactive risk reducing approach that supports security teams by mapping the vulnerable routes in the network to assess risk, identify vulnerabilities, and take counter measures to protect key assets. Researchers have used AI techniques for path modelling using intrusion alerts [60] or vulnerability descriptions [61,62]. Some researchers used all cyber data, including alerts, vulnerabilities, logs and network traffic to simulate attacker/defender activities and take preventive action in real time [63,64].

5.1.4.4. Automated risk analysis and impact assessment. Automated risk analysis and impact assessment strengthen the risk management team by intelligently using the risk data available internally and externally to assess risk and related metrics in real time. AI is a catalyst to accelerate an advancement in risk management by automating the calculation of risk score [65,66], the inference of the probability of a security incident [67], the identification of key vulnerability risk indicators [68,69] and risk assessment and decision analysis [70–72] using log data and threat intelligence within and outside the organization.

5.1.4.5. Predictive intelligence. Predictive intelligence is intelligence that is actionable and relevant in a given context and can be used to anticipate attacks. Intrusion prediction tools are helpful in providing an active defence against future attacks by predicting the type, intensity and target of an intrusion in advance. Researchers are using deep-learning [73–76] approaches to forecast the alerts from malicious sources [73] or on a given target [74–76] using the sequence of previous alerts [73], historic spam e-mail [74], and network traffic [75,76] data.

Malware prediction involves methods to predict and block the malicious files before they execute their payload completely, to prevent malware attacks rather than remedy them. In this direction, Rhode et al. developed a malware prediction model based on a recurrent neural networks (RNNs) model to predict malicious behaviour using machine activity data [77].

Attack prediction is deemed to have excellent potential for proactively advancing cyber resilience. Researchers present attack prediction schemes by utilizing different types of data retrieved from news sites and websites [78], dark web forums [79], national vulnerability databases [79], incident reports [79], and common vulnerabilities and exposure databases [80].

5.1.5. Risk management strategy

A risk management strategy assists the operational risk decisions by establishing the priorities, risk tolerance and constraints. It needs to ensure that the acceptable risk levels are established and documented along with the reasonable resolution times and investment. AI has the potential to transform this category by automating the following activities.

5.1.5.1. Decision support for risk planning. Cybersecurity risk planning involves the implementation of a desired portfolio of countermeasures within a predetermined budget. Formal decision support systems [81–83] and attack graph modelling [84] can help security planners make economic comparisons with the cost of countermeasures and available risk budgets.

The decision-making problem in cybersecurity risk planning is important due to the sensitivity of the risk plan to the decision maker's attitude towards risk and its interplay with the budget available. Thus, the implementation of a decision support system for estimating an uncertain risk faced by an organization under cyberattack, factoring uncertain threat rates, countermeasure costs, and impacts on its assets is very important. In this direction, Rees et al. [81] used a genetic algorithm to find the best combination of countermeasures to block or mitigate security attacks, allowing the user to determine the preferred trade-off between the investment cost and the resulting risk. Paul & Wang [82] and Paul & Zhang [83] have used robust optimisation to investigate the optimal balance between the prevention, detection and containment safeguards to deal with cybersecurity uncertainty.

Zheng et al. [84] used the attack graph model for the identification of a portfolio of security controls to reduce risk. Their model is utilized to choose the best possible set of controls to ensure that the overall cost does not go above the organizational budget.

5.1.6. Supply chain risk management

Supply chain risk management supports risk decisions specifically related to identifying, assessing and managing supply chain risks. The effective management of cybersecurity risks in a supply chain requires a comprehensive view of the threats and vulnerabilities, cost-effective supply chain risk planning strategies, and an assessment of the cyber resilience of the supply chain. Researchers are actively using AI techniques to automate threat analysis and prediction [85], optimal cybersecurity investment [86–88], and an assessment of the cyber resilience [89] of the supply chain.

5.1.6.1. Cyber supply chain security. Cyber supply chain security requires a secure integrated network between the incoming and outgoing chain's subsystems. Therefore, it is essential to understand and predict threats using both internal and threat intelligence resources to limit the disruption of the business. Yeboah-ofori et al. integrated cyber threat intelligence data and used machine learning techniques to predict cyberattack patterns on cyber supply chain systems [85].

Optimizing cybersecurity investments is an important area in the Industry 4.0 supply chain cybersecurity to quickly detect, mitigate and balance the impact of security breaches with the available budget. In this direction, Sawik [86–88] proposed different models to determine an optimal cybersecurity investment with a limited budget and a portfolio of security controls to balance the cybersecurity in the supply chain.

An assessment of the cyber resilience of a supply chain is a crucial task to make the supply chain safe from cyber intrusions and to secure a competitive business advantage. In 2021, Rahman et al. [89] proposed an integrated and comprehensive approach based on Dempster-Shafer (D-S) theory as a methodology for building a framework for evaluating the cyber resilience of an additive manufacturing supply chain.

5.2. Protect

The **protect** function helps plan and implement appropriate controls to limit or contain the impact of a potential cybersecurity event. This includes a number of technical and procedural controls to proactively protect against internal and external cyber threats. AI can improve the resilience of the system by authenticating users, devices and other assets, monitoring the user behaviour, automated access control, adaptive training, data leakage prevention & integrity monitoring, automated information protection and processes and provision of protective solutions to proactively secure the system. A summary of each primary study focusing on the protect function is provided in Table 5. Solution categories along with a detailed overview of the AI use cases in each category are presented below.

5.2.1. Identity management, authentication and access control

Identity management, authentication and access control is responsible for limiting the access to assets and associated facilities to authorized users, processes or devices, and to authorized activities. AI can be used for the management and protection of physical and remote access using intelligent user authentication, intelligent device authentication, automated access control using authorizations, and access permissions to prevent unauthorized access and its consequences.

5.2.1.1. AI-supported user authentication. AI can improve user authentication with physical biometrics [90], behavioural biometrics [91–94], or multi-factor authentication [95,96] instead of easily compromisable usernames, passwords and even one-time text tokens.

Physical biometrics refers to the innate physical characteristics of users that can be used for identification such as fingerprints, iris and biosignals. In 2020, Siam et al. presented a PPG (photoplethysmography) based biometric human authentication system using deep learning [90]. In contrast, behavioural biometrics are related to the uniquely identifiable and measurable patterns in human activities that can provide continuous and user-friendly security. There are several behavioural biometrics, including usage behaviour [91–93] and gait [94]. The use of behaviour patterns related to a user's interaction with its own device is the main basis of the continuous authentication systems. In this direction, Valero et al. suggested using mobile functions and usage data, e.g., accelerometer, gyroscope, magnetometer, and statistics from interactions with different applications, to determine whether the current user is the same as the one previously authenticated [91]. In 2019, Sánchez et al. designed and implemented a continuous authentication mechanism based on the behavioural profiles of users according to their interaction with different office devices for smart offices using a cloud computing paradigm and a random forest (RF) algorithm [92]. In addition, these solutions are gradually being used in federated identity management solutions, which further increases interest in them [93]. Gait authentication is a non-intrusive, transparent and continuous method of authentication in mobile devices that captures the information needed to verify the authenticity of the user as the person walks. In 2022, Alobaidi et al. [94] investigated the feasibility of gait authentication in the uncontrolled real world by using time and frequency-domain feature extraction of each gait cycle.

Multi-factor authentication is a layered approach to secure data and applications that require a combination of two or more credentials to verify the identity or login of users. Researchers [95,96] used the keystroke dynamics as a second layer of security for web-user authentication and device authentication.

5.2.1.2. AI-supported device authentication. Intelligent device authentication is the process of authenticating devices based on their credentials or behaviour in the network to ensure the security of machine-to-machine communication. Researchers are actively working in the field of sensor identification and authentication to ensure the security of cyber-physical systems or the automotive sector. Channel [97] and sensor [98,99] imperfections are used to find the transient and steady-state parameters as an input to the machine learning model for sensor identification.

5.2.1.3. Automated access control. Automated access control restricts system access to authorized users based on situations or their roles and regulations within an organization. Researchers are actively using AI techniques for maintenance of the access control state [100], role mining [101] and situation-aware decision making [102,103] to prevent

Table 5
Summary of the primary studies focused on the protect function.

Solution Category	Use Case	Contribution	AI domain	Author
Identity Management, Authentication and Access Control	AI-supported user authentication	Physical biometric-based authentication	Learning	Siam et al. [90]
		Behavioural biometric-based authentication	Learning	Valero et al. [91], Sanchez et al. [92], Martin et al. [93]
		Behavioural biometric-based authentication	Perception	Alobaidi et al. [94]
		Multifactor authentication	Reasoning	Rahman et al. [95], Shaot & Schmidt [96]
	AI-supported device authentication	Sensor identification & authentication	Learning	Hafeez et al. [97], Baldini et al. [98]
		Source authentication of distributed synchrophasors	Learning	Cui et al. [99]
	Automated access control	Role-based access control	Planning	Benedetti and Mori [100], Abolfathi et al. [101]
		Attribute-based access control	Planning	Chukkapalli et al. [102]
		Attribute-based access control	Reasoning	Leander et al. [103]
Awareness and Training	Adaptive security awareness and training	Adaptive cybersecurity training	Communication	Tan et al. [104]
		Recommender system for secure coding	Communication	Nembhard et al. [105]
		Secure coding awareness	Learning	Gasiba et al. [106]
Data Security	Data leakage prevention	Monitoring data access, data movement and user activity	Learning	Le and Zincir-Heywood [107], Kim et al. [108], Al-Shehari et al. [109]
		Automated data sensitivity detection	Communication	Alzhrani et al. [110], Guo et al. [111]
		Advanced persistent threat detection	Learning	Li et al. [112], Alghamdi & Reger [113]
	Intelligent e-mail protection	Malicious spam email detection	Learning	Gallo et al. [114]
		Malicious spam email detection	Communication	Wu et al.[115].
		Phishing email detection	Communication	Gualberto et al. [116], Nguyen et al. [117]
	Malicious domain blocking & reporting	Website design features for malicious website detection	Learning	Cohen et al. [118]
		Domain-based features for malicious website detection	Learning	Marques et al. [119], Yu et al. [120], Spaulding & Mohaisen [121]
		URL-based features for malicious website detection	Learning	Indrasiri et al. [122], Vinayakumar et al. [123]
		Hybrid features for malicious website detection	Learning	Li et al. [124], Alotaibi [125]
Information Protection Processes & Procedures	AI-powered backup	Dynamic backup scheduling	Reasoning	Qin et al. [126]
		Intelligent backup scheduling	Learning	Van de Ven et al. [127]
	AI-enhanced vulnerability management plan	Context-based vulnerability risk scoring	Reasoning & Learning	Zeng et al. [128]
		Vulnerability exploitation trends	Learning	Yin et al. [129]
		Vulnerability exploitation trends	Communication	Yin et al. [130]
Protective Technology	Log analysis	Evidence extraction	Learning	Bai et al. [131]
		Data presentation technique	Communication	Afzaliseresht et al. [132]
		Handle variety & interoperability issue	Communication	Torre-Abaitua et al. [133], Eljasik-Swoboda and Demuth [134]
		Automated security analysis of heterogenous log data	Learning	Sisiaridis and Markowitch [135]
	IPS	IPS for electronic control units	Learning	De Araujo-Filho et al. [136]
		IPS for IoT network	Learning	Constantinides et al. [137]
	Anti-virus/Anti-malware	Analysis of modus operandi of malware	Learning	De Lima et al. [138]
		Dynamic data analysis	Learning	Marques et al. [139]
	Protection by deception	Decoy text generation	Planning	Karuna et al. [140]

unauthorized access and its consequences.

Role based access control (RBAC) grants access to different users based on their roles in the organization. Benedetti and Mori have proposed the use of AI techniques to update and maintain the access control state when exceptions or violations are reported [100]. They mainly work on providing an optimized plan of actions to reconfigure the RBAC state to facilitate the maintenance process. Abolfathi et al. proposed a scalable and optimal role mining approach to extract user-role and role-permission relations from existing access control lists [101].

Attribute-based access control takes into account various pre-configured attributes of characteristics that can be related to the user, environment and accessed resource. Chukkapalli et al. [102] and Leander et al. [103] tested the performance of situation-aware decision making for attribute-based access control in smart fisheries and smart manufacturing systems, respectively.

5.2.2. *Awareness and training*

This solution category involves the cybersecurity awareness and training provided to personnel and partners to carry out their information security duties and responsibilities in compliance with policies and procedures. AI methods can be used for adaptive and personalized cybersecurity training, awareness or recommendations through the automatic selection of content using natural language processing algorithms [104,105] or by providing a machine learning-enabled intelligent coach for solution-guiding hints [106].

5.2.2.1. Adaptive security awareness & training. Adaptive training and awareness are important to overcome the challenges of outdated training content, the selection of training material, and the selection of an acceptable training approach. Tan et al. [104] created an adaptive web-based learning system that receives up-to-date training content from DBpedia and provides automated content selection based on a learner's prior knowledge of information security. Nembhard et al. [105] and Gasiba et al. [106], on the other hand, help programmers by topic modelling or by using serious games techniques to recommend or raise awareness of secure coding practices.

5.2.3. Data security

Data security governs information management in line with the risk strategy for protecting sensitive information. This includes protecting data at rest and in transit, as well as managing the lifecycle of the assets that hold the information, including its decommissioning or disposal. Researchers are actively using AI techniques for data leakage prevention, intelligent e-mail protection, malicious domain blocking and reporting, and agent-based integrity monitoring to ensure data confidentiality, integrity and availability.

5.2.3.1. Data leakage prevention. Data leakage prevention deals with the detection and protection of data breaches, exfiltration or the unwanted destruction of data. AI techniques are used for monitoring data access, data movement and user activity [107–109], automated data sensitivity detection [110,111], and advanced persistent threat (APT) detection [112–114] to prevent data leakage.

The identification of the authorized individuals and how they use sensitive information provide accurate insights for data leakage prevention by observing their concerning behaviours or activities. Researchers [107–109] are actively using AI techniques to monitor user activity to identify their anomalous behaviour in terms of a spike in activities or unusual activities by correlating the data received from multiple sources. In this direction, researchers are using an insider threat test dataset provided by CERT to provide insights for data leakage prevention using different temporal representations of user activity [107] or by daily activity summaries, e-mail contents and email network [108]. In contrast, Al-shehari and Alsowail [109] proposed a model that is only applicable for the identification of data leakage events during the sensitive period before an employee leaves an organization.

Automated data sensitivity detection is a method of identifying and classifying data by analysing, labelling and organizing the data into relevant categories (confidential, private and public) based on shared characteristics. It can empower the data leakage prevention techniques with the ability to monitor users' actions towards only particular relevant portions of sensitive data, rather than tracking all the data at all times. Alzhrani et al. [110] proposed an automated classification technique enabled by security similarity (ACESS) for mitigating the threat of sensitive data leakage from insiders. In 2021, Guo et al. [111] pointed out that sensitive information often resides in unstructured data, making unintentional data leakage easier. Therefore, they proposed a content and context-based sensitive information identification method using BiLSTM and an attention mechanism.

Advanced persistent threats (APTs) are a type of targeted cyberattack that lasts for a long time and is unnoticed by the target network's defences. The main purpose of this type of attack is to steal data rather than cause any damage. Researchers are working on the efficient capture of telemetry from endpoints, networks and clouds to integrate and analyse this diverse telemetry to uncover anomalies, indicators of compromise (IoCs) and other behaviour of interest [112,113].

5.2.3.2. Intelligent e-mail protection. Intelligent e-mail protection is a class of software solutions to prevent sophisticated e-mail focused cyberattacks. Traditionally, a spam e-mail was used as a tactic for hawking goods and services by sending unsolicited e-mails to bulk lists. Today, however, it is actively being used to spread malware, steal authentication credentials, or commit financial frauds. AI techniques are being used for automated protection against malicious spam e-mails.

Researchers are using supervised classification [114] and deep-learning based [115] techniques to identify spam in real time using dynamic incoming e-mail data, including general, view, subject, attachment, and content-related features. However, Gualberto et al. [116] and Nguyen et al. [117] confined their research efforts to phishing e-mail detection by analysing the content of e-mails.

5.2.3.3. Malicious domain blocking & reporting. Malicious domain blocking and reporting provides the next level of security for e-mail protection by catching any malicious network traffic resulting from opening the spam e-mail or attachment. Researchers use AI to identify suspicious websites for each DNS lookup and block access to malicious websites associated with malware, phishing, ransomware and other cyber threats.

The detection of malicious websites works by training machine learning algorithms with a rich collection of malicious and non-malicious website features. These features can be divided into four main categories: website design features [118], domain-based features [119–121], URL-based features [122,123], and hybrid features [124, 125].

In 2021, Cohen et al. [118] presented a new website categorization method to identify malware or crack websites based on the automated scraping and processing of thousands of visual and non-visual design features. Domain name-based features are very popular for the detection of malicious websites in the research domain. Researchers are using supervised machine learning models [119] and deep-learning [120,121] techniques to detect malicious domain names using classical domain-name features. Features extracted from the URL strings, including linguistic, lexical, contextual and statistical information are used to determine the malicious websites. Indrasiri et al. [122] and Vinayakumar et al. [123] provide the detection and analysis of malicious websites by using URL-based features as an input to the ensemble machine learning and deep-learning models, respectively.

Some researchers are also using hybrid features for malicious website identification to comprehend botnet detection [124] or phishing website detection [125]. These techniques use a combination of features related to domain name structure and DNS response such as resolution source, daily resolution amount, etc.

5.2.4. Information protection, processes and procedures

This area deals with the protection of information sources and assets consistent with the defined security policies, processes and procedures. It includes the protection of information, and the establishment, management and implementation of response, recovery and vulnerability management plans. Researchers are working on the application of AI techniques for AI-powered backup, and an AI-enhanced vulnerability management plan to maintain the processes and procedures to manage information protection.

5.2.4.1. AI-powered backup. AI-powered backup solutions are emerging to back up critical data and software components according to priorities and requirements for ensuring efficient backup. AI techniques are being used for dynamic backup scheduling [126] and optimized backup scheduling [127].

Qin et al. [126] designed a dynamic backup system with intelligent scheduling algorithms to improve the stability and predictability of the backup environment. The proposed system schedules the backup efficiently by determining which backup starts first and which storage is assigned to that backup for improving the efficiency. In contrast, Van de Ven et al. [127] used a two-dimensional Markov chain to model data backups and study the optimization of backup scheduling. At each time slot, the proposed technique examines a probabilistic backup policy to initiate the backup, regardless of the backup size.

5.2.4.2. AI-enhanced vulnerability management plan. A vulnerability management plan is a framework designed to proactively reduce the exposure to risk that has the potential to disrupt and impact the system. With the rise in reported vulnerabilities in recent years, it is more important to align the vulnerability management plan with the system's requirements and critical success factor. Researchers are using AI techniques to determine context-based, vulnerability risk scores and vulnerability exploitation trends to protect the assets and information systems in real time.

Context-based vulnerability risk scoring will help the analyst to prioritize the risk in the context of particular assets or information systems and enable them to take protective action. Zeng et al. [128] proposed a new risk prioritization method to assess the vulnerability risk by integrating the attacker model to capture the attacker's preference for exploiting vulnerabilities. The risk score is defined by the criticality of exploitation and the likelihood of the exploitation using the logic-reasoning engine.

A vulnerability exploitation trend will help the analyst to prioritize patching and remediation by predicting the vulnerabilities that will most likely be exploited. Researchers are using novel AI techniques to predict the exploitability [129] and solve the class imbalance problem to improve the performance of machine learning algorithms [130]. Here, Yin et al. [129] focused on the problem of vulnerability exploitation prediction using transfer learning to help experts prioritize the application of a patch. Yin et al. [130] also proposed a novel sequential batch-learning technique, called real-time dynamic concept adaptive learning, to address the concept-deviation and dynamic class imbalance issues in exploitability prediction.

5.2.5. Protective technologies

Protective technologies provide the security and resilience of systems and assets. These technologies use particular tamper-evident features to identify and deter attempts to breach, change, penetrate and extract information from the organization's assets. AI can be used to provide protective solutions in the form of log analysis tools, intrusion prevention systems, anti-virus/anti-malware solutions and protection by deception.

5.2.5.1. Log analysis. Log analysis is the process of reviewing computer generated event logs to proactively identify bugs, security concerns or other risks. AI-powered log analysis tools can automate the routine and repeated tasks to efficiently handle large amounts of distributed log data.

Bai et al. [131] tested the performance of a variety of supervised machine learning approaches for detecting the evidence of malicious remote desktop protocol (RDP) sessions using windows RDP event logs. Afzaliseresht et al. [132], on the other hand, proposed a new approach for data presentation using a storytelling technique to generate a natural language report for recognizing cyber threat information according to the users' level of knowledge.

Researchers are also working on providing solutions to solve variety and interoperability issues in log management. De la Torre-Abaitua et al. [133] and Eljasik-Swoboda and Demuth [134] worked on addressing the variety issue using intelligent methods for extracting and processing textual data from different sources for the acceptable log feature representation with the aid of information retrieval approaches. Similarly, Sisiaridis and Markowitch [135] worked on the security analytics of the heterogenous log data derived from different network sensors by employing automated feature extraction and feature selection techniques.

5.2.5.2. Intrusion prevention system. Intrusion prevention systems monitor the network traffic and then take an appropriate action to thwart the attack by reporting, blocking, dropping or resetting the connection. Researchers have proposed unsupervised isolation forest [136] and self-organizing incremental neural networks and SVM based intrusion prevention systems [137] for embedded systems in automotive electronics and IoT networks, respectively.

5.2.5.3. Anti-virus/Anti-malware solutions. AI-powered anti-virus/anti-malware solutions can analyse thousands of files and extract useful features to classify them as benign or malware. Researchers have created anti-virus programs for detecting malware using the features retrieved from the executables [138] or dynamic data analysis [139] as an input to artificial neural networks (ANNs) or recurrent neural network (RNN) models, respectively.

5.2.5.4. Protection by deception. Protection by deception is an advanced technique to protect critical documents after an attacker has penetrated the network. AI has been used to generate credible fake text documents to mislead cyberattacks. Karuna et al. [140] proposed the creation of decoy files to divert the adversary away from the real target when the adversary is already in the system. Their decoy text-creation approach uses a genetic algorithm that manipulates real document's comprehensibility to hard-to-comprehend, but believable fake documents.

5.3. Detect

The **detect** function enables the timely discovery of the cybersecurity events by developing and implementing appropriate activities to identify their occurrence. This function is crucial for the security as prompt detection will minimize the disruption. It includes activities for the timely detection of intrusions and anomalies along with the impact assessment, implementation of security continuous monitoring to verify the effectiveness of protective measures, and appropriate maintenance of detection processes to ensure the awareness of cyber events. AI can improve the detection speed by monitoring the internal and external information sources and swiftly correlating this information to detect the unusual activities to minimize the repercussions. Table 6 presents a summary of the main contribution of each research study to the detect function, along with the details of the solution category, AI use cases and the AI domain used. The solution categories along with a detailed review of the AI use cases are provided below.

5.3.1. Anomalies and events

The solutions in this area address the detection and classification of anomalous activities by establishing and managing baselines for operations and dataflows collected from multiple sources. These baselines are then used to detect and analyse events to understand attack targets and methods.

5.3.1.1. Intrusion detection system (IDS). An IDS is a set of tools and techniques to monitor a system and network traffic to analyse the anomalous and suspicious activities, with the aim to detect possible intrusions targeting the system. In the state-of-the-art research, IDS was implemented from three perspectives: binary classification, multi-category classification or both. Binary classification assumes two labels: normal and attack. Multi-category classification, on the other hand, deals with the problem of classification into three or more classes. In the case of IDSs, the multi-category classification differentiates between different types of attacks and provides users with more information to deal with the attack. Benchmark datasets are needed for the effective development and evaluation of both binary and multi-category intrusion detection systems. Therefore, only those studies that used benchmark datasets to evaluate their approach are included in this review.

In this direction, researchers used the ADFA-WD and ADFA-WD: SAA datasets for system call-based IDS [141], Aegean Wi-Fi (AWID) wireless networks dataset [142,172,184], BGP RIPE dataset for routing information services [143], datasets provided by the Canadian Institute of Cybersecurity [151-153,178,180-184,192-194], CIFAR-10 dataset of images [147,148], CTU-13 dataset for various botnet scenarios [149,150], Ethereum classic dataset depicting attacks on an open-source blockchain-based distributed computing framework with a smart contract [151], ICS Cyberattack gas-pipeline dataset [152,153,178,188,189], KDD99 and NSL-KDD datasets of network traffic records [146,148,154-160,168,176,177,179,180,182,184,186,187,190-192,194], Coburg intrusion detection datasets (CIDDS) of labelled flow-based datasets from anomaly detection [156], UNSW-NB15 dataset of raw traffic files for different types of attack [155,157,158,160-163,176,190,191],

Table 6
Summary of the primary studies focused on the detect function.

Solution Category	Use Case	Contribution	AI Domain	Author
Anomalies & Events	Intrusion Detection System	Binary classification	Learning	Ajayi and Gangopadhyay [141], Li et al. [143], Almlani et al. [144], Corsini et al. [145], Kumar et al. [147], Maimó et al. [149], Le et al. [150], Saveetha & Maragatham [151], AL-Hawawreh et al. [152], Zhang et al. [154], Blanco et al. [155], Nguyen et al. [156], Alhowaide et al. [157], Dutta et al. [161], Perez et al. [162], Singh et al. [163], Catillo et al. [164], Zhao et al. [165], Nedeljkovic & Jakovljevic [166], Liu et al. [168], Vidal et al. [169], Latif et al. [170]
		Binary classification	Planning & Learning	Granato et al. [142]
		Hyperparameter tuning of binary classifier	Learning	Choras & Pawlicki [146]
		Binary classification	Perception & Learning	Wu et al. [148]
		Binary classification	Planning & Learning	Vavra et al. [153]
		Solve class imbalance problem in dataset for binary classification	Learning	Binbusayyis and Vaiyapuri [158]
		Feature extraction from dataset for binary classification	Learning	Herrera-Semenets et al. [159], Rashid et al. [160], Elnour et al. [167], Leevy et al. [171]
		Multi-category classification	Learning	Iwendi et al. [173], Toupas et al. [174], Jagtap et al. [178], Asif et al. [179], Liu et al. [180]
		Feature extraction from dataset for multi-category classification	Learning	Abdulhammed et al. [172], D'hooge et al. [175], Shafiq et al. [183]
		Solve class imbalance problem in dataset for multi-category classification	Learning	Huang & Lei [176], Gupta et al. [177]
		Feature extraction from dataset for multi-category classification	Planning & Learning	Blanco et al. [181]
		Hyperparameter tuning of multi-category classifier	Learning	Pawlicki et al. [182]
		Binary and multi-category classification	Learning	Mikhail et al. [184], Basnet & Ali [185], Diallo and Patras [186], Ullah & Mahmoud [188], Li et al. [189], Zhang et al. [190]
		Solve class imbalance problem in dataset for binary and multi-category classification	Learning	Gupta et al. [187]
		Data visualization and binary and multi-category classification	Learning	Zong et al. [191]
		Feature extraction from dataset for binary and multi-category classification	Learning	Ieracitano et al. [192], Liu et al. [193], Xuan et al. [194]
Security continuous monitoring	Security monitoring	Data processing and correlation	Learning	Grammatikis et al. [195], Fausto et al. [196]
		Situational awareness	Learning	Kodituwakku et al. [197], Nikoloudakis et al. [198], Zhang et al. [199], Marino [200]
Detection Processes	Dark web investigation	Sentiment analysis	Communication	Al-Rowaily et al. [201], Deb et al. [202]
		Proactive threat intelligence	Learning	Ishikawa et al. [203]
		Proactive threat intelligence	Communication	Pantelis et al. [204], Schäfer et al. [205]
		Exploring discussion	Communication	Fang et al. [206], Huang & Ban [207]
	Automated assessment of different threat intelligence sources	Automated analysis of reports	Communication	Kim et al. [208], Sarhan & Spruit [209]
		Tweet processing pipeline for threat information extraction	Communication & Reasoning	Alves et al. [210]
		Tweet processing pipeline for threat information extraction	Communication	Dionísio et al. [211]
		Mining of tweets for threat information extraction	Communication	Saura et al. [212]
		Vulnerability intelligence	Communication	Georgescu et al. [213]
		Identification of threat evolution	Communication & Reasoning	Sleeman et al. [214]
		Generation of structured cyber threat intelligence records	Communication	Sun et al. [215]
		Issue threat warning	Communication	Sapienza et al. [216]
		Analysis of Chinese threat intelligence	Communication	Tsai et al. [217]
	Multi-lingual threat intelligence	Threat intelligence tool for Russian language	Communication	Ranade et al. [218]
	AI-powered honeypots	Honeypot-based bot detection	Learning	Memos & Psannis [219]
		Early-warning system based on distributed network of honeypots	Learning	Chatziadam et al. [220]

BOT-IoT dataset of normal and botnet traffic in an IoT network [157, 183], benchmark dataset of Hadoop logs for log-based anomaly detection [164], SEA dataset for behaviour logs of Unix users [165], Secure water-treatment dataset (SWaT) of network data gathered from different sensors and actuators in water-treatment plant [153,166,167,178], UGR'16 dataset of well-labelled traffic collected from tier-3 internet service provider [168], DARPA'99 dataset with an online and offline collection of real/synthetic samples in an experimental environment [169], UCM 2011 dataset with the real traffic traces [169], and TON_IoT dataset for new generations of Industry 4.0/IoT, Industrial internet of things (IIoT) [170].

Binary classification is the basic classification type and is the most studied in the literature for the intrusion detection problem. Most researchers worked on applying different machine learning classifiers [148-152,154-164,168-173,175-177] for misuse detection, but a few also worked on solving the hyperparameter optimization of classifiers [146], the class imbalance problem in datasets [158], and feature extraction from the dataset [159,160,167,171].

In multi-category classification, the dataset contains multiple disjoint classes and the data belonging to each class are given the same label. In the multi-category classification problem, there are several other classes of traffic besides normal, including denial of service (DoS), distributed denial of service (DDoS), remote-to-local or user-to-root attacks. Multi-category classification for intrusion detection also explores the application of different classifiers [173,174,178,179,180], feature extraction problem [172,175,181,183], hyperparameter tuning of classifiers [182], and handling the class imbalance problem in datasets [176, 177].

Some researchers have used both binary and multi-category classification in their approaches [191–201]. They have proposed different classifiers for the intrusion detection problem [184-186,188-190] and address the class imbalance [187], three-dimensional data visualization [191] and feature extraction [192–194] problems for the datasets.

5.3.2. Security continuous monitoring

Security continuous monitoring is real-time monitoring of information systems and assets to gain a clear insight into their environment and detect security events. AI can be used to automate monitoring by providing security intelligence using a dynamic, heterogenous information network that will process the logs of data generated from the monitoring of physical environments, networks, service providers, users and systems containing sensitive information.

5.3.2.1. Security monitoring. Security monitoring is a process that involves the collection, analysis and presentation of data from a variety of sources, with the aim to develop a universal solution to reveal the perpetrator's modes of operation and intentions. In this direction, researchers are actively working on processing and correlating data from heterogeneous sources [195,196] and situational awareness [197–200] for understanding security information.

An important problem in this area is processing massive, dynamic and heterogenous security information from an evolving collection of sources and algorithms. Thus, AI techniques are being used for a detailed analysis to track the associated security events in a fine-grained and reliable way. Grammatikis et al. presented a security information and event management system specifically tailored for the smart grid to detect, normalize and correlate cyberattacks and anomalies against a range of smart-grid application layer protocols [195]. Similarly, Fausto et al. [196] focused on the integration of logs belonging to both the physical and cyber domains and correlated their data together to detect potential anomalies in critical infrastructures.

Situational awareness provides a holistic and specific view of large-scale networks to enable security analysts and researchers to identify, process and comprehend information in real time. For this, Kodituwakku et al. [197] introduced a platform to process and visualize real-time, large-scale network data to not only monitor and study network flow data, but also to develop novel analytics. Similarly, Nikoloudakis et al. [198] presented an automated, situational awareness platform that uses real-time awareness features provided by the software-defined network (SDN) paradigm to perform a vulnerability assessment on network-enabled entities, their assignment to a connectivity appropriate slice and the continuous monitoring of the underlying infrastructure. Researchers [199,200] are actively working on maintaining holistic situational awareness (cyber and physical systems) in cyber-physical systems due to the tight integration of cyber and physical systems in mission-critical applications.

5.3.3. Detection processes

The detection processes involve activities to ensure the maintenance and readiness of detection procedures to reliably provide the threat intelligence and awareness of cybersecurity events. This involves activities to continuously improve and test the detection processes for efficient working. AI can be used to provide proactive vigilance on the internet using automated threat intelligence extraction from various web and internal resources. The resources include the dark web, threat intelligence sharing platforms and honeypots. The following use cases detail the use of AI techniques for the maintenance of detection processes.

5.3.3.1. Dark web investigation. A dark web investigation is a process that monitors internet resources in the dark web related to cybercrime. This process continuously monitors criminal forums and the black market to detect illegal activities and take appropriate measures to minimize risk. Researchers perform a dark web investigation by making a sentiment analysis of textual data from dark web forums [201,202], threat intelligence using dark web data [203–205], and identification of key attackers, their assets and areas of expertise [206,207].

Automated analysis tools are needed to identify the potential threats by analysing the language and targets of hackers without manually monitoring the large volume of posts made on the dark web. Sentiment analysis is used to automate the mining of opinion, views and emotions from the text using natural language processing (NLP). In this direction, a bilingual lexical resource (BiSAL) for the sentiment analysis of English and Arabic texts related to cyber threats, radicalism and conflicts was reported by Al-Rowaily et al. [201]. Deb et al. [202] constructed a methodology for predicting malicious cyber events by exploiting the behaviour of malicious actors via a sentiment analysis of posts on hacker forums. These forums on both the surface web and dark web have some predictive power that can be used as signals outside the network to predict attacks using time-series models.

Proactive threat intelligence is identifying and addressing the security risks before an attack occurs by collecting information from hacker forums and marketplaces offering products and services focusing on malicious attacks. Ishikawa et al. [203] presented a machine learning scheme to track attack activities and the evolution of infected devices using a dark web traffic analysis. Their analysis is based on the exploration of the underlying correlation between targeted network services, as indicated by the destination port numbers of scanning packets. In contrast, Pantelis et al. [204] and Schafer et al. [205] worked on devising specialized information retrieval techniques for the dark web to support cyber threat textual mining and business intelligence.

Exploring the discussion on dark web forums is an important problem for extracting key ideas to uncover popular topics, emerging threats, and key actors in the attacker community for the benefit of cybersecurity professionals. Researchers [206,207] have used topic modelling to extract topics, track the evolution of topics, and identify key hackers with their speciality topics and uncover their role in the underground market.

5.3.3.2. Automated assessment of different threat intelligence sources. An automated assessment of different threat intelligence sources will help

extract useful information from various sources, such as vulnerability databases, twitter, news sites, incident reports, and research articles to take timely actions to ensure the overall security of the system. This involves processing evidence-based knowledge from multiple sources about threats and actors to improve security and the decision-making process. The researchers are working to solve the problem of the quantity and heterogeneity of cyber threat intelligence sources and their formats to provide actionable intelligence.

Security professionals face a basic challenge when analysing cybersecurity reports, as immeasurable amounts of cyber information are produced daily, necessitating automated information extraction technologies to facilitate data retrieval and query. In this direction, researchers [208,209] have presented innovative methods for extracting the information from cyber threat intelligence reports using named-entity recognition to help security analysts gain accurate threat information as quickly as possible.

Timely and relevant information extraction from open-source intelligence (OSINT) published daily by users, security organizations and researchers is critical for maintaining a high level of security. Twitter is an important OSINT platform and a hub of cybersecurity intelligence due to its natural aggregation capability, timeliness, centre of public and private opinions and the presence of the most important cybersecurity feeds (e.g., NVD, ExploitDB, CVE, Security Focus). In this direction, tweet processing pipelines [210,211] and the mining of tweets to extract security issues [212] and relevance score of sources for intelligence verification [213] have been proposed.

Vulnerability intelligence is extracting the information about software and system vulnerabilities from public vulnerability datasets (e.g., CVE and NVD). This will help in the identification of vulnerabilities and attack vectors to prioritize security efforts and patching schedules. Georgescu et al. [213] proposed an automated system for diagnosing and detecting potential vulnerabilities in IoT systems using an IoT specific ontology and content-based extraction of the most appropriate vulnerabilities, considering the ongoing changes in the CVE database and the current situation of the IoT system.

The explosive growth of cybersecurity information on designated platforms, the dark web and social sites requires the development of automated tools for the identification of threat evolution [214], the generation of structured cyber threat intelligence records [215], threat warning [216], and cyber threat intelligence (CTI) analysis from local threat intelligence sources [217]. To identify the threat evolution, Sleeman et al. [214] used dynamic topic modelling to show the evolution of key topics in a time-stamped collection of cybersecurity documents. The growth and unstructured nature of the information shared on open-source threat intelligence publishing platforms (OSTIPs) make it challenging to automatically gather the CTI records. In this context, Sun et al. [215] proposed an automatic way to generate structured CTI data from OSTIPs using a combination of machine learning and natural language processing to achieve accurate, structured and detailed data that can be readily used by security tools and analysts for threat mitigation. The early threat warning systems help in defending against cyberattacks by giving a timely notification about possible incidents and security issues using information feeds. Sapienza et al. [216] mine the tweets of security experts and cybersecurity related blogs to issue a cyber threat warning. The threat intelligence tools targeting a specific language to help threat intelligence professionals gain a better insight into cyber threats in their local language for country-specific knowledge. In this direction, Tsai et al. [217] developed an automated system to analyse Chinese cyber threat intelligence to increase the threat intelligence visibility. This includes the development of an automatic classification system, a recommendation system and threat-labelling techniques.

5.3.3.3. Multi-lingual threat intelligence. The multilingual nature of the internet demands the translation of threat intelligence sources to draw a reliable conclusion. Here, third party translation engines are not suitable due to their lack of cybersecurity terminology as well as their inadequate privacy and confidentiality policies. Ranade et al. [218] highlighted the importance of developing threat intelligence tools for non-English languages. However, they only worked on Russian in their threat intelligence tool.

5.3.3.4. AI-powered honeypots. The main objective of using a honeypot is to study the techniques and behaviour of cyber attacks to improve the existing security system and be prepared for these types of attacks. AI-powered honeypots use a machine learning algorithm to predict the probability of an attack using data from multiple honeypots [219] or threat intelligence data from dark web sites [220] to prevent large-scale security events as early as possible. A method that uses a combination of machine learning and a honeynet-based detection method was proposed to determine whether an IoT device could be a component of a botnet [219]. In contrast, Chatziadam et al. [220] presented an early-warning intrusion detection system based on a distributed network of honeypots that uses darknets for data collection.

5.4. Respond

The **respond** function creates a roadmap for managing and limiting the impact of a potential cybersecurity event. This function is critical as it represents the first line of defence in incident handling and develops risk mitigation approaches for the future. This function includes planning ahead to develop effective processes to address the problem, analyse the incidents to determine their cause, scope and impact, incident containment, and the coordination of communication during and after an attack. By using AI techniques for response activities, incidents can be resolved more quickly and with less time and effort for security analysts. A summary of the primary studies focusing on the respond function is provided in Table 7. A detailed explanation of various cybersecurity solutions and AI use cases in each category are described below.

5.4.1. Response planning

This category is about planning well-maintained response procedures to follow during and after an incident to limit its scope and impact. This includes defining a contingency plan that captures various attack scenarios with the appropriate response action and incorporating lessons learned from ongoing incident response activities to update the plan. AI can be used to automate the response planning by establishing a dynamic case management tool to record, execute and update the contingency plan.

5.4.1.1. Dynamic case management. Dynamic case management tools are based on historical security breaches to record different attack scenarios and recommend appropriate response actions before an incident occurs. This helps in planning the response activities for specific types of breaches and in knowledge management for recording after an incident is closed. Research in this field has focused on automated response recommendations by matching the most similar incident from a knowledge manager using case-based reasoning and revise the knowledge manager after the incident.

Researchers use a case-based cybersecurity incident resolution system where domain experts describe the precedent model to save and retrieve precedents from the knowledge base. Researchers actively use a hierarchical structure [221,222], machine learning [223,224] and an ontological approach [225] to formalize the base of precedents. Kim et al. [221] described a hierarchical structure containing the attributes of the RFM (recency, frequency and monetary) technique to find similar cases for a rapid response to a security breach. Their approach considers the circumstances of the security event using its frequency and different attribute values. In contrast, Jiang et al. [222] used the hierarchical structure to store attributes from potential attack scenarios, such as the

Table 7
Summary of the primary studies focused on the respond function.

Solution Category	Use Case	Contribution	AI domain	Author
Response planning	Dynamic case management	Case-based incident resolution system using RFM	Reasoning	Kim et al. [221]
		Case-based incident resolution system using attack situations	Reasoning	Jiang et al. [222]
		Incident object description exchange format (IODEF) to retain, reuse and share the problem-solving experience	Reasoning	Nunes et al. [223]
		Incident resolution recommendation system	Learning	Kraeva & Yakhyaeva [224]
		Representation of security incident	Reasoning	Ping et al. [225]
Communications	Automated Responsibility Allocation	Adaptive and dynamic decision-making model for resource allocation	Learning	Shah et al. [226]
	Collaboration Support System	Asynchronous collaboration support systems	Learning	Lin et al. [227]
		Synchronous collaboration support systems	Reasoning	Thomas et al. [228]
Analysis	Automatic Incident Characterization	Severity assignment using multi-class categorization	Learning	Decastro-Garcia et al. [229]
	Alert Triage	Automated knowledge inference	Learning	Husak et al. [230]
		Alert grouping	Learning	Manganiello et al. [231]
		Alert prioritization	Learning	Dey et al. [232]
	Forensic Analysis	Intelligent attribution	Learning	Chen [233]
		Anomaly identification in forensic timeline	Communication	Studiawan & Sohel [234]
		Evidence correlation from different forensic devices	Communication & Reasoning	Amato et al. [235]
		Decision framework for optimizing forensic investigation	Reasoning	Nisioti et al. [236]
Mitigation	Automated Isolation	Attack localization on physical network	Learning	Sakhnini et al. [237]
		Isolation and replacement of infected devices	Learning	Maimó et al. [238]
	Automated Remediation	Selection of an optimal set of countermeasures	Learning	Nespoli et al. [239]
		Recommendation system for security analysts	Learning	Husak et al. [240]
		Recommendation system to prevent malware spread	Learning	Husak [241]
Improvements	Long-Term Improvements	Automated knowledge extraction	Communication	Piplai et al. [242], Peng et al. [244]
		Automated knowledge extraction	Learning	Woods et al. [243]

target organization, the attacker information, the affected resource, and its potential impact on the target. Nunes et al. [223] proposed the use of case-based reasoning with the incident object description exchange format to retain, reuse and share the problem-solving experience of cybersecurity incident resolution. The K-nearest neighbour algorithm was used to compute the case similarity. Kraeva and Yakhyaeva [224] proposed a recommendation system that maps security incidents into embeddings using neural networks and then finds the nearest incident embedding to recommend a similar case for resolution recommendation. Ping et al. [225] used an ontological approach to provide a standardized representation of security incidents to formalize the precedent base.

5.4.2. Communications

This activity helps to coordinate the communication between the stakeholders during and after a security incident. This includes the communication to support collaboration between security analysts during an attack, as well as cross-sector threat intelligence sharing to improve the response capability of the protection team during an emergency. This activity will also ensure the allocation of contingency roles and responsibilities when a response is needed. AI can support this activity in the following two use cases.

5.4.2.1. Automated responsibility allocation. Automated responsibility allocation can serve as an intelligent and adaptive decision support tool to assist security operation centre (SOC) managers in assigning incident response duties based on the nature of an incident, staff expertise and availability. There is no specific research study in the SLR that addresses this issue. However, Shah et al. [226] approached the problem of making the best decisions to allocate resources (time or additional workforce) to maintain an optimal level of operational effectiveness in a cybersecurity operation centre in the face of disruption due to a number of factors, such as a higher alert-generation rate, new alert patterns and analyst absenteeism. They developed a stochastic, dynamic programming-based, adaptive and dynamic decision-making model that is solved via reinforcement learning.

5.4.2.2. Collaboration support system. A collaboration support system is an information system that facilitates the efficient sharing of data, information and knowledge amongst various actors involved in incident response. These actors can be teams and employees inside or outside the organization. AI techniques are being used to support a cyber defence, collaborative, analysis support system [227,228] for both cross-sector threat intelligence sharing and community information sharing between security analysts.

These collaboration support systems fall into two main categories: asynchronous and synchronous. Asynchronous collaboration support systems do not provide real-time communication and team members can view information when it is convenient for them and jump in and out of the conversation as required. One such type of collaboration support system based on a message board was proposed for cross-sector threat intelligence sharing by Lin et al. [227]. They designed a multi-agent system-based monitoring mechanism in the blackboard-sharing module to solve the concurrency problems of threat intelligence sharing and to improve the execution efficiency of tasks. In contrast, synchronous collaboration systems provide real-time communication to support a real-time response by allowing security analysts to quickly exchange their findings and introduce effective task division. In this direction, Thomas et al. [228] proposed a system that provides visualization, real-time communication and the efficient conversion of massive amounts of data handled by numerous analysts to provide a comprehensive understanding of threats and the corresponding response actions.

5.4.3. Analysis

Analysis is the process of reviewing the security incident and response activities to ensure that the correct process is followed for handling an incident. This involves collecting and analysing the information about the incident to support incident characterization and alert the investigation to determine the severity and impact of the incident. It also supports the recovery activities by forensic analysis to collect and preserve the evidence for future litigation. AI can support the analysis process over the following use cases.

5.4.3.1. Automatic incident characterization. Incident characterization addresses the processes required to identify the incident category in accordance with the response plan. This includes the identification of incident criticality and its relationship with other incidents to automatically prioritize the incidents for further investigation. Decastro-Garcia et al. [229] proposed automated, multiclass, categorization models using machine learning techniques to assign the severity for different types of cybersecurity events.

5.4.3.2. Alert triage. Alert processing and triage is a way of efficiently and accurately investigating intrusion alerts to prioritize and analyse their relationship and determine whether or not they will be escalated to an incident response. AI techniques can be used to enable an effective alert triage by providing the automated knowledge inference [230], alert grouping [231] and alert prioritization tools [232] to identify threat alerts and escalate them for further investigation.

Knowledge inference is the process of applying logical rules to the knowledge base in order to evaluate and interpret new information. Based on this, Husak et al. [230] proposed a method that uses sequential rule mining to extract knowledge from shared intrusion detection alerts and use it to create predictive and customized blacklists.

Alert grouping and alert prioritization can address the problem of threat-alert fatigue by aggregating or triaging alerts of significance from massive threat-alert logs. The use of self-organizing maps and unsupervised clustering algorithms was proposed to group the security alerts that are likely to belong to the same attack scenario [231], while the notion of an anomaly score for each attribute of analysed events was defined to measure the priority of security events and find potentially anomalous events [232].

5.4.3.3. Forensic analysis. Forensic analysis is a post-mortem technique to establish a timeline for the attack and shed light on the extent and source of the breach, as well as the tools and methods employed to fully eradicate the threat and prevent it from recurring. This investigation also connects the dots between the fragments of evidence the attacker left behind to create a footprint that can be used as evidence in court or to support a prosecution. AI techniques can be used to help the incident response team in intelligent attribution [233], anomaly identification in a forensic timeline [234], evidence correlation from different forensic devices [235] and a decision framework for optimizing forensic investigation [236].

Intelligent attribution helps in inferring the cause of security events by discovering the relationships between different entities and events. Contextual based learning can be used to capture the status and offer hints for source attribution [233].

A forensic timeline provides information about the activities that occurred prior to, during and following a cybersecurity incident. A security incident documented in the forensic timeline can be considered as an anomaly that needs to be identified. A deep autoencoder can be used to build a baseline model for normal events in log files and an anomaly threshold for reconstructed values based on the constructed baseline to identify anomalies [234].

The main goal of forensic investigators is to detect and analyse the fraudulent activities in order to prepare a report for a court case. Investigators have various forensic tools at their disposal to examine devices, but these tools generate data in different formats and further complicate the analysis process. Therefore, Amato et al. [235] presented a new semantics-based method that assists forensic investigators during the analysis process by correlating the evidence found using different forensic tools.

The emergence of new adversarial techniques and technologies challenges the ability of a forensic analysis team to investigate incidents in a timely and effective manner with limited resources. To address these challenges, a novel model for a decision support system to recommend inspection actions using threat intelligence information harvested from the repository of known adversarial tactics, techniques and procedures for the optimization of forensic investigation was proposed [236]. This model takes into account the probability relationship and proximity values between the potential attack activities, current investigation findings, and the available budget for the investigation.

5.4.4. Mitigation

Mitigation involves a set of activities to prevent the expansion of security events and remediate their effects to eliminate any long-term consequences of the security breach. This is a critical step that not only contains the incident, but also mitigates the new vulnerabilities or documents them as an accepted risk. The following use cases explain the use of AI techniques in security incident mitigation.

5.4.4.1. Automated isolation. Automated isolation is based on the principle of automatically isolating a device or set of devices in response to the detection of an indicator of compromise (IoC). This includes disconnecting the device or set of devices upon infection or finding the high-risk infected users to give them more attention. The automated isolation works by localizing the attack on the physical network and by isolating or replacing the infected devices or users on the network using network function virtualization (NFV) and software-defined networking (SDN).

The localization of an attack to specific features or measurements in the system is essential to assist cybersecurity professionals in mitigating the impact of the attack in communication networks by linking the attacks to particular locations within the physical system. Sakhnini et al. [237] proposed an attack classification and localization model for the physical layer in the smart grid. The model uses ensemble and representational learning for attack classification and the implementation of a Chi-squared algorithm to correlate the attack scenario with the specific features and localize the attack to a set of specific measurements or a specific location in the system.

A real-time mitigation system is also needed in integrated clinical environments to deal with cybersecurity incidents quickly and efficiently. Maimo et al. [238] have proposed the use of machine learning techniques to detect and categorize the spread phase of ransomware attacks and the use of the NFV and SDN paradigms to contain ransomware propagation by isolating and replacing the infected devices.

5.4.4.2. Automated remediation. Automated remediation is a guided problem resolution process that automatically performs remedial actions with simple scripts or powerful context-aware recommendation systems. AI techniques are being used to select the optimal set of countermeasures for threat elimination [239], in recommendation systems to assist security analysts with a resilient configuration and the orchestration of security tools [240], and for tracing the attacker's lateral movement [241].

Developing rapid and effective response capabilities against disruptive cyberattacks is a fundamental pillar of defensive cybersecurity, and the selection of the best set of countermeasures to respond to threats is an important research problem in determining the best reaction in a fully automatic fashion. Nespoli et al. [239] proposed a new cybersecurity reaction technique based on artificial immune systems to select and enforce the best combination of atomic countermeasures for the assets of the protected system that are exposed to risk.

Another important research problem are decision support systems and intelligent recommender systems that can help security analysts quickly and effectively defend resources and services against cyberattacks. Recently, Husak et al. [240] proposed a recommender system to suggest the most robust configuration for a critical infrastructure and then orchestrate the network security tools for rapid mitigation measures. The recommendation system was used to swiftly stop the spread of malware or to track the lateral movement of the attacker [241].

5.4.5. Improvements

Improvements help ensure lessons are learned from incident detection and response activities. This includes updating the response plans and strategies according to the lessons learned. Researchers have used AI for automated knowledge extraction [242–244] based on incident reports.

5.4.5.1. Long-term improvements. Knowledge extraction from incident and threat intelligence reports provides credible information to security analysts that can be used to detect or find patterns indicative of cyber-attacks. Piplai et al. [242] proposed a system to extract knowledge from after action reports, aggregate it by grouping together comparable items, and display the retrieved knowledge in cybersecurity knowledge graphs. These graphs help security analysts to find similarities between different cyberattacks. In contrast, Woods et al. [243] described a knowledge extraction process based on data mining techniques to identify subsets of the indicator and incident landscapes for which the complete incident information might be useful to security analysts and decision makers. Peng et al. [244] worked on extracting knowledge about threat actions based on the conditional co-occurrence degree from threat related articles. Their work also applies machine learning to classify different threat related articles based on the similarity features of the article.

5.5. Recover

The primary goal of the **recover** function is to maintain resilience planning and the timely restoration of capabilities or services that were impaired due to a cybersecurity incident. This encourages a prompt return to normal operations in order to lessen the impact of a cybersecurity event and distil important information in the form of lessons learned. This function can serve as a roadmap for returning to normal operations with the help of the following categories of cybersecurity solutions. Table 8 summarizes the main contributions of primary studies focused on the recover function.

5.5.1. Recovery planning

Recovery planning involves the maintenance, testing and execution of the processes and procedures to restore the systems or assets affected by cybersecurity incidents. This entails the timely restoration of lost data and capacities that were impaired to ensure that everything is functioning as needed. AI-based recovery planning can automate the data and system recovery along with the deletion of malware or contaminated data in the face of a cybersecurity event. However, no significant research work in this domain was found in the SLR.

5.5.2. Improvements

This solution category involves a review of the security event to improve the recovery planning processes by learning from the security breach. It involves an amendment to the recovery plan and processes based on lessons learned and a review of existing strategies to match the security goals and objectives. AI can be used to automatically review the existing strategies, incident report and audit logs to find opportunities for improvements from the latest breach for future response planning.

Table 8
Summary of the primary studies focused on the recover function.

Solution Category	Use Case	Contribution	AI domain	Author
Improvements	Analysis and aggregation of incident reports	Post-mortem analysis of vulnerabilities	Communication	Meyers and Meneely [245]
		Aggregation of cybersecurity incident reports	Planning & Learning	Carriegos et al. [246]

5.5.2.1. Analysis and aggregation of incident reports. The analysis and aggregation of security incident data and reports can provide valuable insights that can be used to provide recommendations and directions to advance cybersecurity to the next level. However, the management and analysis of incident data is an arduous and time-consuming task. AI techniques can be used for efficient data collection, aggregation, information extraction, visualization and the prediction of heterogenous incident data [245,246]. Meyers and Meneely [245] developed an automated methodology for the post-mortem analysis of vulnerabilities to find complex relationships between them using natural language processing. Carriegos et al. [246] proposed an effective method to aggregate cybersecurity incident reports to determine and define accurate measures of cybersecurity incidents to make predictions and deploy security policies accordingly.

5.5.3. Communication

Communication helps in recovery by coordinating communication activities between internal and external parties. Although there is no significant research article in this area in the SLR, the ***provision of propriety platform*** for sharing information on the latest security breaches or threats is a research area with high potential to help ensure the cybersecurity of critical infrastructure.

6. Descriptive analysis

After the state-of-the-art analysis, the statistical distribution of the primary studies is shown in terms of the taxonomy, the AI technique used, the publication type, the publication year, and the geographical distribution of the research advances to answer RQ3.

6.1. Distribution by article type

Of the 236 articles selected for the SLR, 101 articles (43%) are sourced from conference proceedings and 135 (57%) from peer-reviewed journals, as shown in Fig. 5.

6.2. Distribution by publication year

The time span of this review was 2010 to February 2022. As Fig. 6 illustrates, AI for cybersecurity was a relatively under researched topic till 2016, with only a handful of studies published in peer-reviewed journals and at conferences. Only in the last four years (2018 to 2021) has there been an increase in interest in AI as a cybersecurity research

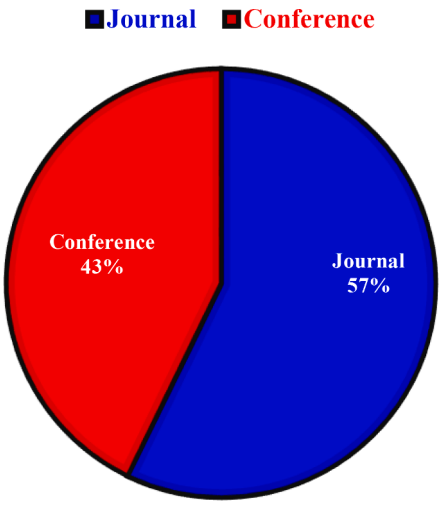


Fig. 5. Article distribution by type of publication.

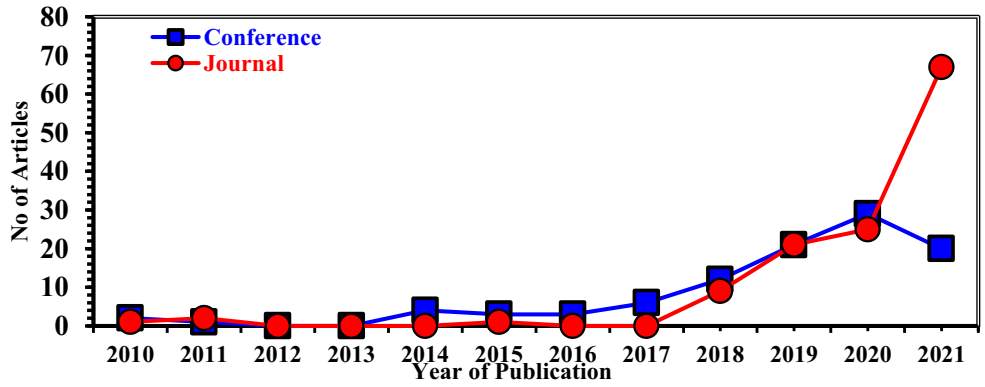


Fig. 6. Annual distribution of articles.

topic. Fig. 6 also shows the implications of Covid-19 in terms of a smaller number of conference publications in 2021 compared to previous years. In contrast, the number of journal publications in 2021 increased by nearly 2.6 times compared to the previous year.

6.3. Distribution by geographical region

The geographical distribution of the authors of the referenced articles is presented here according to the five continents: Asia, America, Europe, Africa, and Oceania. The joint articles by authors from different continents are presented as a collaborative region. For the selected journal publications (see Fig. 7(a)), 30% of the researchers are located in Europe, followed by 22% researchers in Asia, and 22% in America. Oceania has relatively few research papers on the topic, only 4%. Africa has zero journal publications in the selected pool. The remaining 22% of articles are collaborative efforts by researchers from different continents.

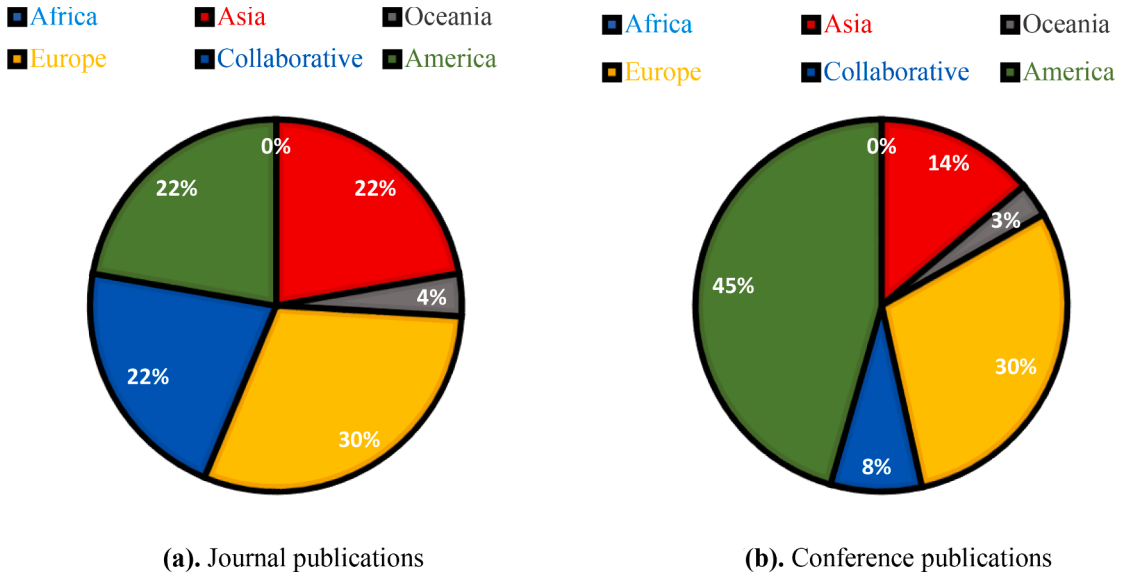
Regarding conference publications (see Fig. 7(b)), 45% of the authors are located in America, followed by 30% in Europe and 14% in Asia. Oceania has very few conference publications on the topic, only 3%. Similar to journal publications, Africa has zero publications in the selected studies. The remaining 8% of the articles are the collaborative work of researchers from different continents.

6.4. Distribution by NIST cybersecurity function

The primary studies were read in full and relevant data were extracted and summarized in Section 5. Each of the primary studies was determined to have a focus on a specified cybersecurity function. The identified studies were classified into five main groups, based on the cybersecurity function that is in their primary focus: identify, protect, detect, respond and recover. Fig. 8 illustrates the distribution of studies regarding these five identified categories. amongst the journal publications (see Fig. 8(a)), 36% of the studies focused on the detection of anomalies and cybersecurity incidents. Detection governed by machine learning algorithms can lead to automatic attack detection and defence in a timely manner. The second most popular category with 28% is identify, followed by protect and respond, 25% and 10%, respectively. Very few studies reported in 2021 and 2022 focused on the use of AI for recovery.

Conference publications show the same trend with different percentages, as shown in Fig. 8(b). Most conference articles on the application of AI for cybersecurity were published in three main categories: identify (40%), detect (31%) and protect (17%). Respond and recover accounted for the remaining 12%, at 11% and 1%, respectively.

Figs. 9(a) and 9(b) map the percentages during the past 6 years. The figures show that the identify, protect and detect categories have increasing numbers of publications with years, whereas, the respond and recover functions have only started getting attention recently.



(a). Journal publications **(b).** Conference publications

Fig. 7. Geographical distribution of primary studies related to AI for cybersecurity.

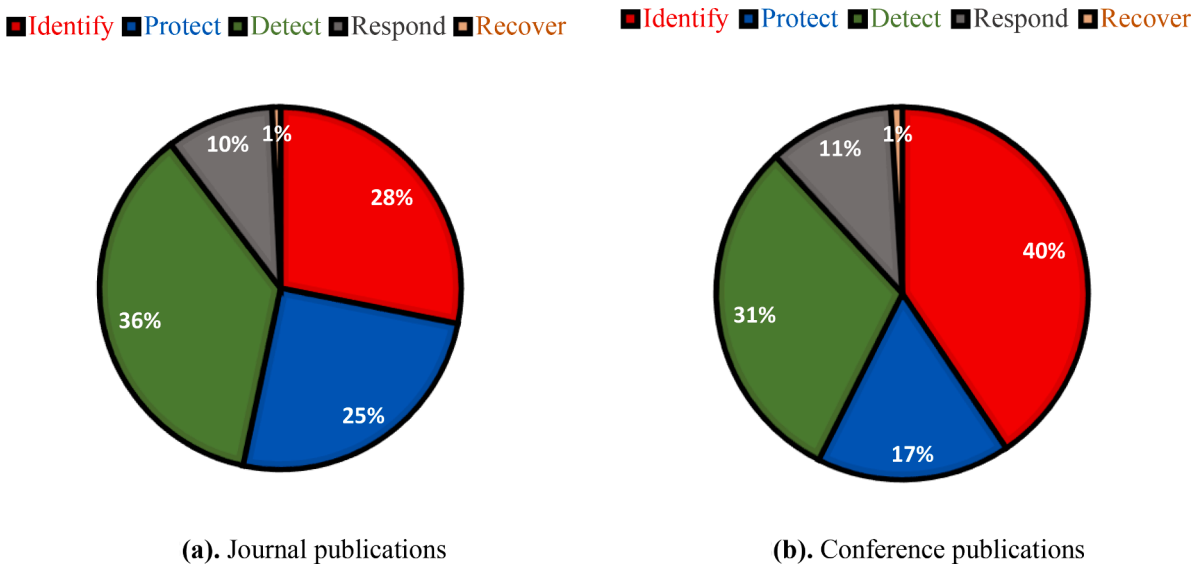


Fig. 8. Distribution of primary studies with respect to the NIST functions.

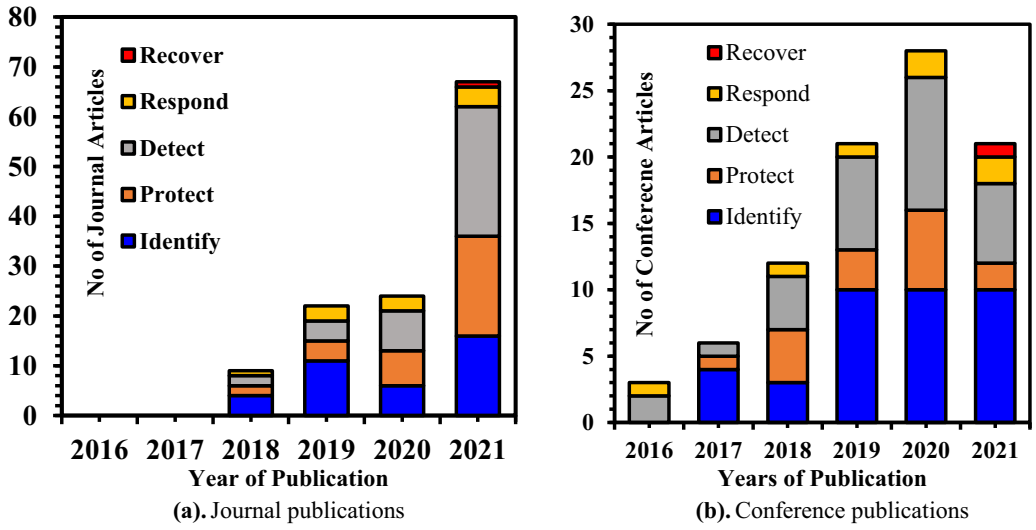


Fig. 9. Distribution of primary studies with respect to the NIST functions during the past 6 years for journal and conference publications.

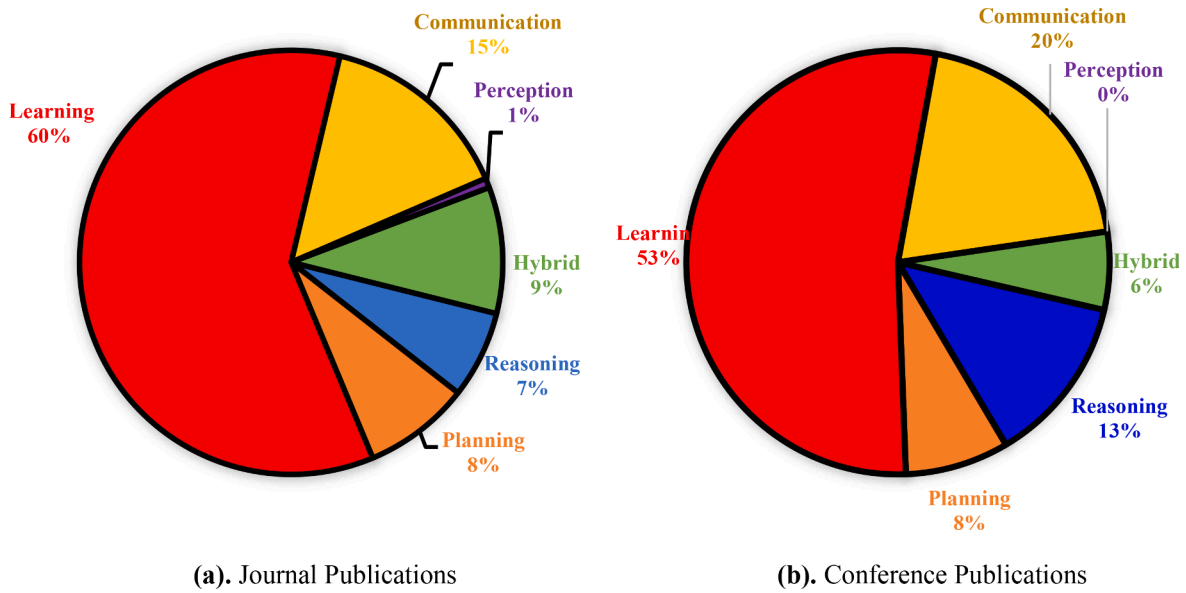
6.5. *Distribution by the AI technique used*

Another feature examined in the qualitative analysis was the AI technique used in the studies. AI technique refers to the methodological approaches that can be specified, including algorithms, architectures, data or knowledge formalisms and algorithms. The core AI domains, i.e., reasoning, planning, learning, communication and perception, proposed by AI Watch are used for the analysis. The primary studies in the reasoning domain tackle the way machines transform data into knowledge or infer facts from data. Planning studies focus on automated planning for the design and execution of strategies with carefully optimized solutions. The studies in the learning domain address solutions for automatic learning, predicting, adapting, and reacting to changes. Studies in the communication domain emphasize the machine's capacity to recognize, interpret, comprehend or produce information in spoken or written human conversations. The primary studies in the perception domain address sensing the surrounding environment with vision and hearing. The AI-domain focus was determined for each of the primary studies. Figs. 10(a) and 10(b) illustrate the distribution of studies regarding the five identified AI domains in journals and at conferences.

Figs. 11(a) and 11(b) show the shares of AI techniques used in the past 6 years in journals and conference publications. It is clear from the figures that learning is the most widely used technique for cybersecurity applications, followed by communication.

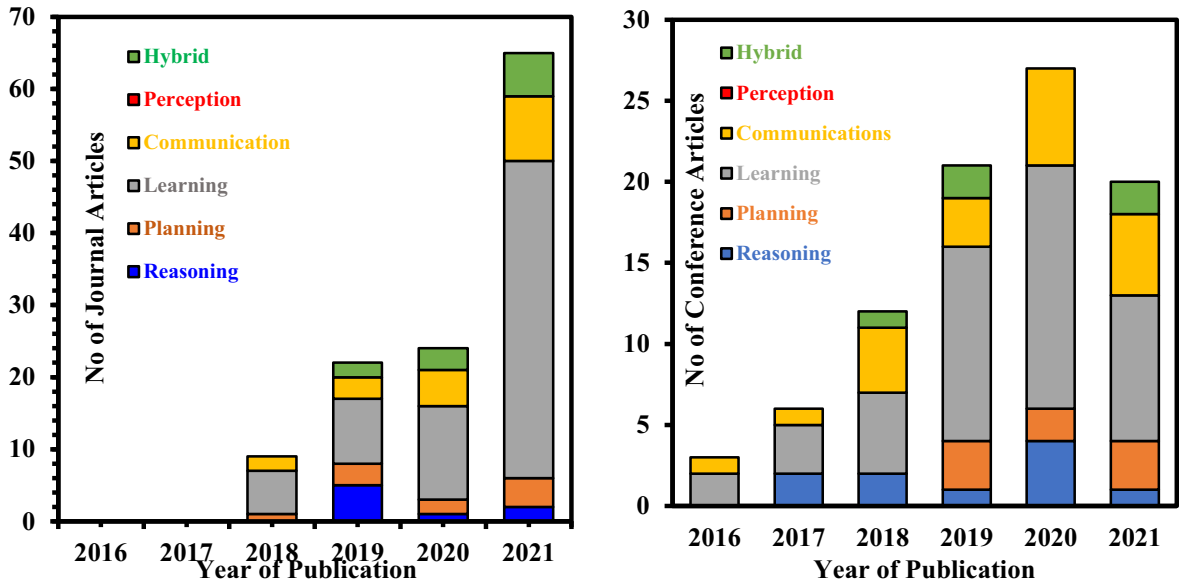
7. Research gaps

To answer the fourth and final research question of this paper (RQ4), the literature relevant to our research questions was scrutinized to highlight potential research gaps and identify opportunities for future AI for cybersecurity research. A key element in conducting AI for cybersecurity research is identifying emerging application areas, appropriate resources (e.g., data sources and management, computational infrastructure, etc.), and advanced AI techniques for the successful adoption of AI for cybersecurity. This section provides useful directions for future research in four main areas: (i) emerging areas of cybersecurity applications, (ii) data representation, (iii) advanced AI methods for cybersecurity, and (iv) research and development of new infrastructure.



(a). Journal Publications **(b).** Conference Publications

Fig. 10. Distribution of primary journal and conference studies with respect to the AI techniques used.



11(a). Journal publications **11(b).** Conference publications

Fig. 11. Distribution of use of AI domain during the past 6 years.

7.1. Emerging areas of cybersecurity applications

Advancing AI for cybersecurity requires a solid foundation of application domains. Moreover, the future outlook necessitates both the automation of new cybersecurity activities and the continuing enhancement of existing ones. The emerging application areas that the research community can address are described below:

- **Automated Retrieval of key risk indicators:** There are no research articles in the SLR corpus that address the automated retrieval of risk indicators in real time. Thus, it is a temping future research direction to develop an early-warning system to indicate risk development over time due to policy violations, red flags or other symptoms. The automatic retrieval of key risk indicators, such as the presence of an unpatched system, the number of attempted breaches, the mean time between failures, etc., and converting them into knowledge will be beneficial to prevent a cybersecurity breach by timely remediation of the risk.
- **Detection of new attacks:** Defending against zero-day attacks is one of the most challenging aspects of modern cybersecurity. A zero-day attack is a cyberattack that targets a new, not yet widely known, software vulnerability. Of course, defending against something you do not know exists is a major hurdle. For this, complete visibility across the entire information technology environment, including endpoints, networks and cloud, is needed.
- **Predictive Intelligence:** Predictive analytics can facilitate automated decision making for routine cybersecurity tasks, including attack path prediction, malware prediction, data triage, spam filtering, vulnerability classification, security prediction, and mission mapping. Despite the fact that these tasks are widely used and clearly defined, many of the commonly used techniques are either not automated or have a high rate of false positives. Advanced

AI-based predictive analytics can be used to help tackle some of these particular issues. Deep Bayesian forecasting, burst detection, deep generative modelling with temporal constraints, time-based neural graph networks, and other techniques are amongst the promising predictive analytics techniques. Each strategy can be improved by incorporating industry-specific policies, information, tasks and needs. Security operation centre (SOC) analysts and CTI professionals, particularly those at the tactical and operational levels inside their businesses, are amongst those who can profit from improved forecasts.

- **Multi-lingual threat intelligence:** In the SLR corpus, only one study analysed the threat intelligence understanding across unfamiliar languages. The multi-lingual nature of the internet makes it more challenging for the cybersecurity community to continue to strategically mine threat intelligence from social media, blogs and dark web markets. Non-English-speaking nations are both the top sources and the top targets of cyberattacks, according to the Symantec research [247]. Therefore, studies should be conducted utilizing datasets with non-English content to assess the effectiveness of text mining in other languages. Additionally, the majority of pre-processing tools and libraries only support the English language. New research can provide tools that target other languages or create language translators for private security settings.
- **AI-powered Cyber Defence and Resilience:** Organizations can apply the right security controls automatically using data from their analytics. Examples include automated threat modelling, automated patching, automated remediation and mitigation, and automated network segmentation and reorganization. SOC analysts and operators can benefit greatly from intelligent job automation. Enhanced AI agents, reinforcement learning, actor critical networks, selected defensive, adversarial learning techniques, and Bayesian networks are some of the modern AI techniques that enable these tasks. Future studies can examine how each technique can produce suitable cyber protection for various industries (e.g., scientific cyber infrastructure, enterprise IT, sensor-based environments, etc.).
- **Data Breach Prevention and Discovery:** In recent years the prevalence of data breaches has resulted in losses for both businesses and consumers. This topic needs to receive considerable attention from cybersecurity researchers as the majority of the existing research concentrates on observing the behaviour and activities of insiders or the analysis of telemetry from endpoints to detect the presence of insider threats and advanced persistent threats (APTs). However, the research on the identification of sensitive data to prevent accidental data exposure is still lacking. The machine learning, and natural language processing-based AI techniques can be used for sensitive data discovery, as well as monitoring and control of the data flow across end points to prevent or analyse the data leakages in big-data scenarios. The studies should also consider the analysis of dark websites to discover any accidental data exposure so as to regain control and save the reputation.
- **Fake Document Generation**: The protection of critical digital assets such as intellectual property and national security data is of utmost importance in this era of cyberwarfare. In the SLR corpus only one study reported the use of the automated generation of believable and interactive fake documents to protect critical documents after a network's penetration. This fake document generation using AI is a relatively new concept to safeguard sensitive material by falsifying the information to create numerous fake versions of any document. Thus, this topic needs to receive increased attention from the cybersecurity community.
- **Context driven Alert Processing & Triage:** Thousands of warnings and events that are meticulously gathered for threat analysis overwhelm security teams every day, necessitating the attention of experienced threat intelligence experts. Current research articles have attempted to overcome this problem using high-level management to correlate security alerts, considering their logical relationship, including their prioritization, before forwarding them to users. However, they still lack consideration of different event contexts in a particular network environment. Researchers can use different language models to implement representation learning for the event context and work on developing adaptive methods for dynamic network environments. In addition, further efforts are needed in the area of data visualization and online update capacity to visualize the relationships between alerts and develop an efficient triage system for alerts.
- **AI-Powered Incident Response:** As the typical time an attacker needs from initial compromise to complete takeover of an enterprise's infrastructures has been greatly shortened, it is also critical to automate response efforts to effectively mitigate, contain or outsmart attacks. Incident response process automation requires the documentation of knowledge acquired from past security incidents and events triggered by the application of a particular solution along with the recording of new threat patterns and their traits over time. This knowledge can then be used to create automated incident response playbooks that can make recommendations, conduct resource allocation or allocate responsibilities based on their expertise, availability or case history. These automated security playbooks will fuel the proactive defence and help in staying ahead of increasingly sophisticated and ever-evolving threats. Moreover, in future, the sharing of these standardized security playbooks on threat intelligence platforms will allow different organizations to consume such information in response to an incident at a machine time.

7.2. Data representation

For AI to be effective, it is crucial to have good data. The biggest challenges are the selection of appropriate datasets for training and handling the variety and velocity of the data. Therefore, the importance of data representation and quality, recency mining, and context awareness for the training and modelling of AI models for cybersecurity applications is described below:

- **Refined Data Representations.** The representation of data is critical to the performance of AI algorithms. A flattened feature vector is the most popular method for describing cybersecurity so far. Despite being widely used, this method ignores important linkages that are evident in the data it depicts (e.g., sequences). For this reason, this representation, when used in production contexts, can produce noticeably worse outcomes. Future researchers in AI for cybersecurity could carefully evaluate how cybersecurity data exists in the environment they are interested in and carefully choose a suitable alternative that best depicts the phenomenon of interest to mitigate this difficulty. For instance, file systems and applications within virtual machines can be visualized as both trees or graphs to capture their dependencies (due to their hierarchical nature). Grids, sequences and non-Euclidean representations are additional potential representations (e.g., tensors, cubes). Additionally, key data properties, organizational needs, and pertinent social behavioural economics theories may all be taken into consideration when choosing an acceptable representation.
- **Context-awareness in cybersecurity:** Current research on cybersecurity typically starts with the pertinent cyber data, which contains several low-level features. Such datasets can be used to apply data mining and machine learning approaches to find a coherent pattern that accurately explains them. However, to decide whether or not suspicious activity is present, broader contextual information such as temporal and spatial relationships between events, or connections and dependencies can be used. For instance, while security professionals may not view individual connections as malevolent, other methodologies may classify them as DoS assaults. Therefore, the inability to foresee dangers or assaults using contextual knowledge is a significant weakness of earlier efforts in cybersecurity. Context-

aware adaptive cybersecurity solutions could therefore be another study area in the field of AI for cybersecurity research.

- **Incremental Learning & Recency Mining:** To provide data-driven decisions, machine learning-based security models often use a lot of static data. However, the behaviour patterns of users and malicious adversaries might not be static and can vary greatly over time. Therefore, for predictive analytics in tackling common cybersecurity tasks, like data triage, spam filter, vulnerability categorization and mission mapping, recent behavioural patterns and accompanying machine learning rules are more likely to be interesting and significant than older ones. As a result, another challenge in AI for cybersecurity research could be effectively applying the idea of recency analysis in cybersecurity solutions.

7.3. Advanced AI methods for cybersecurity

The more sophisticated AI techniques are needed to realize the full potential of the aforementioned data sources, application areas and data representations. amongst the many alternatives, three crucial new techniques – multiple data source analysis, explainable AI (XAI), and augmented intelligence (human-AI interfaces) – can have a major impact on the development of practical and useable AI for cybersecurity.

- **Multiple Data Source Analysis:** A major drawback of the current AI research and practice landscape for cybersecurity is the isolated use of individual datasets. This is frequently brought on by a lack of access to multiple datasets (common in academics) or a failure to comprehend the relationships between different datasets. A lack of simultaneous processing of multiple datasets can lead to an incomplete assessment of the environment. Future AI research for cybersecurity could try to take advantage of the characteristics of different data sources in a more comprehensive way to address this problem. Entity matching based on deep learning, short-text matching algorithms (like deep structured semantic models), multi-view methods (like multi-source), and multi-task learning techniques are the promising approaches for multiple data source analysis. The successful fusion of multiple datasets can lead to newly derived attributes, enhanced risk management (such as vulnerability assessments), and a comprehensive understanding of an organization's cybersecurity posture.
- **Application of Explainable AI (XAI):** Knowing how and why an algorithm made its initial conclusion is crucial in the field of cybersecurity. Unfortunately, current AI-based algorithms lack transparency in their decision-making process. Despite offering unparalleled performance in high-impact cybersecurity applications like dark web investigations, vulnerability assessments, and others, they are notorious for their "black box" nature. Future AI research for cybersecurity can look at how interpretable and explicable AI can improve an algorithm's performance and open their black-box character to minimize these constraints, increasing their acceptance and trustworthiness for important cybersecurity stakeholders.
- **Augmented Intelligence** (Human-AI Interfaces): According to many cybersecurity experts, AI-based algorithms and systems should not be used exclusively for cybersecurity decision-making. Instead, to enable better decision-making processes, AI-based approaches should be firmly interwoven with human action (for instance, having a security analyst as an active member in the analysis process). These methods, also referred to as augmented intelligence or human-AI interfaces, have the potential to significantly outperform the use of an algorithm or a single human. The breadth, scope and depth of human-AI interaction in critical and fundamental cybersecurity tasks has not yet been adequately researched, but this is sorely needed. Such research would necessarily need to take a multidisciplinary approach, specifically incorporating perspectives from psychology, cognitive science, human-computer interaction, and other fields.

7.4. Research & development of a new infrastructure

AI is gradually becoming a crucial component of cybersecurity to benefit organizations of all sizes and in a variety of industries to increase the efficiency of their cybersecurity. Therefore, it is necessary to research and develop new infrastructures to support the AI technology by dealing with enormous volumes of internal system data along with external security research feeds to provide the real-time cybersecurity considering global and internal security events. Key research gaps for the successful implementation of AI in cybersecurity by organizations and cybersecurity researchers are described below:

- **Lack of threat intelligence platforms (at national and international levels):** Cyber reality is very complicated and dynamic; there are always new threats, and attacks are specifically designed to circumvent a known potential scenario. Therefore, propriety platforms are needed to enable collaboration between designated peers to discuss and share the latest threat data. The design of threat intelligence platforms to offer a flexible, adaptable and networked method of sharing threat information that relies mostly, but not exclusively, on designated information sharing hubs at national and international levels is lacking at present. The government, owners and operators of critical infrastructure, as well as other entities, will benefit from the effective and efficient sharing of accurate, usable, timely, and relevant threat information shared by threats intelligence platforms. By enhancing situational awareness and facilitating effective risk-informed decision-making, such sharing improves the security and dependability of critical infrastructure.
- **Lack of new, real-time or broader datasets:** Datasets are the most important component in AI for cybersecurity. The majority of the available datasets are outdated and might not be sufficient to comprehend the most recent behavioural patterns of different cyberattacks. This survey noticed that many studies applied the AI techniques to the same dataset. For example, most studies used DARPA98, KDD99, NSLKDD, and CICIDS2017 for detection. These studies clearly lack an evaluation of their techniques using the most recent and different datasets. Additionally, the validation of specific context-research across numerous datasets enables analysis across different scenarios.

8. Limitations

The presented SLR provides valuable information on the intersection between cybersecurity and AI techniques, along with the identification of research gaps to feed future research. Nevertheless, our study misses the articles that are published in scientific databases other than Scopus or used different keywords. Also, recent publications (after February 2022) are not included in the studied literature due to time spent on analysing the selected primary studies to obtain reliable results.

9. Conclusions

This SLR study examines the current state-of-the-art research on AI applications for cybersecurity. This was achieved by identifying 236 primary studies out of 2395 related articles from the Scopus database over a 13-year period (2010 to February 2022). The presented study discusses the different AI techniques applied in the cybersecurity domain and which cybersecurity activities have taken advantage of the AI technology. The selected literature is analysed in terms of (i) the presented taxonomy of AI in cybersecurity, (ii) the frequency of publication by year, (iii) the frequency of publication by geographical region, (iv) the cybersecurity contribution type, and (v) the type of AI technique used.

This SLR examined the "how" and "what" of the existing research on AI applied to cybersecurity with an in-depth exploration of specific use cases and the theoretical basis of the research. This study contributes to

the body of knowledge by analysing the evolution of AI applications in the cybersecurity domain and identifying research gaps. The evolution of AI in cybersecurity was studied with respect to different functions, solution categories, specific use cases, and the type of AI technique used. The results of the analysis revealed that the number of publications is increasing, but more attention must be paid to the acquisition and representation of historical data related to different cybersecurity functions to implement practical AI-based cybersecurity solutions. The main contribution of this study is the classification of the primary studies to integrate the state of literature in this area and to comprehend the significance of AI for cybersecurity. In addition, the article has proposed future research directions to address emerging issues for the successful adoption of AI for cybersecurity.

CRediT authorship contribution statement

Ramanpreet Kaur: Conceptualization, Methodology, Validation, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Dušan Gabrijelčič:** Methodology, Project administration. **Tomaž Klobučar:** Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Tomaž Klobučar reports financial support was provided by the Slovenian Research Agency. Tomaž Klobučar reports financial support was provided by the Government Information Security Office of the Republic of Slovenia.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by the Slovenian Research Agency, ARRS (V2–2147 and P2–0037) and the Government Information Security Office of the Republic of Slovenia (V2–2147).

References

[1] Bhardwaj, M.D. Alshehri, K. Kaushik, H.J. Alyamani, M. Kumar, Secure framework against cyber-attacks on cyber-physical robotic systems, J. Electron. Imaging 31 (6) (2022), 061802-061802.
[2] P. Chithaluru, A.T. Fadi, M. Kumar, T. Stephan, Computational intelligence inspired adaptive opportunistic clustering approach for industrial IoT networks, IEEE Internet Things J (2023), https://doi.org/10.1109/JIOT.2022.3231605.
[3] M. Barrett, Technical Report, National Institute of Standards and Technology, Gaithersburg, MD, USA, 2018.
[4] I. Wiafe, F.N. Koranteng, E.N. Obeng, N. Assyne, A. Wiafe, S.R. Gulliver, Artificial intelligence for cybersecurity: a systematic mapping of literature, IEEE Access 8 (2020) 146598–146612.
[5] Z. Zhang, H. Ning, F. Shi, F. Farha, Y. Xu, J. Xu, F. Zhang, K.K.R. Choo, Artificial intelligence in cyber security: research advances, challenges, and opportunities, Artif. Intell. Rev. 55 (2022) 1029–1053.
[6] J. Martínez Torres, C. Iglesias Comesaña, P.J. García-Nieto, Machine learning techniques applied to cybersecurity, Int. J. Mach. Learn. Cybern. 10 (10) (2019) 2823–2836.
[7] T.C. Truong, I. Zelinka, J. Plucar, M. Čandík, V. Šulc, Artificial intelligence and cybersecurity: past, presence, and future, in: Artificial intelligence and evolutionary computations in engineering systems, 2020, pp. 351–363.
[8] S. Samoili, M.L. Cobo, E. Gomez, G. De Prato, F. Martinez-Plumed, B. Delipetrev, A.I. Watch, Technical report, Joint Research Center (Seville site), 2020.
[9] High-Level Expert Group on Artificial Intelligence. (HLEG AI), A definition of AI: main capabilities and disciplines, (2019). Retrieved from Brussels https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341.
[10] D. Zhao, A. Strotmann, Analysis and visualization of citation networks, Synthesis lectures on information concepts, retrieval, and services, 7 1 (2015) 1–207.
[11] V.G. Promyslov, K.V. Semenkov, A.S. Shumov, A clustering method of asset cybersecurity classification, IFAC-PapersOnLine 52 (13) (2019) 928–933.
[12] K. Millar, A. Cheng, H.G. Chew, C.C. Lim, Operating system classification: a minimalist approach, in: 2020 International Conference on Machine Learning and Cybernetics (ICMLC), 2020, pp. 143–150.
[13] A. Aksoy, M.H. Gunes, Automated iot device identification using network traffic, in: IEEE International Conference on Communications (ICC), 2019, pp. 1–7.
[14] A. Sivanathan, H.H. Gharakheili, F. Loi, A. Radford, C. Wijenayake, A. Vishwanath, V. Sivaraman, Classifying IoT devices in smart environments using network traffic characteristics, IEEE Trans. Mobile Comput. 18 (8) (2018) 1745–1759.
[15] I. Cvitić, D. Peraković, M. Periša, B. Gupta, Ensemble machine learning approach for classification of IoT devices in smart home, Int. J. Machine Learn. Cybernetics 12 (11) (2021) 3179–3202.
[16] H. Cam, Online detection and control of malware infected assets, in: IEEE Military Communications Conference (MILCOM), 2017, pp. 701–706.
[17] H.I. Kure, S. Islam, M. Ghazanfar, A. Raza, M. Pasha, Asset criticality and risk prediction for an effective cybersecurity risk management of cyber-physical system, Neural Comput. App. 34 (1) (2022) 493–514.
[18] M. Vega-Barbas, V.A. Villagrá, F. Monje, R. Riesco, X. Larriva-Novo, J. Berrocal, Ontology-based system for dynamic risk management in administrative domains, Appl. Sci. 9 (21) (2019) 4547.
[19] B. Tozer, T. Mazzuchi, S. Sarkani, optimizing attack surface and configuration diversity using multi-objective reinforcement learning, in: IEEE 14th international conference on machine learning and applications, 2015, pp. 144–149.
[20] L.E. García-Hernández, A. Tchernykh, V. Miranda-López, M. Babenko, A. Avetisyan, R. Rivera-Rodriguez, G. Radchenko, C.J. Barrios-Hernandez, H. Castro, A.Y. Drozdov, Multi-objective configuration of a secured distributed cloud data storage, in: Latin American High Performance Computing Conference, 2019, pp. 78–93. Sep.
[21] M. Sharifi, F. Eugene, J.G. Carbonell, Learning of personalized security settings, in: IEEE International Conference on Systems, Man and Cybernetics, 2010, pp. 3428–3432.
[22] D. Bringhenti, G. Marchetto, R. Sisto, F. Valenza, F.J. Yusupov, Towards a fully automated and optimized network security functions orchestration, in: 4th International Conference on Computing, Communications and Security (ICCCS), 2019, pp. 1–7.
[23] Á.J. Varela-Vaca, R.M. Gasca, J.A. Carmona-Fombella, M.T. Gómez-López, AMADEUS: towards the AutoMAteD secUrity teSting, in: Proceedings of the 24th ACM Conference on Systems and Software Product Line, 2020, pp. 1–12.
[24] Á.J. Varela-Vaca, R.M. Gasca, R. Ceballos, M.T. Gómez-López, P.B. Torres, CyberSPL: a framework for the verification of cybersecurity policy compliance of system configurations using software product lines, Appl. Sci. 9 (24) (2019) 5364.
[25] Y. Liu, A. Sarabi, J. Zhang, P. Naghizadeh, M. Karir, M. Bailey, M. Liu, Cloudy with a chance of breach: forecasting cyber security incidents, in: 24th USENIX Security Symposium (USENIX Security 15), 2015, pp. 1009–1024.
[26] Z. Zhan, M. Xu, S. Xu, A characterization of cybersecurity posture from network telescope data, in: International Conference on Trusted Systems, 2014, pp. 105–126, 2014.
[27] S.N.G. Gourisetti, M. Mylrea, E. Gervais, S. Bhadra, Multi-scenario use case-based demonstration of buildings cybersecurity framework webtool, in: 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017, pp. 1–8.
[28] L.V. Stepanov, A.S. Koltsov, A.V. Parinov, Evaluating the cybersecurity of an enterprise based on a genetic algorithm, in: International Russian Automation Conference, 2020, pp. 580–590.
[29] V.L. Narasimhan, Using deep learning for assessing cybersecurity economic risks in virtual power plants, in: 2021 7th International Conference on Electrical Energy Systems (ICEES), 2021, pp. 530–537.
[30] H.H. Nguyen, D.M. Nicol, estimating loss due to cyber-attack in the presence of uncertainty, in: 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2020, pp. 361–369.
[31] C. Ponsard, V. Ramon, M. Touzani, Improving cyber security risk assessment by combined use of i* and Infrastructure Models, in: the 14th International iStar Workshop, 2021, pp. 63–69.
[32] O. Odegbile, S. Chen, Y. Wang, Dependable policy enforcement in traditional non-sdn networks, in: 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), 2019, pp. 545–554.
[33] F.D. Nembhard, M.M. Carvalho, T.C. Eskridge, Towards the application of recommender systems to secure coding, EURASIP J. Inf. Security 1 (2019) 1–24.
[34] S. Liu, G. Lin, Q.L. Han, S. Wen, J. Zhang, Y. Xiang, DeepBalance: deep-learning and fuzzy oversampling for vulnerability detection, IEEE Trans. Fuzzy Syst. 28 (7) (2019) 1329–1343.
[35] S. Jeon, H.K. Kim, AutoVAS: an automated vulnerability analysis system with a deep learning approach, Comput. Secur. 106 (2021), 102308.
[36] P. Huff, K. McClanahan, T. Le, Q. Li, A recommender system for tracking vulnerabilities, in: The 16th International Conference on Availability, Reliability and Security, 2021, pp. 1–7.
[37] D. Iorga, D.G. Corlatescu, O. Grigoresc, C. Sandescu, M. Dascalu, R. Rughinis, Yggdrasil—early detection of cybernetic vulnerabilities from Twitter, in: 23rd International Conference on Control Systems and Computer Science (CSCS), 2021, pp. 463–468.
[38] T. Saha, N. Aaraj, N. Ajjarapu, N.K. Jha, SHARKS: smart hacking approaches for RisK scanning in internet-of-things and cyber-physical systems based on machine learning, IEEE Trans. Emerg. 10 (2) (2021) 870–885.

[39] Y. Wang, Z. Wu, Q. Wei, Q. Wang, NeuFuzz: efficient fuzzing with deep neural network, IEEE Access, 7 36340–36352.
[40] J. Wang, B. Chen, L. Wei, Y. Liu, Skyfire: data-driven seed generation for fuzzing, in: 2017 IEEE Symposium on Security and Privacy (SP), 2017, pp. 579–594.
[41] P. Godefroid, H. Peleg, R. Singh, Learn&Fuzz: machine learning for input fuzzing, in: 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE), 2017, pp. 50–59.
[42] C. Cummins, P. Petoumenos, A. Murray, H. Leather, Compiler fuzzing through deep learning, in: Proceedings of the 27th ACM SIGSOFT International Symposium on Software Testing and Analysis, 2018, pp. 95–105.
[43] H. Xu, Y. Wang, S. Fan, P. Xie, A. Liu, DSmith: compiler fuzzing through generative deep learning model with attention, in: 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1–9.
[44] Y. Chen, C.M. Poskitt, J. Sun, S. Adepu, F. Zhang, Learning-guided network fuzzing for testing cyber-physical system defences, in: 34th IEEE/ACM International Conference on Automated Software Engineering (ASE), 2019, pp. 962–973. Nov.
[45] D. She, K. Pei, D. Epstein, J. Yang, B. Ray, S. Jana, NEUZZ: efficient fuzzing with neural program smoothing, in: IEEE Symposium on Security and Privacy (SP), 2019, pp. 803–817.
[46] X. Liu, X. Li, R. Prajapati, D. Wu, DeepFuzz: automatic generation of syntax valid c programs for fuzz testing, in: Proceedings of the AAAI Conference on Artificial Intelligence 33, 2019, pp. 1044–1051.
[47] S. Zhou, J. Liu, D. Hou, X. Zhong, Y. Zhang, Autonomous penetration testing based on improved deep Q-network, Appl. Sci. 11 (2021) 8823.
[48] R. Gangupantulu, T. Cody, A. Rahm, C. Redino, R. Clark, P. Park, Crown jewels analysis using reinforcement learning with attack graphs, in: IEEE Symposium Series on Computational Intelligence (SSCI), 2021, pp. 1–6.
[49] C. Neal, H. Dagdougu, A. Lodi, J.M. Fernandez, Reinforcement learning based penetration testing of a microgrid control algorithm, in: IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), 2021, pp. 0038–0044.
[50] E.R. Russo, A. Di Sorbo, C.A. Visaggio, G. Canfora, Summarizing vulnerabilities' descriptions to support experts during vulnerability assessment activities, J. Syst. Softw. 156 (2019) 84–99.
[51] M. Aota, H. Kanehara, M. Kubo, N. Murata, B. Sun, T. Takahashi, Automation of vulnerability classification from its description using machine learning, in: IEEE Symposium on Computers and Communications (ISCC), 2020, pp. 1–7.
[52] M. Vanamala, X. Yuan, K. Roy, Topic modeling and classification of common vulnerabilities and exposures database, in: International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), 2020, pp. 1–5.
[53] G. Bakirtzis, B.J. Simon, A.G. Collins, C.H. Fleming, C.R. Elks, Data-driven vulnerability exploration for design phase system analysis, IEEE Syst. J. 14 (2019) 4864–4873.
[54] A. Kuppa, L. Aouad, N.A. Le-Khac, Linking CVE's to MITRE ATT&CK techniques, in: 16th International Conference on Availability, Reliability and Security, 2021, pp. 1–12.
[55] S. Chatterjee, S. Thekdi, An iterative learning and inference approach to managing dynamic cyber vulnerabilities of complex systems, Reliab. Eng. Syst. 193 (2020), 106664.
[56] Y. Jiang, Y. Atif, A selective ensemble model for cognitive cybersecurity analysis, J. Netw. Comput. Appl. 193 (2021), 103210.
[57] S. Samtani, S. Yu, H. Zhu, M. Patton, J. Matherly, H. Chen, Identifying SCADA systems and their vulnerabilities on the internet of things: a text-mining approach, IEEE Intell. Syst. 33 (2) (2018) 63–73.
[58] J. Brown, T. Saha, N.K. Jha, GRAVITAS: graphical reticulated attack vectors for internet-of-things aggregate security, IEEE Trans. Emerg. 10 (3) (2022) 1331–1348.
[59] P. Gao, F. Shao, X. Liu, X. Xiao, Z. Qin, F. Xu, P. Mittal, S.R. Kulkarn, D. Song, Enabling efficient cyber threat hunting with cyber threat intelligence, in: IEEE 37th International Conference on Data Engineering (ICDE), 2021, pp. 193–204.
[60] A. Nadeem, S. Verwer, S. Moskal, S.J. Yang, Alert-driven attack graph generation using S-PDFA, IEEE Trans. Dependable and Secure Comput. 19 (2022) 731–746.
[61] H. Binyamini, R. Bitton, M. Inokuchi, T. Yagyu, Y. Elovici, A. Shabtai, A framework for modeling cyber attack techniques from security vulnerability descriptions, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 2574–2583.
[62] G. Falco, A. Viswanatha, C. Caldera, H. Shrobe, A master attack methodology for an AI-based automated attack planner for smart cities, IEEE Access 6 (2018) 48360–48373.
[63] H. Cam, Model-guided infection prediction and active defense using context-specific cybersecurity observations, in: MILCOM 2019-2019 IEEE Military Communications Conference (MILCOM), 2019, pp. 1–6.
[64] A. Wollaber, J. Peña, B. Bleas, L. Shing, K. Alperin, S. Vilvovsky, P. Trepagnier, N. Wagner, L. Leonard, Proactive cyber situation awareness via high performance computing, in: IEEE High Performance Extreme Computing Conference (HPEC), 2019, pp. 1–7.
[65] J.C. Sancho, A. Caro, M. Ávila, A. Bravo, New approach for threat classification and security risk estimations based on security event management, Future Gener. Comput. Syst. 113 (2020) 488–505.
[66] A.A. Tubis, S. Werbińska-Wojciechowsk, M. Góralczyk, A. Wróblewski, B. Ziętek, Cyber-attacks risk analysis method for different levels of automation of mining processes in mines based on fuzzy theory use, Sensors 20 (24) (2020) 7210.
[67] Y. Qin, Y. Peng, K. Huang, C. Zhou, Y.C. Tian, Association analysis-based cybersecurity risk assessment for industrial control systems, IEEE Syst. J. 15 (1) (2020) 1423–1432.
[68] G. Falco, C. Caldera, H. Shrobe, IIoT cybersecurity risk modeling for SCADA systems, IEEE Internet Things J 5 (6) (2018) 4486–4495.
[69] M. Vega-Barbas, V.A. Villagrá, F. Monje, R. Riesco, X. Larriva-Novo, J. Berrocal, Ontology-based system for dynamic risk management in administrative domains, Appl. Sci. 9 (21) (2019) 4547.
[70] M. Kalinin, V. Krundyshev, P. Zegzhda, Cybersecurity risk assessment in smart city infrastructures, Machines 9 (4) (2021) 78.
[71] B. Biswas, A. Mukhopadhyay, S. Bhattacharjee, A. Kumar, D. Delen, A text-mining based cyber-risk assessment and mitigation framework for critical analysis of online hacker forums, Decis. Support Syst. 152 (2022), 113651.
[72] N. Al-Hadhrami, M. Collinson, N. Oren, A subjective network approach for cybersecurity risk assessment, in: 13th International Conference on Security of Information and Networks, 2020, pp. 1–8.
[73] M.S. Ansari, V. Bartoš, B. Lee, GRU-based deep learning approach for network intrusion alert prediction, Future Gener. Comput. Syst. 128 (2022) 35–47.
[74] L. Wang, R. Jones, Big data analytics in cybersecurity: network data and intrusion prediction, in: IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2019, pp. 0105–0111.
[75] H. Al Najada, I. Mahgoub, I. Mohammed, Cyber intrusion prediction and taxonomy system using deep learning and distributed big data processing, in: IEEE symposium series on computational intelligence (SSCI), 2018, pp. 631–638.
[76] W.G. Mueller, A. Memory, K. Bartrem, Forecasting network intrusions from security logs using LSTMs, in: International Workshop on Deployable Machine Learning for Security Defense, 2020, pp. 122–137.
[77] M. Rhode, P. Burnap, K. Jones, Early-stage malware prediction using recurrent neural networks, Comput. Secur. 77 (2018) 578–594.
[78] I. Perera, J. Hwang, K. Bayas, B. Dorr, Y. Wilks, Cyberattack prediction through public text analysis and mini-theories, in: IEEE International Conference on Big Data (Big Data), 2018, pp. 3001–3010.
[79] E. Marin, M. Almukaynizi, P. Shakarian, Inductive and deductive reasoning to assist in cyber-attack prediction, in: 10th Annual Computing and Communication Workshop and Conference (CCWC), 2020, pp. 0262–0268.
[80] N. Polatidis, E. Pimenidis, M. Pavlidis, S. Papastergiou, H. Mouratidis, From product recommendation to cyber-attack prediction: generating attack graphs and predicting future attacks, Evolving Syst. 11 (3) (2020) 479–490.
[81] L.P. Rees, J.K. Deane, T.R. Rakes, W.H. Baker, Decision support for cybersecurity risk planning, Decis. Support Syst. 51 (3) (2011) 493–505.
[82] J.A. Paul, X.J. Wang, Socially optimal IT investment for cybersecurity, Decis. Support Syst. 122 (2019), 113069.
[83] J.A. Paul, M. Zhang, Decision support model for cybersecurity risk planning: a two-stage stochastic programming framework featuring firms, government, and attacker, Eur. J. Oper. Res. 291 (1) (2021) 349–364.
[84] K. Zheng, L.A. Albert, J.R. Luedtke, E. Towle, A budgeted maximum multiple coverage model for cybersecurity planning and management, IISE Trans. 51 (12) (2019) 1303–1317.
[85] A. Yeboah-Ofori, S. Islam, S.W. Lee, Z.U. Shamszaman, K. Muhammad, M. Altaf, M.S. Al-Rakhami, Cyber threat predictive analytics for improving cyber supply chain security, IEEE Access 9 (2021) 94318–94337.
[86] T. Sawik, A linear model for optimal cybersecurity investment in industry 4.0 supply chains, Int. J. Prod. Res. 60 (4) (2022) 1368–1385.
[87] T. Sawik, B. Sawik, A rough cut cybersecurity investment using portfolio of security controls with maximum cybersecurity value, Int. J. Prod. Res. 60 (21) (2022) 6556–6572.
[88] T. Sawik, Balancing cybersecurity in a supply chain under direct and indirect cyber risks, Int. J. Prod. Res. 60 (2) (2022) 766–782.
[89] S. Rahman, N.U. Hossain, K. Govindan, F. Nur, M. Bappy, assessing cyber resilience of additive manufacturing supply chain leveraging data fusion technique: a model to generate cyber resilience index of a supply chain, CIRP J. Manuf. Sci. Technol. 35 (2021) 911–928.
[90] A.I. Siam, A. Sedik, W. El-Shafai, A.A. Elazm, N.A. El-Bahnasawy, G.M. El Banby, A.A. Khalaf, F.E. Abd El-Samie, Biosignal classification for human identification based on convolutional neural networks, Int. J. Commun. Syst. 34 (7) (2021) 1–22.
[91] J.M. Jorquera Valero, P.M. Sánchez Sánchez, L. Fernández Maimó, A. Huertas Celdrán, M. Arjona Fernández, S. De Los Santos Vílchez, G. Martínez Pérez, Improving the security and QoE in mobile devices through an intelligent and adaptive continuous authentication system, Sensors 18 (11) (2018) 3769.
[92] P.M. Sánchez, A. Huertas Celdrán, L. Fernández Maimó, G. Martínez Pérez, G. Wang, Securing smart offices through an intelligent and multi-device continuous authentication system, in: International Conference on Smart City and Informatization, 2019, pp. 73–85.
[93] A.G. Martín, M. Beltrán, A. Fernández-Isabel, I.M. de Diego, An approach to detect user behaviour anomalies within identity federations, Comp. Security 108 (2021), 102356.
[94] H. Alobaidi, N. Clarke, F. Li, A. Alruban, Real-world smartphone-based gait recognition, Comput. Secur. 113 (2022), 102557.
[95] K.A. Rahman, D. Neupane, A. Zaiter, M.S. Hossain, Web user authentication using chosen word keystroke dynamics, in: 18th IEEE International Conference on Machine Learning and Applications (ICMLA), 2019, pp. 1130–1135.
[96] A. Shaout, N. Schmidt, Keystroke identifier using fuzzy logic to increase password security, in: 21st International Arab Conference on Information Technology (ACIT), 2020, pp. 1–8.

[97] A. Hafeez, K. Topolovec, S. Awad, ECU fingerprinting through parametric signal modeling and artificial neural networks for in-vehicle security against spoofing attacks, in: 15th International Computer Engineering Conference (ICENCO), 2019, pp. 29–38.

[98] G. Baldini, R. Giuliani, M. Gemo, F. Dimc, On the application of sensor authentication with intrinsic physical features to vehicle security, Comput. Electr. Eng. 91 (2021), 107053.

[99] Y. Cui, F. Bai, R. Yan, T. Saha, R.K. Ko, Y. Liu, Source Authentication of distribution synchrophasors for cybersecurity of microgrids, IEEE Trans. Smart Grid 12 (5) (2021) 4577–4580.

[100] M. Benedetti, M. Mori, On the use of Max-SAT and PDDL in RBAC maintenance, Cybersecurity 2 (1) (2019) 1–25.

[101] M. Abolfathi, Z. Raghebi, H. Jafarian, F. Banaei-Kashani, A scalable role mining approach for large organizations, in: Proceedings of the 2021 ACM Workshop on Security and Privacy Analytics, 2021, pp. 45–54.

[102] S.S. Chukkapalli, S.B. Aziz, N. Alotaibi, S. Mittal, M. Gupta, M. Abdelsalam, Ontology driven AI and access control systems for smart fisheries, in: Proceedings of the 2021 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems, 2021, pp. 59–68.

[103] B. Leander, A. Čaušević, H. Hansson, T. Lindström, Access control for smart manufacturing systems, in: European Conference on Software Architecture, 2020, pp. 463–476.

[104] Z. Tan, R. Beuran, S. Hasegawa, W. Jiang, M. Zhao, Y. Tan, Adaptive security awareness training using linked open data datasets, Educ. Inf. Technol. 25 (6) (2020) 5235–5259.

[105] F. Nembhard, M. Carvalho, T. Eskridge, A hybrid approach to improving program security, in: IEEE Symposium Series on Computational Intelligence (SSCI), 2017, pp. 1–8.

[106] T.Espinha Gasiba, U. Lechner, M. Pinto-Albuquerque, Sifu-a cybersecurity awareness platform with challenge assessment and intelligent coach, Cybersecurity 3 (1) (2020) 1–23.

[107] D.C. Le, N. Zincir-Heywood, Anomaly detection for insider threats using unsupervised ensembles, IEEE Trans. Netw. Service Manag. 18 (2) (2021) 1152–1164.

[108] J. Kim, M. Park, H. Kim, S. Cho, P. Kang, Insider threat detection based on user behavior modeling and anomaly detection algorithms, Appl. Sci. 9 (19) (2019) 4018.

[109] T. Al-Shehari, R.A. Alsowail, An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques, Entropy 23 (10) (2021) 1258.

[110] K. Alzhrani, E.M. Rudd, T.E. Boult, C.E. Chow, Automated big text security classification, in: IEEE Conference on Intelligence and Security Informatics (ISI), 2016, pp. 103–108.

[111] Y. Guo, J. Liu, W. Tang, C. Huang, Exsense: extract sensitive information from unstructured data, Comput. Secur. 102 (2021), 102156.

[112] H. Li, J. Wu, H. Xu, G. Li, M. Guizani, Explainable intelligence-driven defense mechanism against advanced persistent threats: a joint edge game and AI approach, IEEE Trans. Dependable Secure Comput. 19 (2) (2022) 757–775.

[113] A.A. Alghamdi, G. Reger, Pattern extraction for behaviours of multi-stage threats via unsupervised learning, in: International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), 2020, pp. 1–8.

[114] L. Gallo, A. Maiello, A. Botta, G. Ventre, 2 Years in the anti-phishing group of a large company, Comput. Secur. 105 (2021), 102259.

[115] D. Wu, W. Shi, X. Ma, A novel real-time anti-spam framework, ACM Trans. Internet Technol. (TOIT) 21 (4) (2021) 1–27.

[116] E.S. Gualberto, R.T. De Sousa, T.P. Vieira, J.P. Da Costa, C.G. Duque, The answer is in the text: multi-stage methods for phishing detection based on feature engineering, IEEE Access 8 (2020) 223529–223547.

[117] M. Nguyen, T. Nguyen, T.H. Nguyen, A deep learning model with hierarchical lstms and supervised attention for anti-phishing, in: 1st Anti-Phishing Shared Task Pilot at 4th ACM IWSPA, 2018, pp. 29–38.

[118] D. Cohen, O. Naim, E. Toch, I. Ben-Gal, Website categorization via design attribute learning, Comput. Secur. 107 (2021), 102312.

[119] C. Marques, S. Malta, J.P. Magalhães, DNS dataset for malicious domains detection, Data Br 38 (2021), 107342.

[120] B. Yu, J. Pan, D. Gray, J. Hu, C. Choudhary, A.C. Nascimento, M. De Cock, Weakly supervised deep learning for the detection of domain generation algorithms, IEEE Access 7 (2019) 51542–51556.

[121] J. Spaulding, A. Mohaisen, Defending internet of things against malicious domain names using D-FENS, in: IEEE/ACM Symposium on Edge Computing (SEC), 2018, pp. 387–392.

[122] P.L. Indrasiri, M.N. Halgamuge, A. Mohammad, Robust ensemble machine learning model for filtering phishing URLs: expandable random gradient stacked voting classifier (ERG-SVC), IEEE Access 9 (2021) 150142–150161.

[123] R. Vinayakumar, K.P. Soman, P. Poornachandran, evaluating deep learning approaches to characterize and classify malicious URL's, J. Intell. Fuzzy Syst. 34 (3) (2018) 1333–1343.

[124] W. Li, J. Jin, J.H. Lee, Analysis of botnet domain names for IoT cybersecurity, IEEE Access 7 (2019) 94658–94665.

[125] B. Alotaibi, M. Alotaibi, Consensus and majority vote feature selection methods and a detection technique for web phishing, J. Ambient. Intell. Humaniz. Comput. 12 (1) (2021) 717–727.

[126] Y. Qin, B. Hoffmann, D.J. Lilja, Hyperprotect: enhancing the performance of a dynamic backup system using intelligent scheduling, in: IEEE 37th International Performance Computing and Communications Conference (IPCCC), 2018, pp. 1–8.

[127] P.M. Van de Ven, B. Zhang, A. Schörgendorfer, Distributed backup scheduling: modeling and optimization, in: IEEE INFOCOM 2014-IEEE Conference on Computer Communications, 2014, pp. 1644–1652.

[128] Z. Zeng, Z. Yang, D. Huang, C.J. Chung, LICALITY–Likelihood and criticality: vulnerability risk prioritization through logical reasoning and deep learning, IEEE Trans. Netw. Service Manag. 19 (2) (2021) 1746–1760.

[129] J. Yin, M. Tang, J. Cao, H. Wang, Apply transfer learning to cybersecurity: predicting exploitability of vulnerabilities by description, Knowl. Based Syst. 210 (2020), 106529.

[130] J. Yin, M. Tang, J. Cao, H. Wang, M. You, A real-time dynamic concept adaptive learning algorithm for exploitability prediction, Neurocomputing 472 (2022) 252–265.

[131] T. Bai, H. Bian, M.A. Salahuddin, A. Abou Daya, N. Limam, R. Boutaba, Rdp-based lateral movement detection using machine learning, Comp. Commun. 165 (2021) 9–19.

[132] N. Afzaliseresht, Y. Miao, S. Michalska, Q. Liu, H. Wang, From logs to stories: human-centred data mining for cyber threat intelligence, IEEE Access 8 (2020) 19089–19099.

[133] G. De la Torre-Abaitua, L.F. Lago-Fernández, D. Arroyo, A compression-based method for detecting anomalies in textual data, Entropy 23 (5) (2021) 618.

[134] T. Eljasik-Swoboda, W. Demuth, Leveraging clustering and natural language processing to overcome variety issues in log management, in: ICAART, 2020, pp. 281–288.

[135] D. Sisiaridis, O. Markowitch, Reducing data complexity in feature extraction and feature selection for big data security analytics, in: 1st International Conference on Data Intelligence and Security (ICDIS), 2018, pp. 43–48.

[136] P.F. De Araujo-Filho, A.J. Pinheiro, G. Kaddoum, D.R. Campelo, F.L. Soares, An efficient intrusion prevention system for CAN: hindering cyber-attacks with a low-cost platform, IEEE Access 9 (2021) 166855–166869.

[137] C. Constantinides, S. Shiaeles, B. Ghita, N. Kolokotronis, A novel online incremental learning intrusion prevention system, in: 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2019, pp. 1–6.

[138] S.M. de Lima, H.K. Silva, J.H. Luz, H.J. Lima, S.L. Silva, A. de Andrade, A.M. da Silva, Artificial intelligence-based antivirus in order to detect malware preventively, Prog. Artif. Intell. 10 (1) (2021) 1–22.

[139] P. Marques, M. Rhode, I. Gashi, Waste not: using diverse neural networks from hyperparameter search for improved malware detection, Comput. Secur. 108 (2021), 102339.

[140] P. Karuna, H. Purohit, S. Jajodia, R. Ganesan, O. Uzuner, Fake document generation for cyber deception by manipulating text comprehensibility, IEEE Syst. J. 15 (1) (2020) 835–845.

[141] O. Ajayi, A. Gangopadhyay, DAHID: domain adaptive host-based intrusion detection, in: IEEE International Conference on Cyber Security and Resilience (CSR), 2021, pp. 467–472.

[142] G. Granato, A. Martino, L. Baldini, A. Rizzi, Intrusion detection in Wi-Fi networks by modular and optimized ensemble of classifiers, in: IJCCI, 2020, pp. 412–422.

[143] Z. Li, A.L. Rios, L. Trajković, Machine learning for detecting anomalies and intrusions in communication networks, IEEE J. Sel. Areas Commun. 39 (7) (2021) 2254–2264.

[144] M. Almiani, A. AbuGhazleh, Y. Jararweh, A. Razaque, DDoS detection in 5G-enabled IoT networks using deep Kalman backpropagation neural network, Int. J. Mach. Learn. Cybern. 12 (11) (2021) 3337–3349.

[145] A. Corsini, S.J. Yang, G. Apruzzese, On the evaluation of sequential machine learning for network intrusion detection, in: 16th International Conference on Availability, Reliability and Security, 2021, pp. 1–10.

[146] M. Choraś, M. Pawlicki, Intrusion detection approach based on optimised artificial neural network, Neurocomputing 452 (2021) 705–715.

[147] K.S. Kumar, S.A. Nair, D.G. Roy, B. Rajalingam, R.S. Kumar, Security and privacy-aware artificial intrusion detection system using federated machine learning, Comput. Electr. Eng. (96)107440.

[148] J.C. Wu, S. Lu, C.S. Fuh, T.L. Liu, One-class anomaly detection via novelty normalization, Comput. Vis. Image. Underst. 210 (2021), 103226.

[149] L. Fernández Maimó, A. Huertas Celdrán, M. Gil Pérez, F.J. García Clemente, G. Martínez Pérez, Dynamic management of a deep learning-based anomaly detection system for 5G networks, J. Ambient. Intell. Humaniz. Comput. 10 (8) (2019) 3083–3097.

[150] D.C. Le, A.N. Zincir-Heywood, M.I. Heywood, Data analytics on network traffic flows for botnet behaviour detection, in: IEEE symposium series on computational intelligence (SSCI), 2016, pp. 1–7.

[151] D. Saveetha, G. Maragatham, Design of Blockchain enabled intrusion detection model for detecting security attacks using deep learning, Pattern Recognit. Lett. 153 (2022) 24–28.

[152] M. Al-Hawawreh, E. Sitnikova, F. den Hartog, An efficient intrusion detection model for edge system in brownfield industrial Internet of Things, in: Proceedings of the 3rd International Conference on Big Data and Internet of Things, 2019, pp. 83–87.

[153] J. Vávra, M. Hromada, L. Lukáš, J. Dworzecki, Adaptive anomaly detection system based on machine learning algorithms in an industrial control environment, Int. J. Crit. Infrastruct. 34 (2021), 100446.

[154] Y. Zhang, L. Wang, W. Sun, R.C. Green II, M. Alam, Distributed intrusion detection system in a multi-layer network architecture of smart grids, IEEE Trans. on Smart Grid 2 (4) (2011) 796–808.

[155] R. Blanco, P. Malagón, S. Briongos, J.M. Moya, Anomaly detection using Gaussian mixture probability model to implement intrusion detection system, in: International Conference on Hybrid Artificial Intelligence Systems, 2019, pp. 648–659.

[156] G.N. Nguyen, N.H. Le, M. Viet, K. Elhoseny, B.B. Shankar, A.A. Gupta, Abd El-Latif, Secure blockchain enabled Cyber–physical systems in healthcare using deep belief network with ResNet model, J. Parallel. Distrib. Comput. 153 (2021) 150–160.

[157] A. Alhowaide, I. Alsmadi, J. Tang, Ensemble detection model for IoT IDS, Internet of Things 16 (2021), 100435.

[158] A. Binbusayyis, T. Vaiyapuri, Unsupervised deep learning approach for network intrusion detection combining convolutional autoencoder and one-class SVM, Appl. Intell. 51 (10) (2021) 7094–7108.

[159] V. Herrera-Semenets, L. Bustio-Martínez, R. Hernández-León, J. van den Berg, A multi-measure feature selection algorithm for efficacious intrusion detection, Knowl. Based Syst. 227 (2021), 107264.

[160] M. Rashid, J. Kamruzzaman, T. Imam, S. Wibowo, S. Gordon, A tree-based stacking ensemble technique with feature selection for network intrusion detection, Appl. Intell. 52 (2022) 9768–9781.

[161] V. Dutta, M. Choraś, R. Kozik, M. Pawlicki, Hybrid model for improving the classification effectiveness of network intrusion detection, in: Computational Intelligence in Security for Information Systems Conference, 2019, pp. 405–414.

[162] S.I. Pérez, S. Moral-Rubio, R. Criado, A new approach to combine multiplex networks and time series attributes: building intrusion detection systems (IDS) in cybersecurity, Chaos, Solit. Fractals 150 (2021), 111143.

[163] P. Singh, A. Pankaj, R. Mitra, Edge-detect: edge-centric network intrusion detection using deep neural network, in: IEEE 18th Annual Consumer Communications & Networking Conference, 2021, pp. 1–6.

[164] M. Catillo, A. Pecchia, U. Villano, AutoLog: anomaly detection by deep autoencoding of system logs, Expert Syst. Appl. 191 (2022), 116263.

[165] R. Zhao, Y. Yin, Y. Shi, Z. Xue, Intelligent intrusion detection based on federated learning aided long short-term memory, Phys. Commun. 42 (2020), 101157.

[166] D. Nedeljkovic, Z. Jakovljevic, CNN based method for the development of cyber-attacks detection algorithms in industrial control systems, Comput. Secur. 114 (2022), 102585.

[167] M. Elnour, N. Meskin, K.M. Khan, Hybrid attack detection framework for industrial control systems using 1d-convolutional neural network and isolation forest, in: IEEE Conference on Control Technology and Applications (CCTA), 2020, pp. 877–884.

[168] H. Liu, C. Zhong, A. Alnusair, S.R. Islam, FAIXID: a framework for enhancing ai explainability of intrusion detection results using data cleaning techniques, J. Netw. Syst. Manag. 29 (4) (2021) 1–30.

[169] J.M. Vidal, M.A. Monge, S.M. Monterrubio, EsPADA: enhanced payload analyzer for malware detection robust against adversarial threats, Future Gener. Comput. Syst. 104 (2020) 159–173.

[170] S. Latif, Z.E. Huma, S.S. Jamal, F. Ahmed, J. Ahmad, A. Zahid, K. Dashtipour, M. U. Aftab, M. Ahmad, Q.H. Abbasi, Intrusion detection framework for the internet of things using a dense random neural network, IEEE Trans. Industr. Inform. 18 (9) (2021) 6435–6444.

[171] J.L. Leevy, J. Hancock, R. Zuech, T.M. Khoshgoftaar, Detecting cybersecurity attacks using different network features with lightgbm and xgboost learners, in: IEEE Second International Conference on Cognitive Machine Intelligence (CogMI), 2020, pp. 190–197.

[172] R. Abdulhammed, M. Faezipour, A. Abuzneid, A. Alessa, Enhancing wireless intrusion detection using machine learning classification with reduced attribute sets, in: 14th International Wireless Communications & Mobile Computing Conference (IWCMC), 2018, pp. 524–529.

[173] C. Iwendi, S.U. Rehman, A.R. Javed, S. Khan, G. Srivastava, Sustainable security for the internet of things using artificial intelligence architectures, ACM Trans. Internet Technol. 21 (3) (2021) 1–22.

[174] P. Toupas, D. Chamou, K.M. Giannoutakis, A. Drosou, D. Tzovaras, An intrusion detection system for multi-class classification based on deep neural networks, in: 18th IEEE International Conference on Machine Learning and Applications (ICMLA), 2019, pp. 1253–1258.

[175] L. D' hooge, M. Verkerken, T. Wauters, B. Volckaert, F. De Turck, Hierarchical feature block ranking for data-efficient intrusion detection modeling, Comput. Netw. 201 (2021), 108613.

[176] S. Huang, K. Lei, IGAN-IDS: an imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks, Ad Hoc Netw 105 (2020), 102177.

[177] N. Gupta, V. Jindal, P. Bedi, CSE-IDS: using cost-sensitive deep learning and ensemble algorithms to handle class imbalance in network-based intrusion detection systems, Comput. Secur. 112 (2022), 102499.

[178] S.S. Jagtap, S.S. VS, V. Subramaniyaswamy, A hypergraph based Kohonen map for detecting intrusions over cyber–physical systems traffic, Future Gener. Comput. Syst. 119 (2021) 84–109.

[179] M. Asif, S. Abbas, M.A. Khan, A. Ftima, M.A. Khan, S.W. Lee, MapReduce based intelligent model for intrusion detection using machine learning technique, J. King Saud Univ. - Comput. Inf. Sci. 34 (2022) 9723–9731.

[180] J. Liu, B. Kantarci, C. Adams, Machine learning-driven intrusion detection for Contiki-NG-based IoT networks exposed to NSL-KDD dataset, in: Proceedings of the 2nd ACM workshop on wireless security and machine learning, 2020, pp. 25–30.

[181] R. Blanco, P. Malagón, J.J. Cilla, J.M. Moya, Multiclass network attack classifier using CNN tuned with genetic algorithms, in: 28th International Symposium on Power and Timing Modeling, Optimization and Simulation (PATMOS), 2018, pp. 177–182.

[182] M. Pawlicki, R. Kozik, M. Choraś, Artificial neural network hyperparameter optimisation for network intrusion detection, in: International Conference on Intelligent Computing, 2019, pp. 749–760.

[183] M. Shafiq, Z. Tian, A.K. Bashir, X. Du, M. Guizani, IoT malicious traffic identification using wrapper-based feature selection mechanisms, Comput. Secur. 94 (2020), 101863.

[184] J.W. Mikhail, J.M. Fossaceca, R. Iammartino, A semi-boosted nested model with sensitivity-based weighted binarization for multi-domain network intrusion detection, ACM Trans. Intell. Syst. Technol. 10 (3) (2019) 1–27.

[185] M. Basnet, M.H. Ali, Deep learning-based intrusion detection system for electric vehicle charging station, in: 2nd International Conference on Smart Power & Internet Energy Systems (SPIES), 2020, pp. 408–413.

[186] A.F. Diallo, P. Patras, Adaptive clustering-based malicious traffic classification at the network edge, in: IEEE INFOCOM 2021-IEEE Conference on Computer Communications, 2021, pp. 1–10.

[187] N. Gupta, V. Jindal, P. Bedi, LIO-IDS: handling class imbalance using LSTM and improved one-vs-one technique in intrusion detection system, Comput. Netw. 192 (2021), 108076.

[188] I. Ullah, Q.H. Mahmoud, A hybrid model for anomaly-based intrusion detection in SCADA networks, in: IEEE International Conference on Big Data (Big Data), 2017, pp. 2160–2167.

[189] G. Li, Y. Shen, P. Zhao, X. Lu, J. Liu, Y. Liu, S.C. Hoi, Detecting cyberattacks in industrial control systems using online learning algorithms, Neurocomputing 364 (2019) 338–348.

[190] J. Zhang, F. Li, F. Ye, An ensemble-based network intrusion detection scheme with Bayesian deep learning, in: IEEE International Conference on Communications (ICC), 2020, pp. 1–6.

[191] W. Zong, Y.W. Chow, W. Susilo, Interactive three-dimensional visualization of network intrusion detection data for machine learning, Future Gener. Comput. Syst. 102 (2020) 292–306.

[192] C. Ieracitano, A. Adeel, F.C. Morabito, A. Hussain, A novel statistical analysis and autoencoder driven intelligent intrusion detection approach, Neurocomputing 387 (2020) 51–62.

[193] Q. Liu, D. Wang, Y. Jia, S. Luo, C. Wang, A multi-task based deep learning approach for intrusion detection, Knowl. Based Syst. 238 (2022), 107852.

[194] D. Xuan, H. Hu, B. Wang, B. Liu, Intrusion detection system based on RF-SVM model optimized with feature selection, in: International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), 2021, pp. 1–5.

[195] P. Radoglou-Grammatikis, P. Sarigiannidis, E. Iturbe, E. Rios, S. Martinez, A. Sarigiannidis, G. Eftathopoulos, Y. Spyridis, A. Sesis, N. Vakakis, D. Tzovaras, Spear siem: a security information and event management system for the smart grid, Comput. Netw. 193 (2021), 108008.

[196] A. Fausto, G.B. Gaggero, F. Patrone, P. Girdinio, M. Marchese, Toward the integration of cyber and physical security monitoring systems for critical infrastructures, Sensors 21 (21) (2021) 6970.

[197] H.A. Kodituwakku, A. Keller, J. Gregor, InSight2: a modular visual analysis platform for network situational awareness in large-scale networks, Electronics (Basel) 9 (10) (2020) 1747.

[198] Y. Nikoloudakis, I. Kefaloukos, S. Klados, S. Panagiotakis, E. Pallis, C. Skianis, E. K. Markakis, Towards a machine learning based situational awareness framework for cybersecurity: an SDN implementation, Sensors 21 (14) (2021) 4939.

[199] F. Zhang, H.A. Kodituwakku, J.W. Hines, J. Coble, Multilayer data-driven cyber-attack detection system for industrial control systems based on network, system, and process data, IEEE Trans. Industr. Inform. 15 (7) (2019) 4362–4369.

[200] D.L. Marino, C.S. Wickramasinghe, B. Tsouvalas, C. Rieger, M. Manic, Data-driven correlation of cyber and physical anomalies for holistic system health monitoring, IEEE Access 9 (2021) 163138–163150.

[201] K. Al-Rowaily, M. Abulaish, N.A. Haldar, M. Al-Rubaian, BiSAL–A bilingual sentiment analysis lexicon to analyze Dark Web forums for cyber security, Digital Investig 14 (2015) 53–62.

[202] A. Deb, K. Lerman, E. Ferrara, Predicting cyber-events by leveraging hacker sentiment, Information 9 (11) (2018) 280.

[203] S. Ishikawa, S. Ozawa, T. Ban, Port-piece embedding for darknet traffic features and clustering of scan attacks, in: International Conference on Neural Information Processing, 2020, pp. 593–603.

[204] G. Pantelis, P. Petrou, S. Karagiorgou, D. Alexandrou, On strengthening SMEs and MEs threat intelligence and awareness by identifying data breaches, stolen credentials and illegal activities on the dark web, in: 16th International Conference on Availability, Reliability and Security, 2021, pp. 1–7.

[205] M. Schäfer, M. Fuchs, M. Strohmeier, M. Engel, M. Liechti, V. Lenders, BlackWidow: monitoring the dark web for cyber security information, in: 11th International Conference on Cyber Conflict (CyCon), 2019, pp. 1–21.

[206] Z. Fang, X. Zhao, Q. Wei, G. Chen, Y. Zhang, C. Xing, W. Li, H. Chen, Exploring key hackers and cybersecurity threats in Chinese hacker communities, in: IEEE conference on intelligence and security informatics (ISI), 2016, pp. 13–18.

[207] S.Y. Huang, T. Ban, A topic-based unsupervised learning approach for online underground market exploration, in: 18th IEEE international conference on trust, security and privacy in computing and communications/13th IEEE international conference on big data science and engineering (TrustCom/BigDataSE), 2019, pp. 208–215.

[208] G. Kim, C. Lee, J. Jo, H. Lim, Automatic extraction of named entities of cyber threats using a deep Bi-LSTM-CRF network, Int. J. Mach. Learn. Cybern. 11 (10) (2020) 2341–2355.

[209] I. Sarhan, M. Spruit, Open-cykg: an open cyber threat intelligence knowledge graph, Knowl. Based Syst. 233 (2021), 107524.

[210] F. Alves, A. Bettini, P.M. Ferreira, A. Bessani, Processing tweets for cybersecurity threat awareness, Inf. Syst. 95 (2021), 101586.

[211] N. Dionísio, F. Alves, P.M. Ferreira, A. Bessani, Cyberthreat detection from twitter using deep neural networks, in: International Joint Conference on Neural Networks (IJCNN), 2019, pp. 1–8.
[212] J.R. Saura, D. Palacios-Marqués, D. Ribeiro-Soriano, Using data mining techniques to explore security issues in smart living environments in Twitter, Comput. Commun. 179 (2021) 285–295.
[213] T.M. Georgescu, B. Iancu, M. Zurini, Named-entity-recognition-based automated system for diagnosing cybersecurity situations in IoT networks, Sensors 19 (15) (2019) 3380.
[214] J. Sleeman, T. Finin, M. Halem, Understanding cybersecurity threat trends through dynamic topic modeling, front, Big Data 4 (2021).
[215] T. Sun, P. Yang, M. Li, S. Liao, An automatic generation approach of the cyber threat intelligence records based on multi-source information fusion, Future Internet 13 (2) (2021) 40.
[216] A. Sapienza, S.K. Ernala, A. Bessi, K. Lerman, E. Ferrara, Discover: mining online chatter for emerging cyber threats, in: Companion Proceedings of the The Web Conference, 2018, pp. 983–990.
[217] C.E. Tsai, C.L. Yang, C.K. Chen, A.N.T. CTI, Hunting for Chinese threat intelligence, in: IEEE International Conference on Big Data (Big Data), 2020, pp. 1847–1852.
[218] P. Ranade, S. Mittal, A. Joshi, K. Joshi, Using deep neural networks to translate multi-lingual threat intelligence, in: IEEE International Conference on Intelligence and Security Informatics (ISI), 2018, pp. 238–243.
[219] V.A. Memos, K.E. Psannis, AI-powered honeypots for enhanced IoT botnet detection, in: 3rd World Symposium on Communication Engineering (WSCE), 2020, pp. 64–68.
[220] P. Chatziadam, I.G. Askoxylakis, A. Fragkiadakis, A network telescope for early warning intrusion detection, in: International Conference on Human Aspects of Information Security, Privacy, and Trust, 2014, pp. 11–22.
[221] H.K. Kim, K.H. Im, S.C. Park, DSS for computer security incident response applying CBR and collaborative response, Expert Syst. Appl. 37 (1) (2010) 852–870.
[222] F. Jiang, T. Gu, L. Chang, Z. Xu, Case retrieval for network security emergency response based on description logic, in: International Conference on Intelligent Information Processing, 2014, pp. 284–293.
[223] R.C. Nunes, M. Colomé, F.A. Barcelos, M. Garbin, G.B. Paulus, L.A. Silva, A case-based reasoning approach for the cybersecurity incident recording and resolution, Int. J. Softw. Eng. Knowl. Eng. 11 (12) (2019) 1607–1627.
[224] I. Kraeva, G. Yakhyaeva, Application of the metric learning for security incident playbook recommendation, in: IEEE 22nd International Conference of Young Professionals in Electron Devices and Materials (EDM), 2021, pp. 475–479.
[225] L. Ping, Y. Haifeng, M. Guoqing, An incident response decision support system based on CBR and ontology, in: International conference on computer application and system modeling (ICCASM 2010), 2010, pp. 311–337.
[226] A. Shah, R. Ganesan, S. Jajodia, H. Cam, Dynamic optimization of the level of operational effectiveness of a CSOC under adverse conditions, ACM Trans. Intell. Syst. Technol. 9 (5) (2018) 1–20.
[227] Y. Lin, H. Wang, B. Yang, M. Liu, Y. Li, Y. Zhang, A blackboard sharing mechanism for community cyber threat intelligence based on multi-agent system, in: International Conference on Machine Learning for Cyber Security, 2019, pp. 253–270.
[228] L. Thomas, A. Vaughan, Z. Courtney, C. Zhong, A. Alnusair, Supporting collaboration among cyber security analysts through visualizing their analytical reasoning processes, in: IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2018, pp. 1–6.
[229] N. DeCastro-García, Á.L. Muñoz Castañeda, M. Fernández-Rodríguez, Machine learning for automatic assignment of the severity of cybersecurity events, Comput. Math. Methods Med. 2 (1) (2020) e1072.
[230] M. Husák, T. Bajtoš, J. Kašpar, E. Bou-Harb, P. Čeleda, Predictive cyber situational awareness and personalized blacklisting: a sequential rule mining approach, ACM Trans. Manag. Inf. Syst. 11 (4) (2020) 1–6.
[231] F. Manganiello, M. Marchetti, M. Colajanni, Multistep attack detection and alert correlation in intrusion detection systems, in: International Conference on Information Security and Assurance, 2011, pp. 101–110.
[232] A. Dey, E. Totel, S. Navers, Heterogeneous security events prioritization using auto-encoders, in: International Conference on Risks and Security of Internet and Systems, 2020, pp. 164–180.
[233] J.Q. Chen, Intelligent targeting with contextual binding, in: Future Technologies Conference (FTC), 2016, pp. 1040–1046.
[234] H. Studiawan, F. Sohel, Anomaly detection in a forensic timeline with deep autoencoders, J. Inf. Secur. Appl. 63 (2021), 103002.
[235] F. Amato, A. Castiglione, G. Cozzolino, F. Narducci, A semantic-based methodology for digital forensics analysis, J. Parallel Distrib. Comput. 138 (2020) 172–177.
[236] A. Nisioti, G. Loukas, A. Laszka, E. Panaousis, Data-driven decision support for optimizing cyber forensic investigations, IEEE Trans. Inf. Forensics Secur. 16 (2021) 2397–2412.
[237] J. Sakhnini, H. Karimipour, A. Dehghantanha, R.M. Parizi, Physical layer attack identification and localization in cyber–physical grid: an ensemble deep learning based approach, Phys. Commun. 47 (2021), 101394.
[238] L. Fernandez Maimo, A. Huertas Celdran, A.L. Perales Gomez, F.J. Garcia Clemente, J. Weimer, I. Lee, Intelligent and dynamic ransomware spread detection and mitigation in integrated clinical environments, Sensors 19 (5) (2019) 1114.
[239] P. Nespoli, F.G. Mármol, J.M. Vidal, A bio-inspired reaction against cyberattacks: ais-powered optimal countermeasures selection, IEEE Access 9 (2021) 60971–60996.
[240] M. Husák, L. Sadlek, S. Špaček, M. Laštovička, M. Javorník, J. Komárková, CRUSOE: a toolset for cyber situational awareness and decision support in incident handling, Comput. Secur. 115 (2022), 102609.
[241] M. Husák, Towards a data-driven recommender system for handling ransomware and similar incidents, in: IEEE International Conference on Intelligence and Security Informatics (ISI), 2021, pp. 1–6.
[242] A. Piplai, S. Mittal, A. Joshi, T. Finin, J. Holt, R. Zak, Creating cybersecurity knowledge graphs from malware after action reports, IEEE Access 8 (2020) 211691–211703.
[243] B. Woods, S.J. Perl, B. Lindauer, Data mining for efficient collaborative information discovery, in: Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security, 2015, pp. 3–12.
[244] S. Peng, A. Zhou, S. Liao, L. Liu, A threat actions extraction method based on the conditional co-occurrence degree, in: 7th International Conference on Information Science and Control Engineering (ICISCE), 2020, pp. 1633–1637.
[245] B.S. Meyers, A. Meneely, An automated post-mortem analysis of vulnerability relationships using natural language word embeddings, Procedia. Comput. Sci. (2021) 953–958.
[246] M.V. Carriegos, Á.L. Castañeda, M.T. Trobajo, D.A. De Zaballa, On aggregation and prediction of cybersecurity incident reports, IEEE Access 9 (2021) 102636–102648.
[247] Symantec, internet security threat report, Retrieved from https://www.symantec.com/content/dam/symantec/docs/reports/istr-24-2019-en.pdf (2019).