

Artificial Intelligence in Cybersecurity: A Comprehensive Review and Future Direction

Lizzy Ofusori ^a, Tebogo Bokaba ^b, and Siyabonga Mhlongo ^b

^aCentre for Applied Data Science, University of Johannesburg, Gauteng, South Africa; ^bDepartment of Applied Information Systems, University of Johannesburg, Gauteng, South Africa

ABSTRACT



As cybercrimes are becoming increasingly complex, it is imperative for cybersecurity measures to become more robust and sophisticated. The crux lies in extracting patterns or insights from cybersecurity data to build data-driven models, thus making the security systems automated and intelligent. To comprehend and analyze cybersecurity data, several Artificial Intelligence (AI) methods such as Machine Learning (ML) techniques, are employed to monitor network environments and actively combat cyber threats. This study explored the various AI techniques and how they are applied in cybersecurity. A comprehensive literature review was conducted, including a bibliometric analysis and systematic literature review following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Using data extracted from two main scholarly databases: Clarivate's Web of Science (WoS) and Scopus, this article examines relevant academic literature to understand the diverse ways in which AI techniques are employed to strengthen cybersecurity measures. These applications range from anomaly detection and threat identification to predictive analytics and automated incident response. A total of 14,509 peer-reviewed research papers were identified of which 9611 were from the Scopus database and 4898 from the WoS database. These research papers were further filtered, and a total of 939 relevant papers were eventually selected and used. The review offers insights into the effectiveness, challenges, and emerging trends in utilizing AI for cybersecurity purposes.

ARTICLE HISTORY

Received 29 August 2024
Revised 08 November 2024
Accepted 27 November 2024

Introduction

The rapid advancement in smart technologies, the Internet of Things (IoT) and other computing devices has generated enormous amounts of data that require robust security measures (Sarker et al. 2020). While this advancement has made human life and business practices easier, it has brought about a wave of security breaches (Künzler 2023). Studies have shown that due to the increasing dependency on digitalization, many security incidents, such as

CONTACT Lizzy Ofusori  lofusori@uj.ac.za  Centre for Applied Data Science, University of Johannesburg, PO BOX 524, Auckland Park, 2006, Gauteng, South Africa

© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

data breaches, malware attacks, unauthorized access, denial-of-service (DoS) attacks, zero-day exploits, and social engineering have surged exponentially in recent years (McIntosh et al. 2019; Sarker et al. 2020; Sun et al. 2018). These security incidents pose significant risks to individuals, businesses and corporate organizations (Cremer et al. 2022). According to Künzler (2023), the impacts of cybersecurity breaches are estimated to have cost an average of 4.35 million dollars in 2022. This estimation is based on the International Business Machines Corporation (IBM) survey of roughly 3600 individuals from 550 enterprises (Künzler 2023). The global indicator also estimated the cost of cybersecurity breaches to continuously increase between 2023 and 2028 by a total of 5.7 trillion U.S. dollars (Petrosyan 2023). Furthermore, the World Bank (2024) reported a global estimate of \$5.2 trillion loss in value from cyber attacks between 2019 and 2023. According to Cisco (2022), the most common types of security incidents include data breaches (51.5%), network or system outages (51.%), ransomware events (46.7%), and distributed DoS attacks (46.4%). Consequently, it is imperative for organizations to embrace and enact robust cybersecurity measures to minimize potential losses. As noted by Sarker et al. (2020), a nation's security hinges on governments, businesses, and individual citizens having access to highly secure applications and tools, along with the ability to swiftly detect and prevent cyber threats. Hence, there is a need to leverage Artificial Intelligence (AI) techniques such as Machine Learning (ML) techniques and data-driven insights, to enhance cybersecurity capabilities.

In recent years, cybersecurity has been experiencing significant transformations in both technology and operational aspects within computing systems (Künzler 2023). AI stands out as a driving force behind these changes, with ML and Deep Learning (DL) as fundamental components of AI, and positioned to play a crucial role in uncovering insights from data (Sarker 2023). According to Sarker et al. (2020), AI is an interesting tool capable of providing analytics and intelligence to protect against constantly evolving cyber threats. By rapidly analyzing millions of events and monitoring diverse cyber threats, AI can forecast and proactively address issues (Wirkuttis and Klein 2017). Consequently, AI is increasingly being incorporated into the cybersecurity framework and employed across various scenarios to automate security functions or complement the efforts of human security teams (Hofstetter et al. 2020). AI represents a pivotal technology for current and future information systems, with numerous domains already harnessing its capabilities (Sarker, Furhad, and Nowrozy 2021).

While the benefits of AI are well-documented, the literature reveals a gap in thoroughly evaluating the performance and suitability of various AI techniques for specific cybersecurity tasks (Zhang et al. 2022b). Although AI methods such as ML, DL, and natural language processing (NLP) are increasingly being integrated into cybersecurity solutions, many studies tend to focus on

broad implementations rather than systematically comparing different techniques for targeted applications. For example, intrusion detection systems, malware detection, phishing prevention, and anomaly detection each have unique requirements, but the literature lacks in-depth studies that assess which AI models are most effective for these specific tasks. In addition, Sommer and Paxson (2010) highlight a related concern, noting that many studies apply ML techniques without sufficient assessment of their practical suitability or performance under real-world conditions. They emphasize the need for systematic studies that explore not only the technical effectiveness of different algorithms but also their scalability, interpretability, and adaptability to changing threat landscapes. In other words, while many ML techniques show promise in controlled settings, there is a limited understanding of how they perform when applied to real-time cybersecurity challenges, such as handling novel attacks or minimizing false positives. This gap in the literature presents an opportunity for this research to be addressed. Hence, this paper focuses on providing a comprehensive review of the different AI techniques along with their applications within the cybersecurity domain. While this paper does not delve deeply into every individual technique employed in cybersecurity, it offers a broad overview of AI modeling for cybersecurity. Accordingly, four research questions (RQs) were posed:

RQ1. What are the related scholarly articles about AI and cybersecurity from 2014 to 2024?

RQ2. What are the various AI techniques and datasets used in cybersecurity?

RQ3. What are the AI tools used for data extraction, analysis, and optimization?

RQ4. What are the future research directions for the application of AI in cybersecurity?

The remainder of the paper is structured as follows. Section 2 explains the research methodology employed. Section 3 presents the findings of the study as they relate to the research questions. Section 4 provides an in-depth discussion of the findings based on the evidence presented in Section 3, thus expanding the frontiers of knowledge on the application of AI in cybersecurity.

Research methodology

The study adopted a hybrid review (bibliometric and systematic analysis) of relevant academic literature related to the application of AI in cybersecurity.

Bibliometric analysis is a statistical method used to study patterns and trends within academic literature (Donthu et al. 2021). It involves the statistical analysis of bibliographic data, which includes information such as citations, publication dates, authors, journals, and keywords. Hence, in this study, bibliometric analysis was employed to identify and analyze various aspects, including the completeness of bibliographic metadata of articles related to AI and cybersecurity. The bibliometric analysis was conducted using R-software version 4.4.0 via Biblioshiny, a web-based application that provides a user-friendly graphical interface for the R package Bibliometrix (Patil 2020).

On the other hand, a systematic literature review (SLR) is characterized by a methodical and replicable analytical method synthesizing data directly related to the systematic review question (Mallett et al. 2012). SLR is a process that involves collecting relevant evidence on a given topic according to pre-specified eligibility criteria and provides answers to the formulated research questions (Sarkis-Onofre et al. 2021). To offer a comprehensive review of the application of AI in cybersecurity, this study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) standard (Liberati et al. 2009). The PRISMA framework is a widely used protocol for reporting systematic reviews and meta-analyses (Sarkis-Onofre et al. 2021). It consists of a checklist and a flow diagram to help ensure the transparent and complete reporting of research. As presented in Figure 1, the PRISMA standard includes identification, screening (for eligibility), and included. The identification refers to the total records identified through database searching and other methods (Albhirat et al. 2024). In this study, a comprehensive search was conducted on two major electronic databases, namely, Clarivate's Web of Science (WoS) and Scopus yielding 14,509 research papers, which served as a pool of potential articles for subsequent selection, as shown in Figure 1. The screening refers to the number of records after duplicates are removed and those remaining are screened for relevance and eligibility (Sarkis-Onofre et al. 2021). In this study, 49 duplicates were excluded. Included refers to the final number of studies considered in the review, with a total of 939 records being included in the final analysis.

Data extraction and data analysis

A comprehensive literature review was conducted on the application of AI in cybersecurity from 2014 to 2024. This study obtained bibliometric data from the Scopus and WoS databases which were accessed on 12 June 2024. Using a comprehensive search strategy a total of 14,509 records were identified of which 9611 came from the Scopus database, and 4898 from the WoS platform. Scopus and the WoS were selected due to their comprehensive coverage, high-quality content, advanced analytical tools, and robust features. According to Pranckutė (2021),

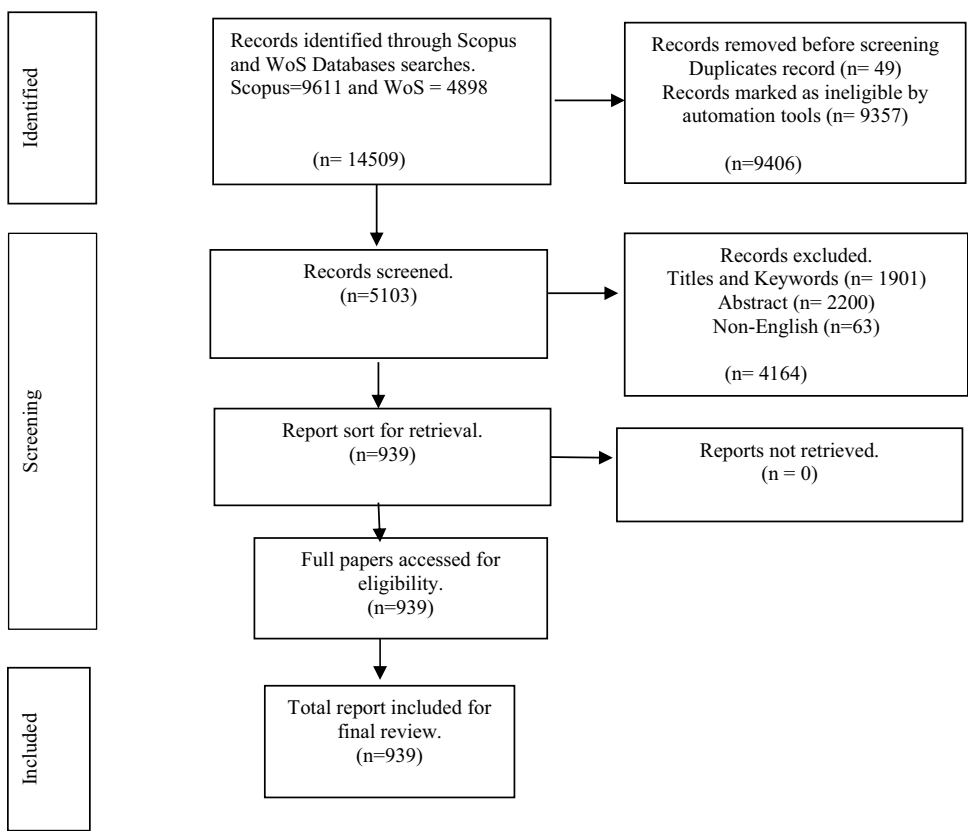


Figure 1. PRISMA systematic literature review (Page et al. 2021).

Scopus and WoS are major abstract and citation databases, providing access to thousands of peer-reviewed journals, books, and conference proceedings across diverse disciplines. These two databases are among the most credible platforms that offer comprehensive access to a wide range of research resources. The search strategy employed the following basic search string applied to all searchable fields of the Scopus and WoS databases:

("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Deep learning" OR "Natural Language Processing" OR "Robotics" AND "Cybersecurity" OR "Cyber-security" OR "Cyber security" OR "Information security" OR "Computer security").

The use of quotation marks was adopted in order to include variations of the search terms. The data extracted for the SLR were synthesized using the core themes identified. Thematic synopsis was crucial for examining how AI techniques are used in cybersecurity.

Inclusion and exclusion criteria

The subject area of our search was limited to the disciplines “Computer Science,” “Information Technology,” “Engineering,” “Decision Science,” “Mathematics” and “Multidisciplinary.” In addition to using the search string to identify articles, further criteria were then applied to refine the search results to include records that met the following conditions: (i) authored exclusively in English; (ii) of all types, excluding News Item, Correction, Meeting Abstract, Data Paper, Book Review, and Letter; and (iii) published by the top 10 publishers ranked by the number of records returned from the search string. These publishers include IEEE, Springer Nature, Elsevier, Sage, Taylor & Francis, MDPI, Wiley, and Emerald Group Publishing. Also, only articles that fell within the period 2014 to 2024 were included in the study.

The date range 2014 to 2024 was chosen to ensure that the literature review captures the most relevant and impactful developments over the past decade. This period aligns with significant advancements in both AI and cybersecurity, including breakthroughs in ML, DL, and the rise of novel cybersecurity threats like ransomware and advanced persistent threats (APTs) (Sarker 2023). By focusing on the last ten years, the review reflects current trends, practices, and challenges while also ensuring that foundational studies and early implementations of key technologies are not overlooked. This balance allows the research to contextualize the latest findings within a decade of evolving innovations and threats, making the analysis both current and comprehensive. The 939 articles obtained were subsequently classified into four major article types (as depicted in Table 1).

Table 1 presents the distribution of the reviewed research articles. It shows the number of papers on the application of AI in cybersecurity from 2014 to 2024. The majority of articles (71.4%) were published in journals, followed by conference proceedings (26.4%), symposiums (1.91%), and workshops (0.31%). It should be noted that 52 of the 670 journal articles were used for the SLR analysis.

Table 1. Distribution of the reviewed research papers.

Reviewed papers	Number of articles	Percentage
Journals	670	71.4%
Conference proceedings	248	26.4%
Symposiums	18	1.91%
Workshop	3	0.31%

Findings

The section presents a hybrid analysis, which includes the bibliometric analysis and the SLR findings. The bibliometric analysis was carried out using

R-software (Biblioshiny) and was used to answer RQ1. The systematic analysis employed a qualitative research design to explore the various AI techniques used in cybersecurity. Qualitative research provides an in-depth analysis and a comprehensive overview by examining patterns and developing insights from the data content (Huberman 2014). The data were then synthesized using the core themes identified.

Bibliometric analysis

The following section answers RQ1 regarding the scholarly articles on AI and cybersecurity.

Completeness of bibliographic metadata

The completeness of bibliographic metadata is essential for the effective management and dissemination of research publications. This metadata is crucial for identifying, cataloging, retrieving, and evaluating research publications. It further enhances the visibility, accessibility, and impact of scholarly works. Figure 2 shows all the key components that are accurately recorded. The majority of the metadata, such as AB-Abstract, AU-Author, DT-Document Type, SO-Journal, LA-Language, PY-Publication Year, TI-Title and TC-Total citation has no missing count and status is “Excellent.” This indicates that the metadata used in this study has been carefully curated and

Completeness of bibliographic metadata - 939 documents from Isi				
Metadata	Description	Missing Counts	Missing %	Status
AB	Abstract	0	0.00	Excellent
AU	Author	0	0.00	Excellent
DT	Document Type	0	0.00	Excellent
SO	Journal	0	0.00	Excellent
LA	Language	0	0.00	Excellent
PY	Publication Year	0	0.00	Excellent
TI	Title	0	0.00	Excellent
TC	Total Citation	0	0.00	Excellent
C1	Affiliation	3	0.32	Good
DE	Keywords	47	5.01	Good
DI	DOI	56	5.96	Good
ID	Keywords Plus	210	22.36	Poor
RP	Corresponding Author	292	31.10	Poor
CR	Cited References	939	100.00	Completely missing

Figure 2. Bibliographic metadata completeness (source: biblioshiny).

meets high standards of data integrity and quality. Furthermore, it shows that the articles from the database can be used for academic research since the researcher has been provided access to complete and accurate bibliographic records. However, while the metadata C1-Affiliation, DE-Keywords and DI-Keyword Plus have a missing count of 3, 47 and 56 respectively, and the status is “Good,” ID and RP status is “Poor” with 210 and 292 missing counts respectively. A “Good” status indicates that, despite the few missing entries, the overall quality and completeness of these metadata fields are acceptable and generally reliable. Thus, in this study, C1, DE and DI indicate that the majority of records have author affiliations, keyword descriptors and digital object identifiers, which are generally considered sufficient and reliable. The “Poor” status indicates the lack of document identifiers and reprint authors which makes it difficult to uniquely identify the records within the database and to contact the responsible author for additional information or copies of the work. Only CR-Cited References is completely missing with 939 missing counts. This implies that none of the 939 records include any information about the references that were cited within those publications.

Authors’ publications per year versus citations per year

Table 2 presents the analysis and evaluation of the authors’ publication output contributing to the application of AI in cybersecurity research

Table 2. Authors’ publication per year versus citations per year.

Author	Year	Freq.	Total Citation (TC)	TCpY
AL-TURJMAN F	2019	1	155	25.833
	2022	3	19	6.333
	2023	2	1	0.5
ALI A	2022	1	51	17
	2023	1	5	2.5
	2024	3	3	3
CHEN Y	2018	1	757	108.143
	2023	4	24	12
KIM J	2020	1	23	4.6
	2021	1	3	0.75
	2022	1	2	0.667
	2023	2	0	0
	2024	1	1	1
LI J	2023	5	15	7.5
	2024	1	0	0
LI X	2019	1	1	0.167
	2023	3	2	1
	2024	3	0	0
LI Y	2018	1	757	108.143
	2021	1	0	0
	2022	2	3	1
	2023	3	18	9
LIU Y	2019	1	1	0.167
	2021	1	1	0.25
	2024	4	4	4
MOHAMED A	2022	1	5	1.667
	2023	5	13	6.5
RAGAB M	2022	2	4	1.333
	2023	3	6	3
	2024	1	0	0

over the past 10 years (2014–2024). This involves examining various metrics and trends, such as the number of publications per year, the total citation count (TC), and the total citation per year (TCpY). These trends provide valuable insights into an author’s research trajectory in their field. For example, while AL-TURJMAN’s publications in the year 2019 was one, the number of citations per year (25.833) highlights that although the publication rate was steady, the citation count per publication was high, demonstrating significant impact and recognition in the field of AI and cybersecurity. AL-TURJMAN’s publications in the year 2022 and 2023 were three and two respectively and did not yield many citations per year (i.e. 6.333 and 0.5 respectively). This highlights a low citation-to-publication ratio.

Country production over time

Table 3 highlights a list of the top 10 countries that have significantly contributed to the application of AI in cybersecurity research in the past 10 years. Leading the list are China (617 articles), the USA (578 articles), India (492 articles), the United Kingdom (226 articles), and Germany (171 articles). This analysis serves as an eye-opener for countries that are falling behind in this critical area. Growth in research output is a key indicator of progress for nearly every country. With technological advancements occurring at an accelerated pace worldwide, it is crucial for each country to stay competitive in this technological race.

Table 3. Country production over time.

Rank	Country	Articles
1	China	617
2	USA	578
3	India	492
4	United Kingdom	226
5	Germany	171
6	Italy	167
7	Australia	133
8	Saudi Arabia	131
9	Spain	91
10	Egypt	70

Country citation

Table 4 is a compilation of the top 10 countries that have been actively engaged in the application of AI in cybersecurity research over the past 10 years in terms of their total citation. China leads with 1379 total citations, followed closely by Australia with 915 citations, Algeria with 744 citations, the USA with 427 citations, and India with 286 citations. With the rapid global technological advancements, it has become imperative for every country to keep pace with the technological landscape. China, despite having the highest total

citations (1379), has a relatively moderate average article citation (12.20). Algeria, with 744 total citations, has the highest average article citation (148.80), indicating fewer but highly cited articles. The USA, while having a substantial number of total citations (427), has a lower average article citation (7.40). India shows a lower performance in both total citations (286) and average article citations (3.20). This comparison highlights the variations in both the volume and impact of research publications from different countries, with some countries producing fewer but more impactful articles, and others producing a larger volume with varying citation impacts.

Table 4. Country citation.

Rank	Country (TC)	Total Citation (TC)	Average Article Citation
1	China	1379	12.20
2	Australia	915	41.60
3	Algeria	744	148.80
4	USA	427	7.40
5	India	286	3.20
6	Canada	255	28.30
7	Saudi Arabia	251	6.30
8	Pakistan	204	34.00
9	Turkey	204	29.10
10	United Kingdom	182	6.70

Most relevant affiliations

This section presents the top 10 universities worldwide that have made substantial contributions to research output in the application of AI in cybersecurity during the period reviewed. As shown in [Table 5](#), the most active universities include King Abdulaziz University (48 articles), Kennesaw State University (42 articles), Graphic Era Deemed to be University (34 articles), King Khalid University (34 articles) and Prince Sattam Bin Abdulaziz University (33 articles). Given that there is a significant global demand for research in technological advancements in cybersecurity, this insight provides valuable insights into the leading institutions in this research field.

Table 5. Most relevant affiliations.

Rank	Affiliation	Articles
1	King Abdulaziz University	48
2	Kennesaw State University	42
3	Graphic Era Deemed to Be University	34
4	King Khalid University	34
5	Prince Sattam Bin Abdulaziz University	33
6	Princess Nourah Bint Abdulrahman University	31
7	King Faisal University	26
8	Khalifa University	24
9	SRM Institute of Science and Technology	19
10	Zhejiang University	18

Most relevant authors

Table 6 presents a compilation of the top 10 authors actively contributing to the field of AI and cybersecurity. The top author publishing in the field is Zhang Y, with nine articles, followed by the authors Li X, Li Y, Wang S, Wu F, and Zhang Z who have each published seven articles, and Al-Turjman, Kim J, Li J, and Liu Y with six publications each. The analysis can serve as a valuable source of reference for future researchers in terms of the leading researchers in the field.

Table 6. Most relevant authors.

Rank	Author	Articles
1	Zhang Y	9
2	Li X	7
3	Li Y	7
4	Wang S	7
5	Wu F	7
6	Zhang Z	7
7	Al-Turjman F	6
8	Kim J	6
9	Li J	6
10	Liu Y	6

Keyword analysis

Keywords constitute one of the major aspects of search engines. The selection of correct keywords is extremely important for a higher number of relevant citations. The keywords that are most searched for when AI and cybersecurity are the main keywords are given below (Table 7 and Figure 3). Analysis shows

Table 7. Keywords word cloud.

Rank	Keywords	Occurrences
1	Cybersecurity	376
2	Machine Learning	256
3	Artificial Intelligence	207
4	Deep Learning	152
5	Intrusion Detection	64
6	Information Security	36
7	Security	25
8	Classification	23
9	Malware Detection	20
10	Internet of Things	19



Figure 3. Keywords word cloud (source: biblioshiny).

the frequency occurrence of the top 10 keywords. The term Cybersecurity (376) appears most frequently, followed by Machine Learning (256), Artificial Intelligence (207), and Deep Learning (152). Cybersecurity, as the most frequently occurring keyword, indicates a high level of interest and research activity in the field of cybersecurity, reflecting its importance in protecting digital systems and data from cyber threats and attacks. Similarly, Machine Learning appears 256 times, making it the second most frequent keyword. This suggests a strong focus on the development and application of algorithms that enable computers to learn from and make predictions based on data. Likewise, Artificial Intelligence is mentioned 207 times, indicating significant ongoing research and development in this domain. Moreover, Deep Learning is cited 152 times. As a subset of ML, DL involves neural networks with many layers (deep neural networks) and is particularly effective in tasks such as image processing and speech recognition. Its relatively high frequency highlights its importance in current research trends. The occurrences of these keywords reflect the current trends and research priorities. High-frequency keywords such as “Cybersecurity,” “Machine Learning,” and “Artificial Intelligence” indicate a strong focus on leveraging advanced technologies to enhance security measures. Keywords such as “Intrusion Detection” and “Malware Detection” highlight specific areas of interest within the broader field. The presence of the “Internet of Things” suggests growing concern over securing emerging technologies.

Thematic map

A thematic map provides a visual representation of key themes within a research field (Cobo et al. 2011). Using author keywords, a thematic map was created as shown in Figure 4. This visualization clusters keywords to reveal research field themes. The map has two dimensions: centrality (x-axis) and density (y-axis). Centrality indicates the importance of a particular theme, while density reflects the theme’s level of development. The 2×2 matrix of the thematic map produced four quadrants, with the bubble size representing the frequency of keyword occurrences.

In the upper left quadrant are the niche themes, which are well-developed but isolated (niche) themes. The clusters in the niche themes indicate areas where there is significant depth in research and development, although they may not yet be central to the broader field (Tennekes 2018). Understanding these niche themes can guide researchers and practitioners to delve into highly specialized and potentially impactful areas of study. The clustering of “machine learning,” “random forest,” and “decision tree,” indicate a closely related set of methodologies within ML (Charbuty and Abdulazeez 2021). Their grouping suggests that these specific algorithms and techniques are

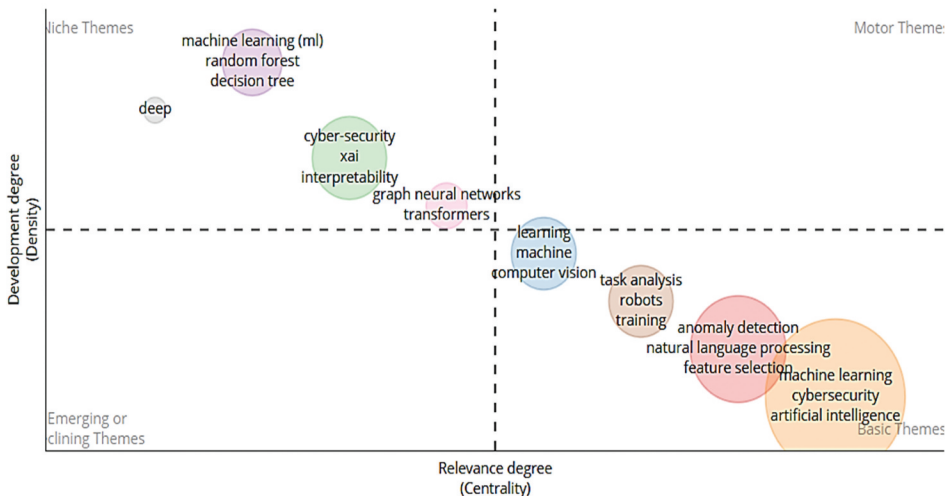


Figure 4. Thematic map (source: biblioshiny).

being extensively developed and studied together, forming a niche within ML research (Tareq and Davud 2024). Similarly, the isolation of “deep” suggests focused and specialized research in DL. Its separation from other terms indicates that it is a broad area with unique, standalone developments or issues that are distinct from other ML techniques (Dong, Wang, and Abbas 2021). Furthermore, the clustering of “cybersecurity,” “xai” (Explainable Artificial Intelligence), and “interpretability,” suggests significant research efforts focused on making AI systems more transparent and trustworthy in the context of cybersecurity (Zhang et al. 2022a). The clustering of these keywords highlights the importance of developing secure AI systems that are also interpretable by humans, a crucial requirement for trust and regulatory compliance (Zhang et al. 2022a). In addition, the clustering of “graph neural networks” (GNN) and “transformers” indicates focused research on these cutting-edge techniques (Aflalo et al. 2023). This grouping suggests advancements and applications in domains that require processing complex structures (GNN) and sequential data (Transformer), indicating their importance and growing adoption in the research community (Aflalo et al. 2023).

The upper right quadrant represents the motor themes which are the most frequently discussed topics in the field. However, the empty “Motor Themes” quadrant in Figure 4 indicates that the field has matured, and no single theme is currently driving significant new development. This could also mean that research is evenly distributed across various established areas without one dominating focus (Tennekes 2018). It also suggests that research efforts are more fragmented, with no single cohesive theme emerging as a dominant driver (Foody 2020). Furthermore, the absence of motor themes could also reflect a recent shift in focus, where previously dominant themes have either

become too broad to be categorized as a single theme or the research community is in transition, looking for new directions (Tennekes 2018).

In the lower right quadrant, we find the themes that are considered basic and transversal, with low levels of development but high levels of centrality and relevancy to the application of AI in cybersecurity. Basic themes are foundational topics that have high centrality but may vary in density. These themes are essential for the field and often connect with various other research areas. In this study, the clustering of “learning,” “machine,” and “computer vision” suggests that research in ML and computer vision is fundamental and interconnected (Khan and Al-Habsi 2020). ML techniques are crucial for developing computer vision applications, and both are core areas of study in AI. Likewise, the clustering of “task analysis,” “robots,” and “training” highlights the importance of understanding and improving task performance in robotics (Janati et al. 2017). Training robots to perform tasks efficiently is a critical aspect of robotics research, and task analysis is essential for optimizing these processes (You et al. 2023). Furthermore, the clustering of “anomaly detection,” “natural language processing” (NLP), and “feature selection” suggests a foundational link between methods for improving the performance of ML models (feature selection) and their applications in NLP and anomaly detection (Bertero et al. 2017). Efficient feature selection enhances the effectiveness of anomaly detection and NLP systems (Pranto et al. 2022). Additionally, the clustering of “machine learning,” “cybersecurity,” and “artificial intelligence” indicates that cybersecurity is a critical application area for AI and ML (Dasgupta, Akhtar, and Sen 2022). AI techniques, particularly ML, are essential for developing advanced cybersecurity measures, highlighting their foundational role in this interdisciplinary research area (Sarker, Furhad, and Nowrozy 2021).

In the lower left quadrant are the emerging or declining themes. In Figure 4, there are no themes/clusters in the lower left quadrant which suggests that there are no currently identifiable new or fading trends in the dataset being analyzed. This could also indicate a period of stability in the research field (Foody 2020). It might also reflect limitations in the dataset as the parameter for this study dataset was adjusted from the default setting of 250 for the number of words to 100. The adjustment was necessary because the default setting produced overlapping clusters that were muddled together making it difficult to see and interpret. Future research may research may readjust the parameters for the datasets to identify emerging or declining patterns.

Thematic evolution

This section presents an analysis of the thematic evolution as it relates to AI in cybersecurity based on author keywords. Exploring thematic evolution provides an interesting broad picture of the development in the field. Such

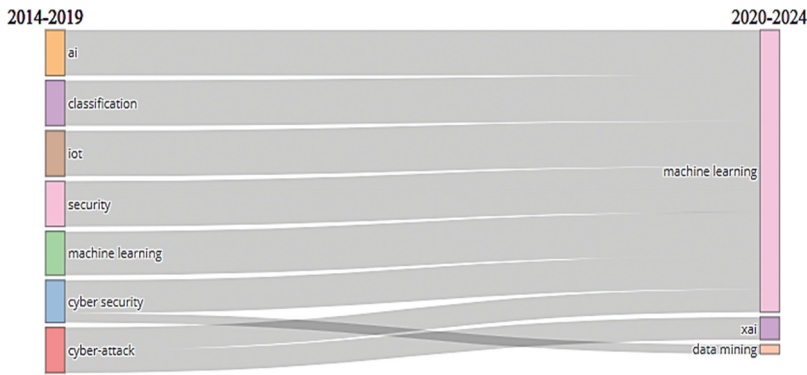


Figure 5. Thematic evolution (source: Biblioshiny).

longitudinal analyses allow for highlighting how topics merge or split into several themes (Madsen, Berg, and Nardo 2023). Figure 5 shows the thematic evolution of AI in cybersecurity research in the period 2014–2024 by dividing it into two time slices or subperiods. Accordingly, we are looking at the evolution of keywords during two different periods (2014–2019 and 2020–2024). The reason for choosing these time slices is that there was an uptick in the volume of research in 2020, coinciding with the COVID-19 pandemic, which some researchers suggested has fueled the shift toward the application of AI in cybersecurity (Hofstetter et al. 2020). Therefore, exploring whether these surges were associated with the emergence and development of new research themes is interesting.

In the first phase, which is the pre-COVID-19 pandemic (2014–2019), the dominant themes are “ai,” “classification,” “iot,” “security,” “machine learning,” “cyber security” and “cyber-attack.” This implies that there is an emphasis on foundational technologies (AI, ML) and their applications (IoT, cybersecurity) in an increasingly digital world (Jallouli et al. 2019). In the second phase, which is during and post-COVID-19 pandemic (2020–2024), the dominant themes are “machine learning,” “xai” and “data mining.” This implies that there is a greater focus on enhancing the interpretability and trustworthiness of AI (XAI) and leveraging data to address unprecedented global challenges (data mining) (Oropeza-Valdez et al. 2024). The critical need for reliable and understandable AI systems during crises highlighted the importance of XAI (Hofstetter et al. 2020; Oropeza-Valdez et al. 2024). Moreover, the pandemic highlighted the necessity of data mining in making informed decisions in real-time (Li et al. 2022). The shift toward ML, XAI, and data mining highlights the need for powerful, transparent, and data-driven solutions in an increasingly complex world.

Trend topics

Figure 6 highlights the trending topics based on author keywords. In the years 2023–2024, the most trending topics are “natural language processing,” “reinforcement learning,” “data privacy,” “machine learning,” “cybersecurity” and “artificial intelligence.” This trend is obviously moving toward the application of AI (Nievas et al. 2024). This implies that the landscape has shifted dramatically in recent years, with a more proactive approach to cybersecurity and a broader application of AI technologies (Sarker, Furhad, and Nowrozy 2021). However, the years 2020–2022 (Figure 6) reflect a reactive approach to cybersecurity, focusing on detecting and responding to threats after they occur. This is based on the trending topics in that period which are “intrusion detection,” “information security,” “intrusion detection systems,” “cyber threat intelligence,” “social media,” “explainability,” “big data,” “data mining” and “malware.” This implies that the period between 2020 and 2022 was characterized by a strong emphasis on defensive cybersecurity strategies (Dawson et al. 2021).

Hence, while 2023–2024 focuses on cutting-edge AI and data privacy, 2020–2022 was more concentrated on foundational cybersecurity measures and the initial phases of AI explainability and big data analytics. These trends reflect a more sophisticated approach to cybersecurity, leveraging AI to predict and prevent threats, rather than simply reacting to them. It is important to note the absence of trending topics between the years 2014–2019. This may imply that the prior research focus might have been different, with different keywords or terminologies used, making it difficult to identify using current search parameters. Furthermore, there can be a significant delay between research completion and publication. This might affect the visibility of older trends.

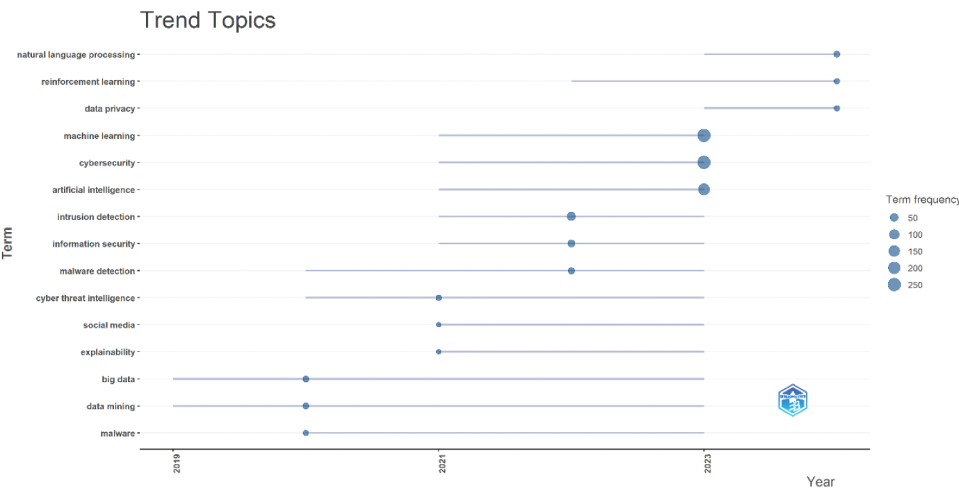


Figure 6. Trend topics.

Annual scientific production

Figure 7 presents the number of articles published each year from 2014 to 2024. The number of articles increased gradually, from 2 in 2014 to 6 in 2017, indicating limited research interest or early exploration in the field during these years. There is noticeable growth in publications, with 17 articles in 2018 and 43 in 2019. This suggests increasing awareness or relevance of the topic. The numbers spike significantly, especially during the pandemic years, from 72 articles in 2020 to 376 in 2023. This suggests a surge in research interest, possibly due to the growing reliance on digital solutions and the increased importance of AI and cybersecurity during and after the COVID-19 pandemic. With 376 articles, 2023 marks the highest number of publications in the dataset, reflecting the topic’s growing importance, rapid development, or increased funding and innovation in these areas. However, there is no report of publication in 2024 as of the time of this analysis. Overall, the table indicates a sharp upward trend, especially in recent years, showing that AI and cybersecurity are becoming central topics of academic and practical importance.

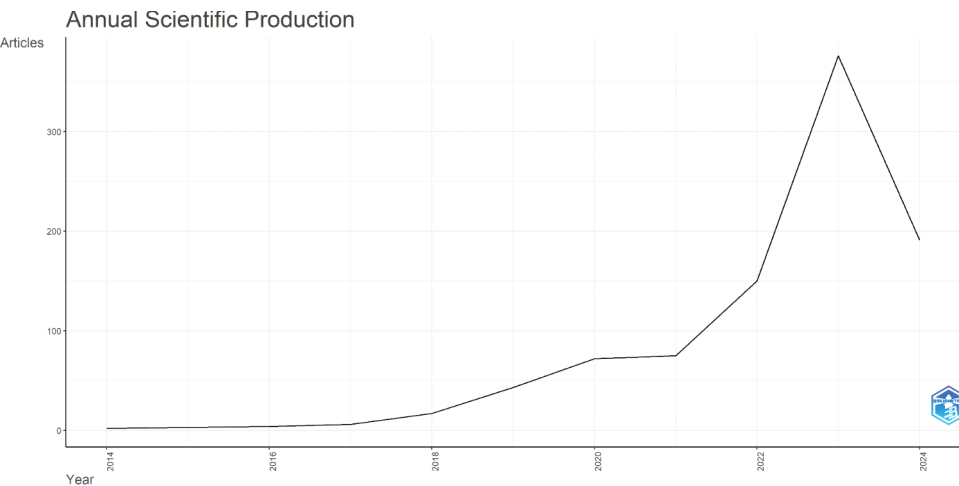


Figure 7. Annual scientific Production.

Systematic analysis

Systematic analysis is a methodical and structured approach to examining a system, process, or problem question (Mallett et al. 2012). It involves breaking down complex systems or issues into smaller, more manageable components and then evaluating each component systematically to gain a comprehensive understanding. In this section, systematic analysis provides a structured and rigorous approach to understanding the various AI techniques used in cybersecurity and how they are applied to curb cyberattacks. Table 8 presents summaries of the top 10 articles identified in the SLR. These

Table 8. Top 10 articles on the application of AI in cybersecurity.

Rank	Authors	Aims/objective	Techniques	Findings	Recommendations
1	Chowdhury et al. (2017)	To explore how Self-Organizing Maps (SOM) can be used to detect botnet activities within network traffic data, specifically focusing on the CTU-13 dataset. To evaluate the efficiency of SOM in terms of computational requirements and performance when applied to large-scale network traffic datasets. To create a reliable and efficient framework using SOM for identifying botnet traffic within the CTU-13 dataset.	SOM	The study found that SOM was effective in distinguishing between botnet traffic and normal traffic. The implementation of SOM achieved high detection rates for botnet traffic within the CTU-13 dataset. This demonstrated SOM's capability to identify various botnet activities accurately. The proposed SOM-based framework was validated through extensive testing on the CTU-13 dataset, showcasing its robustness and reliability in botnet detection.	Given its effectiveness and efficiency, the study recommended implementing SOM in real-time bot detection systems to enhance cybersecurity measures. The authors suggested that further research should be conducted to refine SOM algorithms and explore their application to other types of cyber threats.
2	Kozik et al. (2014)	To improve the detection of web attacks by employing advanced ML techniques. To compare the performance of DT, SVM, and Naive Bayes algorithms in terms of detection accuracy, false positive rates, and computational efficiency.	DT, SVM, Naive Bayes	The algorithms showed high accuracy in detecting various types of web attacks. DT demonstrated high accuracy in classifying network intrusions, making it a reliable algorithm for intrusion detection systems (IDS). SVM showed strong performance, especially in handling complex and non-linear data patterns. This makes it suitable for detecting sophisticated intrusion attempts. Naive Bayes was found to be fast and simple to implement. It requires less computational resources compared to DT and SVM, making it efficient for real-time intrusion detection.	DT was recommended for scenarios where model interpretability and rule extraction are important. However, regularization techniques such as pruning should be applied to avoid overfitting. SVM: Preferred for its robustness and balanced performance, but careful tuning of hyperparameters is essential. It is suitable for applications where computational resources are available for training. Naive Bayes: Suitable for real-time detection systems due to its speed but may require additional techniques to handle feature correlations and improve accuracy. The recommendations focused on algorithm selection, feature engineering, hybrid approaches, regular updates, scalability, and real-time detection, providing a comprehensive framework for improving botnet detection systems in practical network environments.
3	Bhuyan, Bhattacharyya, and Kalita (2015)	To evaluate the effectiveness of various ML algorithms in detecting different types of botnets.	C4.5, SVM, K-Nearest Neighbors (KNN), Bayesian Networks	The findings highlighted the strengths and weaknesses of each algorithm, with C4.5 and SVM emerging as top performers.	

(Continued)

Table 8. (Continued).

Rank	Authors	Aims/objective	Techniques	Findings	Recommendations
4	Bhamare et al. (2016)	To compare the performance of different ML classifiers, including Neural Network Classifier (NNC), Artificial Neural Network (ANN), SVM, Naive Bayes Classifier (NBC), and Gradient Boosting Classifier (GBC). To assess the effectiveness of these algorithms in accurately detecting cyber threats within the ISOT dataset.	NNC, ANN, SVM, NBC, GBC	Both NNC and ANN demonstrated high accuracy and effectiveness in detecting cyber threats. ANN, in particular, showed superior performance due to its ability to model complex patterns in the data. SVM also performed well, especially in handling high-dimensional data, but its performance was slightly lower compared to ANN. The NBC showed good results in terms of speed and simplicity but was outperformed by more complex algorithms like ANN and SVM in terms of accuracy. The GBC showed robust performance, combining accuracy with efficiency in processing the data.	For applications requiring high accuracy and robustness, GBC and SVM were recommended due to their superior performance in handling complex datasets, ability to manage high-dimensional data, and strong generalization capabilities across a variety of tasks. ANN was recommended for scenarios where complex patterns and relationships need to be learned from the data, provided that sufficient computational resources are available. NBC and NNC were recommended for quick, preliminary analysis due to their simplicity and speed, though they may not provide the highest accuracy.
5	Kozik et al. (2014)	To evaluate the effectiveness of the Naive Bayes algorithm in classifying network intrusion data from the KDD 99 dataset. To determine how the application of PCA impacts the performance of the Naive Bayes classifier in terms of accuracy, speed, and resource utilization.	Naive Bayes, PCA algorithm	The Naive Bayes algorithm demonstrated robust performance in classifying different types of network intrusions within the KDD 99 dataset. The application of PCA significantly reduced the dimensionality of the dataset, which in turn improved the processing speed and efficiency of the Naive Bayes classifier.	It was recommended to apply PCA as a preprocessing step before classification, especially when dealing with high-dimensional datasets like KDD 99. This reduces the computational burden and improves classification accuracy. Naive Bayes was suggested as a suitable algorithm for initial deployment due to its simplicity and accuracy, especially in resource-constrained environments.
6	Kato and Klyuev (2014)	To develop a robust and efficient method for detecting DDoS attacks using statistical techniques. To identify statistical patterns associated with DDoS attacks.	Statistical methods	The application of statistical methods such as entropy-based analysis, standard deviation, and correlation analysis proved effective in detecting DDoS attacks. By analyzing network traffic patterns and anomalies, these methods were able to identify abnormal behaviors indicative of DDoS attacks.	It was recommended to implement these statistical methods in real-time detection systems. The ability to process and analyze traffic data in real-time is crucial for timely detection and mitigation of DDoS attacks. It was also recommended that future studies should focus on real-time implementation and validation of these methods in operational network environments.

(Continued)

Table 8. (Continued).

Rank	Authors	Aims/objective	Techniques	Findings	Recommendations
7	Xie, Hu, and Slay (2014)	To enhance the detection of anomalies in network traffic using advanced ML techniques. To compare the effectiveness of the short sequence model against other algorithms (Naïve Bayes, Bayes Network, Decision Stump, RBF Network) using the ECML-PKDD 2007 dataset.	Naïve Bayes, Bayes Network, Decision Stump RBF Network	The short sequence model outperformed all compared algorithms (Naïve Bayes, Bayes Network, Decision Stump, RBF Network) in terms of accuracy and false positive rates. It showed a balanced performance, combining accuracy with computational efficiency, making it suitable for real-time applications. Traditional methods like Decision Stump and Naïve Bayes were less effective in capturing short-term dependencies critical for anomaly detection.	The study recommended exploring hybrid models combining the strengths of the short sequence model with other ML techniques. Further research was suggested to test the model's scalability and robustness on larger and more diverse datasets.
8	Moustafa and Slay (2016)	To evaluate the effectiveness of the k-means clustering algorithm in distinguishing between normal and malicious traffic within this complex UNSW-NB15 dataset. To apply the k-means clustering algorithm to the dataset and assess its ability to form meaningful clusters that separate different types of network traffic.	k-means	The k-means algorithm was able to form clusters that provided a basic separation between normal and malicious traffic. Findings suggest that while k-means alone is effective in differentiating between normal and malicious traffic, combining it with supervised learning methods can significantly enhance the performance of IDS.	The recommendations focused on hybrid approaches, feature selection, real-time implementation, and continuous model updating to create robust and efficient IDS solutions.
9	Hoque, Bhattacharyya, and Kalita (2016)	To evaluate and compare the effectiveness of different information metrics for detecting DDoS attacks. To identify which information metrics are most effective for DDoS detection in terms of accuracy, precision, and other relevant metrics.	Information metrics	The findings reveal that metrics such as entropy-based measures showed strong performance in identifying anomalies associated with DDoS attacks. Other metrics, such as Kullback-Leibler divergence and Mutual Information, also demonstrated good performance but varied depending on the specific characteristics of the dataset. By optimizing the parameters, the genetic algorithm was able to achieve higher detection rates and lower false positive rates. The study findings suggest that incorporating the GA for dimensionality reduction can lead to more accurate and efficient IDS.	The study recommended that a combination of information metrics could potentially offer improved detection capabilities by leveraging the strengths of different metrics. It also recommended further research into the integration of these metrics into real-time DDoS detection systems to enhance their practical applicability. Recommendations focus on adopting the GA for feature selection, combining it with other algorithms, ensuring real-time application and scalability, continuously updating models, and exploring further research opportunities to maintain and improve intrusion detection.
10	Wang and Paschalidis (2016)	To improve the accuracy and efficiency of anomaly detection methods in network security. To identify optimal parameters for the genetic algorithm (GA) to maximize detection performance.	Genetic algorithm		

articles were selected based on their direct relevance to the research questions of the study. Only those articles that provide significant insights related to the key topics of interest are considered. By presenting the summaries the table provides unique perspectives into the application of AI in cybersecurity research. It also provides insights into the diversity of each study. This diversity enriches our knowledge and provides a well-rounded view of the application of AI in cybersecurity (Sarker et al. 2020). Furthermore, it allows for the comparison of different authors, AI techniques, aims/objectives, findings, and recommendations used in AI and cybersecurity research. This includes comparing the effectiveness of ML algorithms, feature selection methods, and evaluation metrics across various datasets and scenarios.

Studies such as Chowdhury et al. (2017) and Kozik et al. (2014) have demonstrated the effectiveness of ML algorithms, including Self-Organizing Maps (SOM), Decision Trees (DT), Support Vector Machines (SVM), and Naive Bayes, in detecting botnet and web attack activities. These algorithms excel at identifying patterns within network traffic, enabling accurate classification of normal and malicious behavior. However, their performance can vary depending on the specific dataset and type of attack (refer to Table 8). To enhance detection capabilities, researchers have also investigated the use of feature selection techniques, such as Principal Component Analysis (PCA), to reduce data dimensionality and improve algorithm efficiency. Studies by Kozik et al. (2014) and Wang and Paschalidis (2016) highlight the benefits of PCA in improving classification accuracy and computational speed.

Beyond ML, statistical methods have shown promise in detecting Distributed Denial of Service (DDoS) attacks. Hoque, Bhattacharyya, and Kalita (2016) explored the use of information metrics like entropy to identify anomalies associated with DDoS attacks. These methods offer complementary approaches to ML, providing additional insights into network traffic patterns. While individual studies have yielded promising results, the consensus among researchers is that a combination of techniques is often necessary for robust and effective intrusion detection. Hybrid approaches, incorporating multiple algorithms and feature engineering, can enhance overall performance. Additionally, the importance of real-time detection, continuous model updates, and scalability cannot be overstated.

AI techniques and datasets used in cybersecurity

This section answers the second research question (RQ2) by exploring the various AI techniques and datasets used in cybersecurity. According to Camacho (2024), the most common AI techniques used in cybersecurity include ML and DL. ML is a subfield of AI that focuses on the development of algorithms and statistical models that enable computers to learn from and

make predictions or decisions based on data (Ladosz et al. 2022). There are three major types of ML, namely, supervised learning, unsupervised learning, and reinforcement learning (RL). Supervised learning is a type of ML where the algorithm learns from labeled data, which means it is given inputs along with corresponding outputs during training measures (Sarker et al. 2020). Algorithms such as DT, SVM), and neural networks are trained on labeled data to detect known threats. Unsupervised learning is a type of ML where the algorithm learns patterns and structures from unlabeled data without human intervention (Dixit and Silakari 2021). Techniques such as clustering (e.g., k-means) and anomaly detection (e.g., Isolation Forests) are used to identify unusual patterns that may indicate novel threats. RL is a ML technique that trains algorithms by trial and error rather than using sample data (Ladosz et al. 2022). Reinforcement ML in cybersecurity has been used in several studies to identify and respond to attacks in real-time, autonomous intrusion detections, cyber-physical systems, and DDoS defenses (Moerland et al. 2023).

DL is a subset of ML that utilizes ANNs with multiple layers to learn complex patterns from data (Sarker et al. 2020). It performs learning on a multi-layer feed-forward neural network consisting of an input layer, one or more hidden layers, and an output layer (Sarker et al. 2020).

It is essential to note that high-quality datasets are necessary for training AI models effectively. Some of these datasets include KDD 99, CTU-13 and CIC-IDS -2018. These datasets and AI techniques collectively advance the field of cybersecurity by providing robust tools and data for identifying, analyzing, and mitigating various cyber threats. Hence, this section highlights the various AI techniques and datasets used in cybersecurity, and these are summarized in Table 9.

- (1) KDD 99 Dataset: This is one of the most widely used datasets for evaluating the performance of anomaly detection methods (Wang and Paschalidis 2016). It was derived from the DARPA 1998 dataset and contains a variety of simulated intrusions in a network environment. The data includes four main categories of attacks, namely, DoS, Remote to Local (R2L), User to Root (U2R), and Probing attacks are simulated (Hoque, Bhattacharyya, and Kalita 2016). Wang and Paschalidis (2016) conducted a study using the KDD 99 dataset to explore the effectiveness of Genetic Algorithms (GAs) for feature selection and intrusion detection. The study findings suggest that incorporating GAs for dimensionality reduction can lead to more accurate and efficient IDS. Their recommendations focus on adopting GA for feature selection, combining it with other algorithms, ensuring real-time application and scalability, continuously updating models, and exploring further research opportunities to maintain and improve intrusion detection. Furthermore, Hoque, Bhattacharyya, and Kalita (2016) conducted a study utilizing the KDD 99 dataset and introduced

the Feature Feature Score (FFSc), an information-theoretic metric, to enhance the analysis of multivariate data for intrusion detection. The findings of the study suggest that the FFSc can significantly improve the accuracy and efficiency of IDS by identifying and utilizing the most relevant features. Also, the study's recommendations focus on adopting the FFSc for feature selection, integrating it with ML models, ensuring continuous monitoring and updates, optimizing for real-time applications, and encouraging further research into similar metrics. Likewise, Kozik et al. (2014) applied Naive Bayes and the PCA algorithm to the KDD 99 dataset. They achieved a false positive rate with the dataset. The study findings emphasized the importance of preprocessing steps like PCA in handling high-dimensional data and recommended hybrid and ensemble approaches to further improve detection rates.

- (2) HTTP CSIC 2010 Dataset: This is a well-known dataset used for research in the field of web security and intrusion detection. It was created by the Computer Security Group (Grupo de Seguridad Informática Corporativa, CSIC) at the Universidad Carlos III de Madrid (Xie, Hu, and Slay 2014). The dataset consists of a collection of HTTP requests and responses captured over a period of time. Xie, Hu, and Slay (2014) conducted a study using the CSIC HTTP 2010 dataset to analyze web application attacks. The study employed several ML algorithms, including Naïve Bayes, Bayes Network, Decision Stump, and Radial Basis Function (RBF) Network, to detect and classify these attacks. The dataset contained a mix of benign and malicious traffic, which was essential for training and evaluating the algorithms. The study highlighted the importance of feature selection in improving model performance.
- (3) CTU-13 Dataset: The CTU-13 (Czech Technical University) dataset is a labeled dataset for botnet traffic analysis that was created by the CTU (Chowdhury et al. 2017). It is widely used for research in network security, particularly in detecting botnet activities. Chowdhury et al. (2017) used SOM to investigate the effectiveness and efficiency of detecting bots in the CTU-13 dataset. With different types of bot behavior, the proposed botnet detection method was able to detect bots with reasonable accuracy. Furthermore, the study approach ensured that bots will be found in small-sized clusters with the majority of nodes (>99%) removed from further consideration.
- (4) CSE-CIC-IDS 2018 Dataset: This is a comprehensive dataset created by a collaboration between the Communications Security Establishment (CSE) and the Canadian Institute for Cybersecurity (CIC) (Kanimozhi and Jacob 2019b). The dataset aims to provide a realistic and extensive source of network traffic data for evaluating

intrusion detection systems (IDS) and other cybersecurity solutions. CSE-CIC-IDS 2018 has the same features as the dataset created in 2017. However, more devices were used in the test environment to better model the attacks. Ferrag et al. (2020) used various DL methods, including recurrent neural networks (RNNs), deep neural networks (DNNs), restricted Boltzmann machines (RBMs), deep belief networks (DBNs), convolutional neural networks (CNNs), deep Boltzmann machines (DBMs), and deep autoencoders (DAs), on the CSE-CIC-IDS2018 and Bot-IoT datasets. They compared the classification accuracy and classification time of these DL methods. Additionally, their study examined DL-based IDS and categorized 35 attack detection datasets from the literature. The study findings reveal that these methods showed high classification accuracy across various attack types. CNNs, in particular, were effective in capturing spatial dependencies in the data. Kim, Shin, and Choi (2019) applied CNN and RNN DL methods to the CSE-CIC-IDS 2018 dataset and compared the performance of these two approaches. In their findings, both CNN and RNN models demonstrated high classification accuracy, but CNN slightly outperformed RNN in this metric. Similarly, Kanimozhi and Jacob (2019b) classified the CSE-CIC-IDS 2018 dataset using various ML methods, including ANN, RF, k-NN, SVM, ADA BOOST, and NB. The study found that Random Forest (RF) and Artificial Neural Networks (ANNs) offered the best performance in terms of accuracy and generalization. In another study, Kanimozhi and Jacob (2019a) achieved a classification accuracy of 99.97% using AI on the CSE-CIC-IDS-2018 dataset.

- (5) ISCX 2012 Dataset: This dataset addresses the need for comprehensive, real-world network traffic data to test and validate various cybersecurity solutions (Injadat et al. 2018). Injadat et al. (2018) proposed an effective framework for anomaly detection that leverages feature selection, dimensionality reduction, and ensemble learning methods. The authors used RF, SVM and k-NN algorithms in their studies. Also, the Bayesian Optimization (BO) technique was used to adjust SVM, RF and KNN parameters. The findings indicated that the framework achieved high accuracy and efficiency, making it suitable for real-time deployment in large-scale network environments.
- (6) ECML-PKDD 2007 Dataset: The dataset is primarily used for research in anomaly detection, user behavior analysis, and web usage mining (Xie, Hu, and Slay 2014). Xie, Hu, and Slay (2014) focused on some technical category that detects anomalies with a short sequence model. The ECML-PKDD 2007 dataset provided diverse examples of web application attacks, including SQL injection, cross-site scripting, and directory traversal. The findings highlight the strengths and

limitations of each algorithm, with RBF Networks and Bayes Networks showing the most promise.

- (7) ISOT Dataset: The ISOT (Information Security and Object Technology) dataset is a compilation of publicly available botnet and normal traffic datasets, encompassing a total of 1,675,424 traffic flows (Bhamare et al. 2016). The malicious traffic within the ISOT dataset was obtained from the French chapter of the Honeynet Project and includes data from the Storm and Waledac botnets. Bhamare et al. (2016) provided a comprehensive evaluation of various ML such as NNC, ANN, SVM, NBC and GBC algorithms on the ISOT dataset for cyber threat detection. Their findings highlighted the effectiveness of ensemble methods like GBC and SVM for achieving high accuracy and robustness. The study also emphasized the importance of feature selection, hyperparameter tuning, and handling imbalanced data to enhance model performance. Their recommendations serve as a valuable guide for selecting and optimizing ML algorithms in cybersecurity.
- (8) ZEUS Dataset: This dataset is often used in cybersecurity research for developing and testing IDS, malware detection algorithms, and other security-related analytics (Bhuyan, Bhattacharyya, and Kalita 2015). Bhuyan, Bhattacharyya, and Kalita (2015) conducted a comprehensive study using various datasets, including Zeus (Snort), Zeus (NETRESEC), Zeus-2 (NIMS), Conficker (CAIDA), and ISOT-Uvic, to evaluate the effectiveness of different ML algorithms for botnet detection. They utilized algorithms such as C4.5, SVM, KNN, and Bayesian Networks. The findings highlighted the strengths and weaknesses of each algorithm, with C4.5 and SVM emerging as top performers.
- (9) CIDDS-001 Dataset: The CIDDS-001 (Coburg Network Intrusion Detection Dataset) dataset was created in an anomaly-based network IDS. Scripts written in network mapper (nmap) and python language were used during the creation of the dataset (Verma and Ranga 2023). Verma and Ranga (2023) classified the CIDDS-001 dataset using KNN, SVM, DT, RF, NB, DL, ANNs, SOMs, and the Expectation Maximization (EM) algorithm. Their findings highlight the superior performance of RF and ANN, along with the importance of robust feature selection and continuous model updates. The study's recommendations emphasize adopting ensemble and DL models, enhancing feature selection, continuous monitoring, optimizing for real-time detection, and conducting further research into advanced models to maintain and improve IDSs.
- (10) CAIDA DDoS 2007 Dataset: The study by Kato and Klyuev (2014) investigated the application of statistical methods to analyze and detect

Table 9. AI techniques and datasets.

Rank	Authors	Dataset used	Techniques	Problem Domain	Evaluation Metrics	Feature Selection
1	Chowdhury et al. (2017)	CTU-13	SOM	DDoS attack detection	Accuracy	Source IP address, destination IP address and protocol-specific features
2	Bhuyan, Bhattacharyya, and Kalita (2015)	Zeus (Snort), Zeus (NETRESEC), Zeus-2 (NIMS), Conficker (CAIDA) and ISOT-Uvic	C4.5, SVM, KNN, Bayesian Networks	Botnet detection	Detection rate(DR), false positive rate	Source IP, destination IP, and basic features (protocols connection)
3	Bhamare et al. (2016)	ISOT	NNC, ANN, SVM, NBC, GBC	Botnet detection	True detection rate, error rate	Network traffic, statistical protocol and time interval
4	Kozik et al. (2014)	KDD Cup 1999	Naïve Bayes, PCA algorithm	Intrusion detection	False positive rate	1) Basic features, (2) Traffic features, and (3) Content features. The basic features are extracted from a TCP/IP connection. The traffic features are divided into two groups (i.e., “same host” features, and “same service” features). The content features concerns suspicious behavior in the data portion
5	Kato and Kiyuev (2014)	CAIDA, DDoS 2007, MIT DARPA	Statistical method	DDoS attack detection	Accuracy	IP address, destination IP address, time interval in seconds between packets, and packet size in bytes from the database
6	Xie, Hu, and Slay (2014)	ECML-PKDD 2007 hTTP, CSIC HTTP 2010	Naïve Bayes, Bayes Network, Decision Stump, RBF network	Web applications attack	False positive rate	N/A
7	Moustafa and Slay (2016)	UNSW-NB15	k-means	Network anomaly detection	Accuracy, DR, FPR	Flow features (e.g., client-to-serve or server-to-client); Basic features (protocols connections); Content features (attributes of TCP/IP, attributes of http services); and Time features (arrival time between packets, start/end packet time, and round-trip time of TCP protocol)
8	Hoque, Bhattacharyya, and Kalita (2016)	KDD 99, CAIDA, TUIDS DDoS	Information metrics	DDoS attack detection	N/A	N/A
9	Wang and Paschalidis (2016)	KDD 99	Genetic algorithm	Intrusion detection	DR	1) The cluster label of the source IP address; 2) the cluster label of the destination IP address; 3) the source port number; 4) the destination port number; 5) the flow duration; 6) the data bytes sent from source to destination; and 7) the data bytes sent from destination to source

(Continued)

Table 9. (Continued).

Rank	Authors	Dataset used	Techniques	Problem Domain	Evaluation Metrics	Network flow features	Feature Selection
10	Ferrag et al. (2020)	CSE-CIC IDS2018	Deep discriminative models. (DNN, RNN, CNN)	Intrusion detection	Accuracy	Network flow features	
11	Injadat et al. (2018)	ISCX2012	SVM, KNN, RF	Anomaly detection	Accuracy, precision, recall, F1-score	Bayesian optimization	
12	Kanimozhi and Jacob (2019b)	CSE-CIC IDS2018	ANN, RF, k-NN, SVM, Adaboost, NB	Network anomaly detection	Accuracy, precision, recall, F1-score	General information, quality of data, data volume, recording environment, and evaluation.	
13	Verma and Ranga (2023)	CIDDS-001	KNN, SVM, DT, RF, NB, DL, ANN, SOMs, EM, k-means	Network intrusion detection	Accuracy	Binary feature encoding	
14	Kim, Shin, and Choi (2019)	CSE-CIC IDS2018	CNN and RNN	Intrusion detection	Accuracy	Traffic features	
15	Moustafa and Slay (2016)	UNSW-NB15 KDD99	DT, LR, NB, ANN, EM clustering	Anomaly detection	Accuracy, FAR	Flow features, basic features, content features, time features	
16	Kanimozhi and Jacob (2019a)	CSE-CIC IDS2018	RF and ANN	Intrusion detection	Accuracy, precision, recall	Network flow features	

DDoS attacks using the CAIDA DDoS 2007 and MIT DARPA datasets. The study highlighted the robustness and scalability of these methods and provided valuable insights into feature selection and anomaly detection. The recommendations focused on real-time implementation, dynamic feature selection, reducing false positives, optimizing performance, and continuous monitoring to enhance the detection and mitigation of DDoS attacks in real-world network environments.

- (11) UNSW-NB15 Dataset: This dataset was created by the Australian Centre for Cybersecurity (ACCS). During its creation, tools such as IXIA PerfectStorm, Tcpdump, Argus, and Bro-IDS were used (Moustafa and Slay 2016). The IXIA tool, which serves as a generator for both normal and abnormal traffic, was deployed on three virtual servers. Moustafa and Slay (2016) examined the complexity of the UNSW-NB15 dataset in their study. For this purpose, in the first step, statistical analysis of qualifications was explained; in the second step, feature correlations were examined; and, in the last step, the performance of the dataset with five classifiers was measured and compared with the KDD99 dataset. The UNSW-NB15 dataset was observed to be more complex than KDD99 dataset.

AI tools and techniques in cybersecurity

The third research question (RQ3) concerned the AI tools used for data extraction, analysis, and optimization. The application domain in Table 10 relies on various types of data and tools to monitor and protect network and system security. For example, IDSs analyze the Network Traffic Data, Log Data or Behavioral Data using AI tools such as SQL, Python or R to identify malicious activities (Sultana and Jilani 2021). Likewise, Imaging and CAPTCHA systems involve processing visual data (image or text data) using AI tools such as Python or R to either display images or distinguish human users from bots (Dinh and Hoang 2023). By leveraging these data types and tools, organizations can effectively process and analyze images, generate robust CAPTCHA systems, and ensure accurate differentiation between human users and automated bots (Challagundla, Reddy Gogireddy, and Reddy Peddavenkatagari 2024). Table 10 presents a summary of the major domains where AI tools and techniques have been utilized for data extraction, analysis, and optimization in cybersecurity models for various purposes.

Intrusion detection systems

SVM and KNN are commonly used in IDS for classification tasks (Sultana and Jilani 2021). The algorithms learn patterns from labeled data to classify network traffic as normal or malicious. SVM aims to find a hyperplane that separates different classes of data, while KNN classifies data based on the

Table 10. Application Domain and usage of AI tools and techniques in cybersecurity.

Application domain	Techniques	AI Tools	Purpose	Authors
Intrusion Detection Systems	SVM and KNN	Python Libraries R Libraries	To build an IDS Feature selection, intrusion detection and classification. To classify various attacks such as DoS, Probe, U2R, and R2L DDoS detection and analysis in SDN-based environment Evaluating host-based anomaly detection systems. To develop an IDS To reduce the false alarm rate.	Sultana and Jilani (2021), Veena et al. (2022),
	K-means and KNN	SQL, Python Libraries, R Libraries		Saheed, Arowolo, and Tosho (2022)
	Naive Bayes	Python Libraries, R Libraries	To build an IDS for multi-class classification.	Yilmaz, Taspinar, and Koklu (2022), Rekha et al. (2020)
	DTs	R Libraries Python Libraries SQL Apache spark MLlib	To detect the malicious code's behavior information by running malicious code on the virtual machine and analyzing the behavior information for intrusion detection.	Ferdiana (2020), Kumari and Mehta (2020), Melvin et al. (2022)
	KNN and Clustering	SQL R Libraries Python Libraries	To develop an IDS.	Bohara et al. (2020)
	ANN and DT	Big data tools (Apache spark MLlib) Python Libraries Specialized tools (WEKA)	To measure the performance of an IDS.	Saranya et al. (2020)
	Adaptive Boosting (AdaBoost)	R Libraries Python Libraries	To improve the performance of classification algorithms.	Divakar et al. (2021)

(Continued)

Table 10. (Continued).

Application domain	Techniques	AI Tools	Purpose	Authors
Network Intrusion Detection System	DTs	R Libraries Python Libraries SQL Apache spark MLlib	Used for feature selection and to build an effective network IDS	Guezaz et al. (2021)
	RF	R Libraries Python Libraries	To build network IDS	Divakar et al. (2021)
	CNN	Python Libraries R libraries	To detect intrusions in network traffic To extract relevant features and identify complex patterns.	Kim et al. (2020), Mohammadpour et al. (2022)
Imaging and CAPTCHA	SVM	Python Libraries R libraries	To classify distorted characters or objects present in CAPTCHA images.	Dinh and Hoang (2023) Kumar, Jindal, and Kumar (2022)
	CNN	Python Libraries R libraries	Leverages convolutional layers to extract hierarchical features from images. To capture spatial relationships and patterns effectively.	Sachdev (2020) Challagundla, Reddy Gogireddy, and Reddy Peddavenkatagari (2024) Wang, Shi, and Uddin (2021)
	Singular Value Decomposition (SVD)	Python Libraries, R Libraries MATLAB	To extract important features from images.	Kaur and Jindal (2020) Ranjan, Patidar, and Kushwaha (2020).
Phishing/malware detection	ANN and CNN	Apache spark MLlib R Libraries Python Libraries	To learn patterns and features from large datasets of phishing emails. To capture spatial relationships and patterns in images.	Soon et al. (2020) Verma et al. (2019) Hassan and Fakharudin (2023).
	SVM	R Libraries Python Libraries R Libraries	To classify instances into phishing or legitimate categories based on features extracted from URLs, email headers, or network traffic.	Anupam and Kumar Kar (2021)
Traffic classification	Q-learning	Python Libraries R Libraries	To detect malicious content without hampering the critical attributes.	Gill et al. (2021)
	K-means clustering	Python Libraries Python Libraries SQL R Libraries	To analyze clusters, identify patterns and distinguish between different classes of traffic.	Kamal et al. (2024) Liao and Li (2022) Jain, Kaur, and Saxena (2022)
	CNN	Python Libraries R libraries	To classify traffic based on its content, enabling fine-grained classification of applications or protocols.	Salman et al. (2021)

(Continued)

Table 10. (Continued).

Application domain	Techniques	AI Tools	Purpose	Authors
Anomaly and DoS detection	SVM	Python Libraries R libraries	To classify network traffic based on features such as packet size, frequency, and protocol type. To find the hyperplane that best separates normal from abnormal instances in a high-dimensional feature space.	Bhati and Shekhar Rai (2020) Abuali, Nissirat, and Al-Samawi (2023)
	DT and KNN	R Libraries Python Libraries SQL	To classify incoming traffic based on the majority class of its nearest neighbours in the feature space. To detect deviations from expected patterns or rules learned from training data.	Alharbi et al. (2021) Ramadhan, Sukarno, and Nugroho (2020)
	PCA	Python Libraries R Libraries MATLAB	To identify relevant features and reduce noise, making it easier to detect anomalies in the data.	Divakar et al. (2021)

majority class among its nearest neighbors (Veena et al. 2022). Naive Bayes is a probabilistic classifier that assumes independence among features (Yilmaz, Taspinar, and Koklu 2022). In IDS, Naive Bayes is used to classify network traffic by computing the probability of a packet being normal or malicious based on its attributes (Rekha et al. 2020). DTs are used in IDS to model the decision-making process based on features extracted from network traffic (Melvin et al. 2022). They recursively partition the feature space based on attribute values, leading to a tree-like structure (Kumari and Mehta 2020). K-Means is a clustering algorithm used in IDS for unsupervised learning tasks (Saheed, Arowolo, and Tosho 2022). It groups network traffic data into clusters based on similarities in feature space. K-Means clustering helps in identifying anomalies and detecting new types of attacks by finding patterns in unlabeled data (Bohara et al. 2020). In IDS, ANN is used for both supervised and unsupervised learning tasks. It learns complex patterns from network traffic data to detect intrusions by adjusting the weights and biases of interconnected neurons (Saranya et al. 2020). Likewise, in IDS, AdaBoost is used to improve the performance of classification algorithms by sequentially training classifiers on different subsets of the training data (Divakar et al. 2021). It focuses on misclassified instances, thus enhancing the overall accuracy of intrusion detection.

Network intrusion detection system (NIDS)

DTs are commonly used in NIDS to classify network traffic based on pre-defined rules learned from historical data (Guezzaz et al. 2021). DTs help in identifying patterns and rules indicative of normal or malicious network behavior (Ferdiana 2020). In NIDS, RF is used to handle large volumes of network traffic data and increase the robustness of intrusion detection by aggregating the predictions of multiple DTs (Divakar et al. 2021). However, convolutional neural networks (CNNs) are DL models commonly used in NIDS for detecting intrusions in network traffic, especially in the context of analyzing packet payloads or network traffic images (Kim et al. 2020). CNNs learn hierarchical representations of features in network data, enabling them to automatically extract relevant features and identify complex patterns associated with different types of network attacks (Mohammadpour et al. 2022).

Imaging and CAPTCHA

SVMs have been used to efficiently separate different classes of images by finding the optimal hyperplane that maximally separates them in feature space (Dinh and Hoang 2023). A recent study by Sachdev (2020), reveals that Wei et al. (2019) developed an SVM-based model to classify segmented characters from CAPTCHA, achieving 99% accuracy. SVMs have been applied in image classification tasks, including CAPTCHA recognition (Kumar, Jindal, and Kumar 2022). In CAPTCHA recognition, SVMs have been used to classify

distorted characters or objects present in CAPTCHA images. Likewise, CNNs are widely used for image recognition and classification tasks, including CAPTCHA recognition. CNNs leverage convolutional layers to extract hierarchical features from images, allowing them to learn complex patterns and variations in CAPTCHA images (Challagundla, Reddy Gogireddy, and Reddy Peddavenkatagari 2024). Compared to traditional methods, the main advantage of CNNs lies in their convolutional layers, where the extracted image features have a strong expressive capability (Wang, Shi, and Uddin 2021). This approach avoids the issues of data preprocessing and manually designed features found in traditional recognition techniques. Furthermore, CNNs have demonstrated superior performance in CAPTCHA recognition due to their ability to capture spatial relationships and patterns effectively. Similarly, Singular Value Decomposition (SVD) has been used in image compression and reconstruction tasks, which can be relevant to CAPTCHA generation and recognition (Ranjan, Patidar, and Kushwaha 2020). By decomposing images into their singular values and corresponding matrices, SVD enables efficient representation and reconstruction of images. In a recent study, Kaur and Jindal (2020) discussed image authentication techniques using SVD. This method was employed to extract important features from images.

Phishing/malware detection

ANN and CNN are two of the most crucial neural networks used for detecting phishing and malware by learning patterns and features from large datasets of phishing e-mails, URLs, or malware samples (Soon et al. 2020). A study by Verma et al. (2019) utilized deep belief networks and ANNs. The ANN achieved 89.95% accuracy with five hidden layers and five nodes per layer, while the deep belief network achieved 96.32% accuracy using similar settings (Hassan and Fakharudin 2023). CNNs can capture spatial relationships and patterns in images, enabling them to detect phishing website layouts or malware signatures effectively. Likewise, SVMs and CNNs have been used for phishing and malware detection by classifying instances into phishing or legitimate categories based on features extracted from URLs, e-mail headers, or network traffic (Anupam and Kumar Kar 2021). Similarly, Q-Learning has been efficient in yielding desirable accuracy for the recognition of malicious content without hampering the critical attributes (Gill et al. 2021). Extensive studies on different and large-scale phishing datasets indicate the usefulness of the Q-Learning-based RL technique in outperforming standard ML models in detection performance (Kamal et al. 2024).

Traffic classification

K-means clustering has been used in traffic classification for unsupervised learning tasks (Liao and Li 2022). The K-means clustering process generates cluster centroids for normal and anomalous traffic, which can then be used to

detect anomalies in new flow records monitored within the same network (Jain, Kaur, and Saxena 2022). Similarly, a recent study has shown that CNN is being utilized in classifying traffic based on its content, enabling fine-grained classification of applications or protocols (Salman et al. 2021). Compared with the traditional classification method, CNN traffic classification can improve the accuracy and reduce the time of classification (Salman et al. 2021).

Anomaly and DoS detection

SVM is one of the most successful techniques in making classifications of intrusive behaviors (Bhati and Shekhar Rai 2020). SVMs classify network traffic based on features such as packet size, frequency, and protocol type (Abuali, Nissirat, and Al-Samawi 2023). They aim to find the hyperplane that best separates normal from abnormal traffic in a high-dimensional feature space. However, Alharbi et al. (2021), claim that the KNN algorithm with fast response capability does not need to train the classifier before use, so this algorithm can be better used for DoS intrusion detection. KNN detects anomalies by comparing new instances to historical data and identifying deviations from normal behavior. Likewise, DTs recursively partition the feature space based on attribute values, leading to a tree-like structure (Ramadhan, Sukarno, and Nugroho 2020). DT identifies anomalies by detecting deviations from expected patterns or rules learned from training data (Ramadhan, Sukarno, and Nugroho 2020). Similarly, PCA is used in anomaly detection for dimensionality reduction and feature extraction (Divakar et al. 2021). It transforms high-dimensional data into a lower-dimensional space while preserving most of the variance. PCA helps in identifying relevant features and reducing noise, making it easier to detect anomalies in the data.

Future directions for the application of AI in cybersecurity

RQ4 investigated the future research directions for the application of AI in cybersecurity. As explained earlier, trending topics in AI and cybersecurity issues have continued to evolve (refer to Figure 6). The order of trending topics indicates shifting priorities within the tech community. While ML is currently at the forefront, cybersecurity and AI are also crucial areas of interest. For instance, ML techniques are increasingly being used to enhance cybersecurity measures. This implies that there is a significant amount of innovation, investment, and application development happening in this field (Sarker et al. 2020). In addition, the thematic map (refer to Figure 4) shows that there is a heavy concentration of basic themes in the application of AI in cybersecurity research, which are central themes in need of more development. Also, many of the themes can be characterized as niche themes, which means that they are specialized or peripheral and in need of a stronger connection to the broader AI in cybersecurity literature. However, in Figure 4,

there are no themes/clusters in the lower left quadrant, which suggests that there are no currently identifiable new or fading trends in the dataset analyzed. This could also indicate a period of stability in the research field (Foody 2020). Moreover, the absence of motor themes in the upper right quadrant suggests a recent shift in focus, where previously dominant themes have either become too broad to be categorized as a single theme or the research community is in transition, looking for new directions (Tennekes 2018). It might also imply limitations in the dataset as the parameter for this study dataset was readjusted from the default setting of 250 for the number of words to 100 (Choudhri et al. 2015). As noted, the readjustment was necessary because the default setting produced overlapping clusters that were muddled together, making it difficult to see and interpret. Future research may readjust the parameters for the datasets to identify emerging or declining patterns as well as motor themes. Since the literature on the application of AI in cybersecurity is growing rapidly, it is likely that some of the current basic and niche themes will move to the upper right quadrant and become motor themes (Madsen, Berg, and Nardo 2023).

Likewise, the thematic evolution (refer to Figure 5) highlights two phases, namely, 2014–2019 (pre-COVID-19 pandemic) and 2020–2024 (during and post-COVID-19 pandemic). In the first phase, the dominant themes were AI, Classification, IoT, Security, ML, Cybersecurity and Cyberattack. This implies that there is an emphasis on foundational technologies (AI, ML) and their applications (IoT, cybersecurity) in an increasingly digital world (Jallouli et al. 2019). In the second phase, which was during and post-COVID-19 pandemic (2020–2024), the dominant themes are ML, XAI and Data mining. This implies that there is a greater focus on enhancing the interpretability and trustworthiness of AI (XAI) and leveraging data to address unprecedented global challenges (Oropeza-Valdez et al. 2024). This suggests that the future directions for the application of AI in cybersecurity involve leveraging advanced AI technologies such as ML to enhance security measures, detect threats, and respond to incidents more effectively (Holzinger 2018). There are emerging algorithms and practical applications that are driving the interest in ML (Sarker, Furhad, and Nowrozy 2021). Likewise, cybersecurity remains a critical concern, due to the increasing number of cyber threats, data breaches, and the need for secure systems as more services go digital. Although there are ongoing efforts to develop better security protocols, tools, and strategies to protect information and systems from malicious activities (Oropeza-Valdez et al. 2024), future research should focus on optimizing AI models for effective deployment across various cybersecurity environments. Also, based on the bibliometric analysis (refer to Figure 6), there is a high demand for expertise in ML and AI, indicating that professionals with skills in this area are in high demand (Verma et al. 2019). Moreover, AI continues to be a major area of focus due to its wide range of applications,

from autonomous vehicles and robotics to natural language processing and decision-making systems (Marr 2019). There is continuous innovation and development in AI technologies, which keeps it a “hot topic” among researchers, developers, and businesses. AI has the potential to transform various industries and aspects of daily life, leading to significant interest in understanding and advancing this technology (Verma et al. 2022). Hence, future research could focus on improving the interpretability and transparency of AI algorithms. This could help build trust in AI-based cybersecurity systems and facilitate better decision-making.

Discussion

The findings of this research reveal both promising developments and challenges in the application of AI within the cybersecurity domain. As AI technologies continue to evolve, they are playing an increasingly critical role in addressing sophisticated cyber threats. However, the analysis also highlights several gaps and underexplored areas in existing literature, which have implications for both researchers and practitioners.

Basic versus niche themes: a field in transition

The thematic map (Figure 4) reveals that many topics within AI and cybersecurity are still categorized as basic or niche themes, suggesting that the field is in a developmental phase. This aligns with Madsen, Berg, and Nardo (2023), who argue that niche themes can indicate specialized areas that need to be further integrated into the broader research landscape. Over time, some of these basic themes (e.g., ML techniques and XAI) are likely to evolve into motor themes, reflecting their growing importance. The lack of motor themes and new clusters identified in the dataset may suggest a period of stability in the field (Foody 2020). However, it could also imply limitations in dataset parameters used during the bibliometric analysis. Future research will need to explore whether these trends represent true stability or a transition toward emerging priorities that are not yet fully captured in current datasets. Also, future research can focus on identifying new research themes by adjusting dataset parameters to better capture emerging trends. Also, exploring how foundational themes (like AI and ML) evolve into more specific areas (e.g., XAI, Data Mining).

Evolving priorities in AI and cybersecurity: shifts in themes and trends

The thematic evolution (Figure 5) illustrates a noticeable shift in research focus between two key phases: pre-COVID-19 (2014–2019) and during/post-COVID-19 (2020–2024). Foundational topics, such as AI, ML, cybersecurity,

and Internet of Things (IoT), dominated earlier research, reflecting the growing dependence on digital technologies and the need for advanced security solutions (Jallouli et al. 2019). The second phase emphasizes XAI and data mining, indicating an emerging need for greater interpretability of AI models. This is aligned with growing concerns about the “black-box” nature of many AI systems, particularly in critical fields like cybersecurity, where trust, transparency, and accountability are essential (Oropeza-Valdez et al. 2024). These shifts in thematic focus highlight the increasing importance of developing AI models that not only detect threats effectively but also provide understandable insights that human operators can act on confidently. Future research should focus on developing XAI models that can clearly explain how decisions are made without compromising performance. This will help build trust among stakeholders and ensure that AI systems are used responsibly and ethically in cybersecurity applications.

Skill gaps and workforce demand: bridging the knowledge divide

The bibliometric analysis (Figure 6) underscores a high demand for AI and cybersecurity expertise, which aligns with Verma et al. (2022). Organizations increasingly require professionals with skills in ML, data science, and cybersecurity, as the integration of AI into security frameworks becomes more widespread. However, the rapid growth of the field has outpaced the development of structured education and training programs, leading to skill shortages. To address this gap, future research could explore interdisciplinary training programs that combine AI and cybersecurity, equipping professionals with the necessary knowledge to manage AI-powered security solutions effectively. In addition, collaborations between academia and industry can foster the development of relevant curricula and hands-on training opportunities.

Trust and transparency: a key challenge for AI in cybersecurity

The findings indicate that XAI has become a focal point in recent years, reflecting the need for transparency in AI-based cybersecurity systems. As AI takes on more responsibilities in threat detection and response, the ability to explain decisions becomes essential for both regulatory compliance and end-user trust (Oropeza-Valdez et al. 2024). This aligns with industry-wide efforts to improve the interpretability of AI models, especially in critical sectors like finance and healthcare, where AI-driven decisions can have significant consequences. Future research should focus on developing XAI models that can clearly explain how decisions are made without compromising performance. This will help build trust among stakeholders and ensure that AI systems are used responsibly and ethically in cybersecurity applications.

Metadata quality

The findings indicate that the high quality of most metadata fields such as abstracts, authors, and titles ensure reliable core research analysis. However, gaps in document identifiers, reprint authors, and cited references present challenges that limit the study's depth particularly in citation analysis and network mapping. While incomplete metadata in areas like keywords and affiliations may slightly affect thematic analyses, the impact on overall trend detection remains minimal due to the availability of other key metadata fields. In addition, the thematic evolution and research areas are primarily captured through the combination of keywords and article abstracts, reducing the dependency on secondary data fields like document identifiers (Donthu et al. 2021). Nevertheless, future research should focus on addressing the gaps identified in the metadata to enhance the depth and scope of analysis.

Limitation of the study

As with all research, validating the methods used is crucial. The primary threat to the validity of this study is the potential omission of relevant papers during the selection process and biases in data extraction. The study relied solely on Scopus and WoS databases because it offers comprehensive access to a wide range of research resources and also have studies from other databases like IEEE, Springer Nature and Elsevier. However, this study may have excluded other databases such as Google Scholar which may have added more insight to the study. Nevertheless, the study provides insights into patterns within the domain, especially since the initial search indicated that the Scopus and WoS digital libraries contain the majority of the necessary papers. Also, the search string used in this study may not have captured all relevant terms, synonyms, and variations related to AI and cybersecurity, potentially leaving out pertinent literature. Similarly, none of the 939 records analyzed included information about cited references, limiting the ability to conduct a comprehensive citation analysis and assess the interconnectedness of research works. Likewise, due to the timeframe specific (2014–2024), the review may not adequately address emerging trends in AI and cybersecurity beyond the specified time frame, which could limit its relevance in a rapidly evolving field. Hence, future research should periodically revisit and update the review to incorporate emerging trends in AI and cybersecurity beyond the specified time frame. This would provide a more dynamic and up-to-date understanding of the field. In addition, future research should consider expanding the search string to include additional terms, variations, and synonyms relevant to AI and cybersecurity to ensure a more comprehensive literature search. Moreover, future research should consider using tools or databases that provide access to citation data to conduct citation analyses, helping to identify

key works and trends within the field. It is also essential that future research should broaden database coverage to include other databases to capture a more diverse range of relevant literature.

Conclusion

This comprehensive review of AI's application in cybersecurity highlights its significant potential to enhance threat detection, response, and overall security posture. While significant progress has been made in leveraging AI for cybersecurity (Khan, Malik, and Nazir 2024), future research must focus on optimizing specific AI techniques, improving algorithm interpretability and transparency, and addressing the challenges of deployment in diverse environments. By continuing to advance AI methodologies and their applications, the cybersecurity field can achieve greater resilience and adaptability against evolving threats.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Lizzy Ofusori  <http://orcid.org/0000-0002-6036-619X>

Tebogo Bokaba  <http://orcid.org/0000-0003-3710-2513>

Siyabonga Mhlongo  <http://orcid.org/0000-0001-8203-5984>

Data availability statement

The authors confirm that all data generated or analyzed during this study are included in this published article.

References

- Abuali, K., L. Nissirat, and A. Al-Samawi. 2023. Advancing network security with AI: SVM-Based deep learning for intrusion detection. *Sensors (Switzerland)* 23 (21):8959. doi: [10.3390/s23218959](https://doi.org/10.3390/s23218959).
- Aflalo, A., S. Bagon, T. Kashti, and Y. Eldar. 2023. Deepcut: Unsupervised segmentation using graph neural networks clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Paris, France, 32–41. IEEE. doi: [10.48550/arXiv.2212.05853](https://doi.org/10.48550/arXiv.2212.05853).
- Albhirat, M., A. Rashid, R. Rasheed, S. Rasool, S. Zulkiffli, and H. Muhammad Zia-Ul-Haq. 2024. The PRISMA statement in enviropreneurship study: A systematic literature and a research agenda. *Cleaner Engineering and Technology* 18 (February):100721. Elsevier. doi: [10.1016/j.clet.2024.100721](https://doi.org/10.1016/j.clet.2024.100721).

- Alharbi, Y., A. Alferaidi, K. Yadav, G. Dhiman, S. Kautish, and J. Xia. 2021. Denial-of-service attack detection over IPv6 network based on KNN algorithm. *Wireless Communications and Mobile Computing* 2021 (1):1–6. doi: [10.1155/2021/8000869](https://doi.org/10.1155/2021/8000869).
- Anupam, S., and A. Kumar Kar. 2021. Phishing website detection using support vector machines and nature-inspired optimization algorithms. *Telecommunication Systems* 76 (1):17–32. doi: [10.1007/s11235-020-00739-w](https://doi.org/10.1007/s11235-020-00739-w).
- Bertero, C., M. Roy, C. Sauvanoud, and G. Trédan. 2017. Experience report: Log mining using natural language processing and application to anomaly detection. In *2017 IEEE 28th International Symposium on Software Reliability Engineering (ISSRE)*, Toulouse, France, 351–60. IEEE. doi: [10.1109/ISSRE.2017.43](https://doi.org/10.1109/ISSRE.2017.43).
- Bhamare, D., T. Salman, M. Samaka, A. Erbad, and R. Jain. 2016. Feasibility of supervised machine learning for cloud security. IN *2016 International Conference on Information Science and Security (ICISS)*, Pattaya, Thailand, 1–5. IEEE. doi: [10.1109/ICISSEC.2016.7885853](https://doi.org/10.1109/ICISSEC.2016.7885853).
- Bhati, B., and C. Shekhar Rai. 2020. Analysis of support vector machine-based intrusion detection techniques. *Arabian Journal for Science & Engineering* 45 (4):2371–83. doi: [10.1007/s13369-019-03970-z](https://doi.org/10.1007/s13369-019-03970-z).
- Bhuyan, M., D. Bhattacharyya, and J. Kalita. 2015. An empirical evaluation of information metrics for low-rate and high-rate DDoS attack detection. *Pattern Recognition Letters* 51:1–7. doi: [10.1016/j.patrec.2014.07.019](https://doi.org/10.1016/j.patrec.2014.07.019).
- Bohara, B., J. Bhuyan, F. Wu, and J. Ding. 2020. A survey on the use of data clustering for intrusion detection system in cybersecurity. *International Journal of Network Security & Its Applications* 12 (1):1. doi: [10.5121/ijnsa.2020.12101](https://doi.org/10.5121/ijnsa.2020.12101).
- Camacho, N. 2024. The role of AI in cybersecurity: Addressing threats in the digital age. *Journal of Artificial Intelligence General Science (JAIGS)* ISSN: 3006-4023 3 (1):143–54. doi: [10.60087/jaigs.v3i1.75](https://doi.org/10.60087/jaigs.v3i1.75).
- Challagundla, B., Y. Reddy Gogireddy, and C. Reddy Peddavenkatagari. 2024. Efficient CAPTCHA image recognition using convolutional neural networks and long short-term memory networks. *International Journal of Scientific Research in Engineering and Management (IJSREM)* 8 (3):1–5. doi: [10.55041/IJSREM29450](https://doi.org/10.55041/IJSREM29450).
- Charbuty, B., and A. Abdulazeez. 2021. Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends* 2 (1):20–28. doi: [10.38094/jastt20165](https://doi.org/10.38094/jastt20165).
- Choudhri, A., A. Siddiqui, N. Khan, and H. Cohen. 2015. Understanding bibliometric parameters and analysis. *Radiographics* 35 (3):736–46. doi: [10.1148/rg.2015140036](https://doi.org/10.1148/rg.2015140036).
- Chowdhury, S., M. Khanzadeh, R. Akula, F. Zhang, S. Zhang, H. Medal, M. Marufuzzaman, and L. Bian. 2017. Botnet detection using graph-based feature clustering. *Journal of Big Data* 4 (1):1–23. doi: [10.1186/s40537-017-0074-7](https://doi.org/10.1186/s40537-017-0074-7).
- Cisco. 2022. Cybersecurity resilience emerges as top priority as 62 percent of companies say security incidents impacted business operations. Accessed June 13, 2024. <https://investor.cisco.com/news/news-details/2022/Cybersecurity-resilience-emerges-as-top-priority-as-62-percent-of-companies-say-security-incidents-impacted-business-operations/default.aspx>.
- Cobo, M., A. Gabriel López-Herrera, E. Herrera-Viedma, and F. Herrera. 2011. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics* 5 (1):146–66. doi: [10.1016/j.joi.2010.10.002](https://doi.org/10.1016/j.joi.2010.10.002).
- Cremer, F., B. Sheehan, M. Fortmann, A. Kia, M. Mullins, F. Murphy, and S. Materne. 2022. Cyber risk and cybersecurity: A systematic review of data availability. *The Geneva Papers on Risk and Insurance-Issues and Practice* 47 (3):698–736. doi: [10.1057/s41288-022-00266-6](https://doi.org/10.1057/s41288-022-00266-6).

- Dasgupta, D., Z. Akhtar, and S. Sen. 2022. Machine learning in cybersecurity: A comprehensive survey. *The Journal of Defense Modeling and Simulation* 19 (1):57–106. doi: [10.1177/15485129209512](https://doi.org/10.1177/15485129209512).
- Dawson, M., R. Baciuc, L. B. Gouveia, and A. Vassilakos. 2021. Understanding the challenge of cybersecurity in critical infrastructure sectors. *Land Forces Academy Review* 26 (1):69–75. doi: [10.2478/raft-2021-0011](https://doi.org/10.2478/raft-2021-0011).
- Dinh, N. T., and V. T. Hoang. 2023. Recent advances of captcha security analysis: A short literature review. *Procedia Computer Science* 218 :2550–62. doi: [10.1016/j.procs.2023.01.229](https://doi.org/10.1016/j.procs.2023.01.229).
- Divakar, S., R. Priyadarshini, R. Kumar Barik, and D. Sinha Roy. 2021. An intelligent intrusion detection scheme powered by boosting algorithm. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 205–09. doi: [10.1109/Confluence51648.2021.9377076](https://doi.org/10.1109/Confluence51648.2021.9377076).
- Dixit, P., and S. Silakari. 2021. Deep learning algorithms for cybersecurity applications: A technological and status review. *Computer Science Review* 39 :100317. doi: [10.1016/j.cosrev.2020.100317](https://doi.org/10.1016/j.cosrev.2020.100317).
- Dong, S., P. Wang, and K. Abbas. 2021. A survey on deep learning and its applications. *Computer Science Review* 40:100379. doi: [10.1016/j.cosrev.2021.100379](https://doi.org/10.1016/j.cosrev.2021.100379).
- Donthu, N., S. Kumar, D. Mukherjee, N. Pandey, and W. Marc Lim. 2021. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research* 133:285–96. doi: [10.1016/j.jbusres.2021.04.070](https://doi.org/10.1016/j.jbusres.2021.04.070).
- Ferdiana, R. 2020. A systematic literature review of intrusion detection system for network security: Research trends, datasets and methods. In *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, Semarang, Indonesia, 1–6. IEEE. doi: [10.1109/ICICoS51170.2020.9299068](https://doi.org/10.1109/ICICoS51170.2020.9299068).
- Ferrag, M. A., L. Maglaras, S. Moschogiannis, and H. Janicke. 2020. Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security & Applications* 50:102419. doi: [10.1016/j.jisa.2019.102419](https://doi.org/10.1016/j.jisa.2019.102419).
- Foody, G. M. 2020. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. *Remote Sensing of Environment* 239:111630. doi: [10.1016/j.rse.2019.111630](https://doi.org/10.1016/j.rse.2019.111630).
- Gill, S., S. Gogte, C. Sharma, P. Pathwar, V. Desai, A. Gupta, A. Vyas, and O. P. Vyas. 2021. Exploring deep reinforcement learning for android malware detection. *EasyChair Preprint* 6594:1–8.
- Guezaz, A., S. Benkirane, M. Azrour, and S. Khurram. 2021. A reliable network intrusion detection approach using decision tree with enhanced data quality. *Security and Communication Networks* 2021 (8):1230593. doi: [10.1155/2021/1230593](https://doi.org/10.1155/2021/1230593).
- Hassan, N. H., and A. S. Fakharudin. 2023. Web phishing classification model using artificial neural network and deep learning neural network. *International Journal of Advanced Computer Science & Applications* 14 (7):535–42. doi: [10.14569/ijacsa.2023.0140759](https://doi.org/10.14569/ijacsa.2023.0140759).
- Hofstetter, M., R. Riedl, T. Gees, A. Koumpis, and T. Schaberreiter. 2020. Applications of AI in cybersecurity. In *2020 second International Conference on Transdisciplinary AI (TransAI)*, Irvine, CA, USA, 138–41. IEEE. doi: [10.1109/TransAI49837.2020.00031](https://doi.org/10.1109/TransAI49837.2020.00031).
- Holzinger, A. 2018. From machine learning to explainable AI. In *2018 world symposium on digital intelligence for systems and machines (DISA)*, Košice, Slovakia, 55–66. IEEE. doi: [10.1109/DISA.2018.8490530](https://doi.org/10.1109/DISA.2018.8490530).
- Hoque, N., D. K. Bhattacharyya, and J. K. Kalita. 2016. A novel measure for low-rate and high-rate DDoS attack detection using multivariate data analysis. In *2016 8th International Conference on Communication Systems and Networks (COMSNETS)*, Bangalore, India, 1–2. IEEE. doi: [10.1109/COMSNETS.2016.7439939](https://doi.org/10.1109/COMSNETS.2016.7439939).

- Huberman, A. **2014**. *Qualitative data analysis a methods sourcebook*. Thousand Oaks, CA: Sage Publications.
- Injadat, M., F. Salo, A. B. Nassif, A. Essex, and A. Shami. **2018**. Bayesian optimization with machine learning algorithms towards anomaly detection. In *2018 IEEE global communications conference (GLOBECOM)*, Abu Dhabi, United Arab, 1–6. IEEE. doi: [10.1109/GLOCOM.2018.8647714](https://doi.org/10.1109/GLOCOM.2018.8647714).
- Jain, M., G. Kaur, and V. Saxena. **2022**. A K-Means clustering and SVM based hybrid concept drift detection technique for network anomaly detection. *Expert Systems with Applications* 193:116510. doi: [10.1016/j.eswa.2022.116510](https://doi.org/10.1016/j.eswa.2022.116510).
- Jallouli, R., M. A. Tobji, D. Bélisle, S. Mellouli, F. Abdallah, and I. Osman. **2019**. Digital economy. In *Emerging Technologies and Business Innovation: 4th International Conference, ICDEc 2019*, 166–76, Beirut, Lebanon.; Springer International Publishing, Cham. doi: [10.1007/978-3-030-30874-2](https://doi.org/10.1007/978-3-030-30874-2).
- Janati, F., F. Abdollahi, S. S. Ghidary, M. Jannatifar, J. Baltes, and S. Sadeghnejad. **2017**. Multi-robot task allocation using clustering method. In *Robot intelligence technology and applications*, ed. J. H. Kim, F. Karray, J. Jo, P. Sincak, and H. Myung, vol. 233, 247. Cham: Springer. doi: [10.1007/978-3-319-31293-4_19](https://doi.org/10.1007/978-3-319-31293-4_19).
- Kamal, H., S. Gautam, D. Mehrotra, and M. S. Sharif. **2024**. Reinforcement learning model for detecting phishing websites. In *Cybersecurity and artificial intelligence: Transformational strategies and disruptive innovation*, ed. H. Jahankhani, G. Bowen, M. S. Sharif, and O. Hussien, 309–26. Cham: Springer. doi: [10.1007/978-3-031-52272-7_13](https://doi.org/10.1007/978-3-031-52272-7_13).
- Kanimozhi, V., and T. P. Jacob. **2019a**. Artificial intelligence based network intrusion detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing. In *2019 international conference on communication and signal processing (ICCSP)*, Chennai, India, 0033–0036. IEEE. doi: [10.1109/ICCSP.2019.8698029](https://doi.org/10.1109/ICCSP.2019.8698029).
- Kanimozhi, V., and T. P. Jacob. **2019b**. Calibration of various optimized machine learning classifiers in network intrusion detection system on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing. *International Journal of Engineering Applied Sciences & Technology* 4 (6):209–13. doi:[10.33564/IJEAST.2019.v04i06.036](https://doi.org/10.33564/IJEAST.2019.v04i06.036).
- Kato, K., and V. Klyuev. **2014**. An intelligent ddos attack detection system using packet analysis and support vector machine. *International Journal of Intelligent Computing Research* 5 (3):464–71. doi:[10.20533/ijicr.2042.4655.2014.0060](https://doi.org/10.20533/ijicr.2042.4655.2014.0060).
- Kaur, S., and A. Jindal. **2020**. Singular value decomposition (SVD) based image tamper detection scheme. In *2020 International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 695–99. IEEE. doi: [10.1109/ICICT48043.2020.9112432](https://doi.org/10.1109/ICICT48043.2020.9112432).
- Khan, A. I., and S. Al-Habsi. **2020**. Machine learning in computer vision. *Procedia Computer Science* 167:1444–51. doi: [10.1016/j.procs.2020.03.355](https://doi.org/10.1016/j.procs.2020.03.355).
- Khan, H. U., M. Z. Malik, and S. Nazir. **2024**. Identifying the ai-based solutions proposed for restricting money laundering in financial sectors: Systematic mapping. *Applied Artificial Intelligence* 38 (1):2344415. doi: [10.1080/08839514.2024.2344415](https://doi.org/10.1080/08839514.2024.2344415).
- Kim, J., J. Kim, H. Kim, M. Shim, and E. Choi. **2020**. Cnn-based network intrusion detection against denial-of-service attacks. *Electronics* 9 (6):916. doi: [10.3390/electronics9060916](https://doi.org/10.3390/electronics9060916).
- Kim, J., Y. Shin, and E. Choi. **2019**. An intrusion detection model based on a convolutional neural network. *Journal of Multimedia Information System* 6 (4):165–72. doi: [10.33851/JMIS.2019.6.4.165](https://doi.org/10.33851/JMIS.2019.6.4.165).
- Kozik, R., M. Choraś, R. Renk, and W. Hołubowicz. **2014**. A proposal of algorithm for web applications cyber attack detection. In *Computer information systems and industrial management*, ed. K. Saeed and V. Snášel, 8838. Berlin, Heidelberg: Springer. doi: [10.1007/978-3-662-45237-0_61](https://doi.org/10.1007/978-3-662-45237-0_61).

- Kumar, M., M. K. Jindal, and M. Kumar. 2022. A systematic survey on CAPTCHA recognition: Types, creation and breaking techniques. *Archives of Computational Methods in Engineering* 29 (2):1107–36. doi: [10.1007/s11831-021-09608-4](https://doi.org/10.1007/s11831-021-09608-4).
- Kumari, A., and A. K. Mehta. 2020. A hybrid intrusion detection system based on decision tree and support vector machine. In *2020 IEEE 5th International conference on computing communication and automation (ICCCA)*, Greater Noida, India, 396–400. IEEE. doi: [10.1109/ICCCA49541.2020.9250753](https://doi.org/10.1109/ICCCA49541.2020.9250753).
- Künzler, F. 2023. Real cyber value at risk: An approach to estimate economic impacts of cyberattacks on businesses. Master thesis, University of Zurich.
- Ladosz, P., L. Weng, M. Kim, and H. Oh. 2022. Exploration in deep reinforcement learning: A survey. *Information Fusion* 85:1–22. doi:[10.1016/j.inffus.2022.03.003](https://doi.org/10.1016/j.inffus.2022.03.003).
- Li, T., Z. Zeng, J. Sun, and S. Sun. 2022. Using data mining technology to analyse the spatiotemporal public opinion of COVID-19 vaccine on social media. *Electronic Library* 40 (4):435–52. doi: [10.1108/EL-03-2022-0062](https://doi.org/10.1108/EL-03-2022-0062).
- Liao, N., and X. Li. 2022. Traffic anomaly detection model using k-means and active learning method. *International Journal of Fuzzy Systems* 24 (5):2264–82. doi: [10.1007/s40815-022-01269-0](https://doi.org/10.1007/s40815-022-01269-0).
- Liberati, A., D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. A. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, and D. Moher. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLOS Medicine* 6 (7):e1000100. doi: [10.1371/journal.pmed.1000100](https://doi.org/10.1371/journal.pmed.1000100).
- Madsen, D. Ø., T. Berg, and M. D. Nardo. 2023. Bibliometric trends in industry 5.0 research: An updated overview. *Applied System Innovation* 6 (4):63. doi: [10.3390/asi6040063](https://doi.org/10.3390/asi6040063).
- Mallett, R., J. Hagen-Zanker, R. Slater, and M. Duvendack. 2012. The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness* 4 (3):445–55. doi: [10.1080/19439342.2012.711342](https://doi.org/10.1080/19439342.2012.711342).
- Marr, B. 2019. *Artificial intelligence in practice: How 50 successful companies used AI and machine learning to solve problems*. New York, USA: John Wiley & Sons.
- McIntosh, T., J. Jang-Jaccard, P. Watters, and T. Susnjak. 2019. The inadequacy of entropy-based ransomware detection. In *Neural Information Processing: 26th International Conference, ICONIP 2019*, ed. T. Gedeon, K. Wong, and M. Lee, 181–89, Sydney, Australia: Springer, Cham. doi: [10.1007/978-3-030-36802-9-20](https://doi.org/10.1007/978-3-030-36802-9-20).
- Melvin, A. A. R., G. J. W. Kathrine, S. S. Ilango, S. Vimal, S. Rho, N. N. Xiong, and Y. Nam. 2022. Dynamic malware attack dataset leveraging virtual machine monitor audit data for the detection of intrusions in cloud. *Transactions on Emerging Telecommunications Technologies* 33 (4):e4287. doi: [10.1002/ett.4287](https://doi.org/10.1002/ett.4287).
- Moerland, T. M., J. Broekens, A. Plaat, and C. M. Jonker. 2023. Model-based reinforcement learning: A survey. *Foundations and Trends* 16 (1):1–118. doi: [10.1561/22000000086](https://doi.org/10.1561/22000000086).
- Mohammadpour, L., T. C. Ling, C. S. Liew, and A. Aryanfar. 2022. A survey of cnn-based network intrusion detection. *Applied Sciences* 12 (16):8162. doi:[10.3390/app12168162](https://doi.org/10.3390/app12168162).
- Moustafa, N., and J. Slay. 2016. The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. *Information Security Journal: A Global Perspective* 25 (1–3):18–31. doi: [10.1080/19393555.2015.1125974](https://doi.org/10.1080/19393555.2015.1125974).
- Nievas, N., A. Pagès-Bernaus, F. Bonada, L. Echeverria, and X. Domingo. 2024. Reinforcement learning for autonomous process control in industry 4.0: Advantages and challenges. *Applied Artificial Intelligence* 38 (1):2383101. doi: [10.1080/08839514.2024.2383101](https://doi.org/10.1080/08839514.2024.2383101).
- Oropeza-Valdez, J. J., C. Padron-Manrique, A. Vazquez-Jimenez, X. Soberon-Mainero, and O. Resendis-Antonio. 2024. Exploring metabolic anomalies in COVID-19 and

- post-COVID-19: A machine learning approach with explainable artificial intelligence. *bioRxiv* 2024.04:15.589583. doi: [10.3389/fmolb.2024.1429281](https://doi.org/10.3389/fmolb.2024.1429281).
- Page, M. J., J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, and S. E. Brennan. 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372:n71. doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71).
- Patil, S. 2020. Global library and information science research seen through prism of biblioshiny. *Studies in Indian Place Names* 40 (49):157–70. <https://www.researchgate.net/publication/339973834>.
- Petrosyan, A. 2023. Estimated cost of cybercrime worldwide 2017-2028. Accessed April 25, 2024. <https://www.statista.com/forecasts/1280009/cost-cybercrime-worldwide>.
- Pranckutė, R. 2021. Web of science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications* 9 (1):12. doi: [10.3390/publications9010012](https://doi.org/10.3390/publications9010012).
- Pranto, M. B., M. H. Ratul, M. M. Rahman, I. J. Diya, and Z. Zahir. 2022. Performance of machine learning techniques in anomaly detection with basic feature selection strategy-a network intrusion detection system. *Journal of Advances in Information Technology* 13 (1):36–44. doi: [10.12720/jait.13.1.36-44](https://doi.org/10.12720/jait.13.1.36-44).
- Ramadhan, I., P. Sukarno, and M. A. Nugroho. 2020. Comparative analysis of K-nearest neighbor and decision tree in detecting distributed denial of service. In *2020 8th International Conference on Information and Communication Technology (ICoICT)*, 1-4, Yogyakarta, Indonesia. IEEE. doi: [10.1109/ICoICT49345.2020.9166380](https://doi.org/10.1109/ICoICT49345.2020.9166380).
- Ranjan, V., K. Patidar, and R. Kushwaha. 2020. An efficient image cryptography mechanism based on the hybridization of standard encryption algorithms. *ACCENTS Transactions on Information Security* 5 (20):42–47. doi: [10.19101/TIS.2020.517004](https://doi.org/10.19101/TIS.2020.517004).
- Rekha, G., S. Malik, A. K. Tyagi, and M. M. Nair. 2020. Intrusion detection in cyber security: Role of machine learning and data mining in cyber security. *Advances in Science Technology and Engineering Systems Journal* 5 (3):72–81.
- Sachdev, S. 2020. Breaking captcha characters using multi-task learning CNN and SVM. In *2020 4th International Conference on Computational Intelligence and Networks (CINE)*, 1–6, Kolkata, India: IEEE. doi: [10.1109/CINE48825.2020.234400](https://doi.org/10.1109/CINE48825.2020.234400).
- Saheed, Y. K., M. O. Arowolo, and A. U. Tosho. 2022. An efficient hybridization of K-means and genetic algorithm based on support vector machine for cyber intrusion detection system. *International Journal on Electrical Engineering and Informatics* 14 (2):426–42. doi: [10.15676/ijeei.2022.14.2.11](https://doi.org/10.15676/ijeei.2022.14.2.11).
- Salman, O., I. H. Elhaji, A. Kayssi, and A. Chehab. 2021. Data representation for CNN based internet traffic classification: A comparative study. *Multimedia Tools & Applications* 80 (11):16951–77. doi: [10.1007/s11042-020-09459-4](https://doi.org/10.1007/s11042-020-09459-4).
- Saranya, T., S. Sridevi, C. Deisy, T. D. Chung, and M. K. A. Khan. 2020. Performance analysis of machine learning algorithms in intrusion detection system: A review. *Procedia Computer Science* 171:1251–60. doi: [10.1016/j.procs.2020.04.133](https://doi.org/10.1016/j.procs.2020.04.133).
- Sarker, I. H. 2023. Machine learning for intelligent data analysis and automation in cybersecurity: Current and future prospects. *Annals of Data Science* 10 (6):1473–98. doi: [10.1007/s40745-022-00444-2](https://doi.org/10.1007/s40745-022-00444-2).
- Sarker, I. H., M. H. Furhad, and R. Nowrozy. 2021. Ai-driven cybersecurity: An overview, security intelligence modeling and research directions. *SN Computer Science* 2 (3):173. doi: [10.1007/s42979-021-00557-0](https://doi.org/10.1007/s42979-021-00557-0).
- Sarker, I. H., A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng. 2020. Cybersecurity data science: An overview from machine learning perspective. *Journal of Big Data* 7 (1):1–29. doi: [10.1186/s40537-020-00318-5](https://doi.org/10.1186/s40537-020-00318-5).
- Sarkis-Onofre, R., F. Catalá-López, E. Aromataris, and C. Lockwood. 2021. How to properly use the PRISMA statement. *Systematic Reviews* 10 (1):1–3. doi: [10.1186/s13643-021-01671-z](https://doi.org/10.1186/s13643-021-01671-z).

- Sommer, R., and V. Paxson. 2010. Outside the closed world: On using machine learning for network intrusion detection. In *2010 IEEE Symposium on Security and Privacy*, 305–16, Oakland, CA, USA: IEEE. doi: [10.1109/SP.2010.25](https://doi.org/10.1109/SP.2010.25).
- Soon, G. K., C. K. On, N. M. Rusli, T. S. Fun, R. Alfred, and T. T. Guan. 2020. Comparison of simple feedforward neural network, recurrent neural network and ensemble neural networks in phishing detection. *Journal of Physics: Conference Series* 1502 (1):012033. doi: [10.1088/1742-6596/1502/1/012033](https://doi.org/10.1088/1742-6596/1502/1/012033).
- Sultana, J., and A. K. Jilani. 2021. Classifying cyberattacks amid covid-19 using support vector machine. In *Security incidents & response against cyber attacks*, ed. A. Bhardwaj and V. Sapra, 161–75. Cham: Springer. doi: [10.1007/978-3-030-69174-5_8](https://doi.org/10.1007/978-3-030-69174-5_8).
- Sun, N., J. Zhang, P. Rimba, S. Gao, L. Y. Zhang, and Y. Xiang. 2018. Data-driven cybersecurity incident prediction: A survey. *IEEE Communications Surveys & Tutorials* 21 (2):1744–72. doi: [10.1109/COMST.2018.2885561](https://doi.org/10.1109/COMST.2018.2885561).
- Tareq, W. Z. T., and M. Davud. 2024. Classification and clustering. *Decision-Making Models*, 351–59. New York: Academic Press. doi: [10.1016/B978-0-443-16147-6.00024-4](https://doi.org/10.1016/B978-0-443-16147-6.00024-4).
- Tennekes, M. 2018. Tmap: Thematic maps in R. *Journal of Statistical Software* 84 (6):1–39. doi: [10.18637/jss.v084.i06](https://doi.org/10.18637/jss.v084.i06).
- Veena, K., K. Meena, Y. Teekaraman, R. Kuppusamy, A. Radhakrishnan, and D. K. Jain. 2022. C SVM classification and KNN techniques for cyber crime detection. *Wireless Communications and Mobile Computing* 2022:1–9. doi: [10.1155/2022/3640017](https://doi.org/10.1155/2022/3640017).
- Verma, A., and V. Ranga. 2023. On evaluation of network intrusion detection systems: Statistical analysis of CIDDs-001 dataset using machine learning techniques. *Pertanika Journal of Science & Technology* 26 (3):1307–32. doi: [10.36227/techrxiv.11454276.v1](https://doi.org/10.36227/techrxiv.11454276.v1).
- Verma, M. K., S. Yadav, B. K. Goyal, B. R. Prasad, and S. Agarawal. 2019. Phishing website detection using neural network and deep belief network. In *Recent findings in intelligent computing techniques, advances in intelligent systems and computing*, ed. P. Sa, S. Bakshi, I. Hatzilygeroudis, and M. Sahoo, vol. 707. Singapore: Springer. doi: [10.1007/978-981-10-8639-7_30](https://doi.org/10.1007/978-981-10-8639-7_30).
- Wang, J., and I. C. Paschalidis. 2016. Botnet detection based on anomaly and community detection. *IEEE Transactions on Control of Network Systems* 4 (2):392–404. doi: [10.1109/TCNS.2016.2532804](https://doi.org/10.1109/TCNS.2016.2532804).
- Wang, Z., P. Shi, and M. I. Uddin. 2021. CAPTCHA recognition method based on CNN with focal loss. *Complexity* 2021 (1):1–10. doi: [10.1155/2021/6641329](https://doi.org/10.1155/2021/6641329).
- Wei, L., X. Li, T. Cao, Q. Zhang, L. Zhou, and W. Wang. 2019. Research on optimization of CAPTCHA recognition algorithm based on SVM. In *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, Zhuhai, China, 236–40. doi: [10.1145/3318299.3318355](https://doi.org/10.1145/3318299.3318355).
- Wirkuttis, N., and H. Klein. 2017. Artificial intelligence in cybersecurity. *Cyber, Intelligence and Security* 1 (1):103–19. doi: [10.1006/jesp.1996.0006](https://doi.org/10.1006/jesp.1996.0006).
- World Bank. 2024. Cybersecurity multi-donor trust fund. Accessed June 13, 2024. <https://www.worldbank.org/en/programs/cybersecurity-trust-fund/overview>.
- Xie, M., J. Hu, and J. Slay. 2014. Evaluating host-based anomaly detection systems: Application of the one-class SVM algorithm to ADFA-LD. In *2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 978–82, Xiamen, China. doi: [10.1109/FSKD.2014.6980972](https://doi.org/10.1109/FSKD.2014.6980972).
- Yilmaz, A. B., Y. S. Taspinar, and M. Koklu. 2022. Classification of malicious android applications using naive Bayes and support vector machine algorithms. *International Journal of Intelligent Systems and Applications in Engineering* 10 (2):269–74.

- You, J., J. Jia, X. Pang, J. Wen, Y. Shi, and J. Zeng. 2023. A novel multi-robot task assignment scheme based on a multi-angle K-means clustering algorithm and a two-stage load-balancing strategy. *Electronics* 12 (18):3842. doi: [10.3390/electronics12183842](https://doi.org/10.3390/electronics12183842).
- Zhang, Z., H. Al Hamadi, E. Damiani, C. Y. Yeun, and F. Taher. 2022a. Explainable artificial intelligence applications in cyber security: State-of-the-art in research. *Institute of Electrical and Electronics Engineers Access* 10:93104–39. doi: [10.1109/ACCESS.2022.3204051](https://doi.org/10.1109/ACCESS.2022.3204051).
- Zhang, Z., H. Ning, F. Shi, F. Farha, Y. Xu, J. Xu, F. Zhang, and K. R. Choo. 2022b. Artificial intelligence in cyber security: Research advances, challenges, and opportunities. *Artificial Intelligence Review* 55 (2):1029–53. doi: [10.1007/s10462-021-09976-0](https://doi.org/10.1007/s10462-021-09976-0).