Artificial Intelligence Review (2025) 58:283
https://doi.org/10.1007/s10462-025-11292-w

Adaptive anomaly detection for identifying attacks in cyber-physical systems: A systematic literature review

Pablo Moriano[1] · **Steven C. Hespeler**[1] · **Mingyan Li**[2] · **Maria Mahbub**[2]

Accepted: 5 June 2025 / Published online: 23 June 2025

Abstract

Modern cyberattacks in cyber-physical systems (CPS) rapidly evolve and cannot be deterred effectively with most current methods, which focus on characterizing past threats. Adaptive anomaly detection (AAD) is among the most promising techniques to detect evolving cyberattacks, with an emphasis on fast data processing and model adaptation. AAD has been researched extensively; however, to the best of our knowledge, our work is the first systematic literature review (SLR) on current research in this field. We present a comprehensive SLR, gathering 397 relevant papers and systematically analyzing 65 of them (47 research and 18 survey papers) on AAD in CPS from 2013 to November 2023. We introduce a novel taxonomy considering attack types, CPS application, learning paradigm, data management, and algorithms. Our findings show that most studies addressed either model adaptation or data processing, but rarely both simultaneously. This indicates a research gap in fully adaptive solutions. We also categorize algorithms, datasets, and attack characteristics, and summarize strengths and weaknesses across the literature. Our review provides a structured and accessible reference for researchers and practitioners, offering insights into key trends and highlighting limitations in current approaches. Finally, we outline several future research directions, including the need for integrated real-time processing and adaptive learning, explainability, and uncertainty quantification in AAD for CPS.

Keywords Cybersecurity · Anomaly detection · Adaptation · Cyber-physical systems

Pablo Moriano and Steven Hespeler have equally contributed to this work.

Extended author information available on the last page of the article

1 Introduction

Modern cyber-physical systems (CPS) including industrial control systems (ICS), vehicles, power grids, and the Internet of Things (IoT), among others, generate vast amounts of high speed data that need to be processed to support decision-making capabilities (Atat et al. 2018; Fei et al. 2019; Kayan et al. 2022). Due to the mission-critical nature that CPS play, securing its operation against malicious threats is essential for guaranteeing daily life activities. However, since information technology (IT) and physical processes are closely linked in CPS, they face a broader range of threats, including both cyber and physical attacks. Here, cyberattacks refer to attacks against the communication and computing components of CPS (Shacham et al. 2004; Pike et al. 2016; Clements et al. 2017). Physical attacks refers to compromising the physical environment of a CPS, subjecting the system to potential malicious data via injection or tampering through sensor and/or actuators (Rutkin 2013; Shoukry et al. 2013; Petit et al. 2015).

Unlike traditional IT systems, CPS must operate under strict timing constraints and resource limitations, while maintaining high reliability and safety in dynamic environments (Lee 2008; Pasqualetti et al. 2013). The tight coupling between physical processes and computational elements creates complex dependencies and feedback loops, which complicate threat detection and response (Cárdenas et al. 2008; Urbina et al. 2016). Furthermore, CPS deployments are often long-lived and lack standardized update mechanisms, making static defenses inadequate against evolving threats (Kundur et al. 2011). These characteristics create a strong need for adaptive detection mechanisms that can continuously learn from operational data and respond to both known and unknown threats.

Among the most used techniques to detect threats in CPS are those based on anomaly detection (Chandola et al. 2009; Mirsky et al. 2018; Moriano et al. 2021, 2022). As opposed to traditional signature-based detection focused on matching patterns from previously seen attacks (Hubballi and Suryanarayanan 2014; Ioulianou et al. 2018; Wu et al. 2019), anomaly detection techniques focus on spotting behavior that looks different from the expected norm (Luo et al. 2021; Shahriar et al. 2023; Moriano et al. 2024). This approach helps on identifying previously unseen attacks as those commonly affecting the cyber and physical components of CPS (Schneider and Böttinger 2018; Xi et al. 2022). Different anomaly detection approaches have been proposed to secure CPS including those based on attack-resilient sensor fusion (Ivanov et al. 2016; Lu et al. 2018), model-based attack detection (Quinonez et al. 2020; Giraldo et al. 2018), and data-based detection (Junejo and Goh 2016; Shin et al. 2017). However, the ability of anomaly detection methods to adapt to detect previously unseen attacks is usually not thoroughly explored. Here, adaptability is closely tied with the concept of "adaptation," which means the ability of an anomaly detection method to anticipate and respond to new and emerging security threats by learning from the experience and the current state of the CPS (Abie 2019; Andrade and Yoo 2019).

Adaptive anomaly detection (AAD) requires two key components: (1) near real-time data processing and (2) a predefined learning mode for model adaptation (Raciti 2013; Settanni et al. 2018; Akowuah and Kong 2021; Biggio 2024). Near real-time processing assists in detecting attacks before they cause consequences, which is crucial in safety-critical CPS. In addition, a predefined learning model (such as full, incremental, or hybrid retrain (Gama et al. 2014)) is needed to adapt the detection model to respond better to unseen attacks. Both components of AAD generally ensure a strong defense against advanced cyberattacks.

This paper presents a systematic literature review (SLR) on AAD in CPS. Our goal is to provide a comprehensive overview of the state of the art in AAD, including their usage across different types of CPS, classification across different learning paradigms, common algorithms, and a discussion of trends and gaps in the literature. To our knowledge, this is the first SLR on AAD in CPS. More specifically, the objectives of our SLR are to: (1) provide researchers with an understanding of current AAD methods in CPS, enabling new researchers to quickly familiarize themselves; (2) highlight gaps and opportunities for future research; and (3) support practitioners in selecting and adapting AAD methods in CPS to fit their needs.

In this SLR, we adopt the term AAD to specifically refer to anomaly detection methods that can adapt to evolving system behaviors and unseen threats through continuous model updates or online learning mechanisms. The "adaptiveness" in AAD denotes the capacity to modify either the data representation, the detection model, or both, in response to changes in the data distribution, attack strategies, or system context. Adaptation can occur through mechanisms such as incremental learning, concept drift handling, or online retraining. It is important to distinguish this definition from active anomaly detection, which, in the broader machine learning literature, often refers to anomaly detection augmented with active learning strategies. In such methods, a model queries an oracle (usually a human expert) to label selected data points with high uncertainty to improve detection performance with minimal labeling effort (e.g., Das et al. 2018; Kim et al. 2023). While both AAD and active anomaly detection aim to enhance detection in dynamic settings, they differ in emphasis: AAD focuses on model self-adaptation to new data or concept drift, whereas active anomaly detection emphasizes efficient data labeling via interactive query strategies. To maintain clarity, throughout this work we refer to adaptive anomaly detection in the context of self-adjusting detection systems for CPS, not involving human-in-the-loop labeling processes unless explicitly stated.

In the past decade, a large number of studies were published covering different aspects of AAD in CPS (Li et al. 2016; Adhikari et al. 2017; Van Wyk et al. 2019; Yasaei et al. 2020; Mowla et al. 2020; Jiao et al. 2022; Ding et al. 2022; Gyamfi and Jurcut 2022; Intriago and Zhang 2023; Cai et al. 2023). A small amount of these surveys have also explored the application of machine learning and data mining techniques to various cybersecurity domains, with an explicit focus on addressing intrusion detection challenges for securing CPS and providing insights into methodologies and best practices (Buczak and Guven 2015; Olowononi et al. 2020). Nonetheless, most studies investigated focus on only one of the key components of AAD (i.e., near real-time data processing or predefined learning mode). Accordingly, results were contradictory and practices heterogeneous. This makes it difficult to contextualize their contribution in terms of how adaptation is carried out.

To fill in the gaps and provide an updated and comprehensive review on the latest developments in AAD in CPS, the present study reviews state-of-the-art works published between 2013 and 2023 (November). The contributions of this article can be summarized as follows:

(1) We review and classify state-of-the-art AAD methods in CPS considering type of application, learning paradigm, data management strategy, and algorithms, along with a comprehensive summary tables (see Tables 2–6).

(2) We introduce a novel AAD taxonomy for CPS that focuses of attack types, applications, and ML algorithms to categorize reviewed works based on learning paradigm and algorithms.
(3) We identify and discuss limitations of reviewed AAD approaches for securing CPS.
(4) We discuss priority future areas of research in this field.

We organize this SLR as follows. In Sect. 2, we contextualize our SLR with respect to other surveys in closely-related areas. Section 3 details the methodology we used to conduct the SLR. Section 4 introduces an AAD taxonomy for CPS and synthesize previous research based on the learning paradigm of the algorithms. In Sect. 5, we discuss our findings and potential future research directions. Finally, we provide a brief summary of this SLR in Sect. 6.

2 Related work

A key distinction must be made between AAD and active anomaly detection, as these terms are occasionally used interchangeably in literature but refer to conceptually different approaches. Active anomaly detection is typically grounded in active learning paradigms, where the system actively selects the most informative data points to query labels for, aiming to optimize model performance with limited labeling resources (Das et al. 2018; Kim et al. 2023). In contrast, this SLR defines adaptive anomaly detection as the use of self-learning or updating strategies that enable the model to respond to evolving environments and cyberattack patterns. This includes strategies such as streaming data adaptation, model retraining, online updating, or drift-aware frameworks (see Results section). While there is potential overlap–such as systems combining active learning with adaptive model updates–the majority of the reviewed CPS literature adopts "adaptive" to imply autonomously updating models without human labeling assistance. This distinction is crucial for interpreting contributions and comparing methods across works.

As the field of AAD in CPS has matured over the last decade, several survey papers have been published with both broad and narrow scope in this field. The works discussed in this section were identified and compiled during the comprehensive literature search detailed in Section 3.

Saad et al. (2019) analyzed cybersecurity challenges in smart grids. They examined vulnerabilities, security needs, detection techniques, countermeasures, and secure communication protocols. Their study emphasized integrating advanced communication and computing technologies to improve reliability and efficiency. They proposed solutions to mitigate cyberattacks, enhancing the security of power networks. Zhang et al. (2019) summarized recent advances in false data injection attacks (FDIA) targeting smart grid state estimation. They reviewed FDIA construction methods, detection strategies, and defense mechanisms. They also outlined future directions, such as applying FDIAs to alternating current (AC) state estimation and using data-driven models for FDIA detection. Syrmakesis et al. (2022) classified cyber-resilience methods for protecting smart grids. They reviewed approaches addressing cyberattacks and anomalies. Their taxonomy supports research on cyber-resilience and identifies promising directions for smart grid security.

Rojas and Rauch (2019) conducted a systematic review of 165 papers on CPS production trends. They highlighted the importance of connectivity and control systems in manufacturing. Their work grouped papers into six categories, including cybersecurity enablers for smart manufacturing. Zeadally et al. (2019) reviewed 12 self-adaptive approaches for managing large-scale CPS. They discussed the strengths, weaknesses, and implementation techniques for self-adaptive mechanisms across CPS architectural layers, including physical, network, and cyber. Rosenberg et al. (2021) reviewed 58 papers on adversarial attacks and defenses in cybersecurity. They proposed a taxonomy based on attack stages, goals, and capabilities. Their work highlighted future research needs in adversarial machine learning (ML). Jamal et al. (2023) reviewed ML and deep learning (DL) techniques for CPS security. They identified challenges, reviewed existing methods, and proposed future directions for artificial intelligence (AI)-based CPS protection against cyber threats. Pekaric et al. (2023) reviewed 21 papers on safety and security in self-adaptive systems. They found that current approaches rarely model both aspects together and often rely on simulations with simplified use cases. Koay et al. (2023) listed vulnerabilities and cyberattacks in ICS. They compared ML methods for attack detection and highlighted challenges like limited datasets and risks of adversarial attacks. They proposed urgent research directions.

Mahapatra et al. (2020) created a taxonomy for secure IoT communication. They classified architectures, communication topologies, and secure transmission methods, including cluster-based and blockchain-based techniques. They identified research challenges and opportunities to enhance IoT security. Stoyanova et al. (2020) reviewed IoT forensics and its challenges, such as device diversity, encryption, and cloud reliance. They examined theoretical models and proposed frameworks using blockchain for evidence integrity. They emphasized the need for standardized forensic processes and readiness strategies. Dai and Boroomand (2022) reviewed 58 papers on security issues in big data systems. They focused on AI techniques like DL and multi-agent systems for detecting and mitigating attacks. They mapped these methods to security strategies and evaluation models. Huang et al. (2022) reviewed reinforcement learning (RL) for cyber-resilience. They discussed vulnerabilities, including posture- and information-related issues, and applications like moving target defense. They also analyzed vulnerabilities in RL systems under adversarial attacks. Alaghbari et al. (2022) surveyed complex event processing (CEP) in CPS security. They highlighted how combining CEP with ML enhances intrusion detection. They also discussed open issues in CEP applications for cybersecurity. Lin et al. (2023) surveyed security and privacy issues in 5 G-industrial IoT (IIoT) factories. They reviewed approaches using DL, RL, and blockchain. Their work identified research opportunities for securing 5 G-enabled industrial systems.

Grimm et al. (2021) reviewed 50 papers on adaptive and intelligent security for vehicles and fleets. They identified open research areas and developed a taxonomy of contextual information categories. Their work aims to guide future developments in automotive security. Strandberg et al. (2022) conducted a systematic review of 67 papers on automotive digital forensics. They categorized papers into technical solutions and surveys and mapped forensic data to security properties and stakeholders. Their findings are relevant to CPS like smart cities.

Cooper et al. (2023) reviewed anomaly detection methods for power system state estimation. They connected traditional data-driven methods with modern approaches addressing

new cyber threats. They proposed directions for future research, including dynamic load profiles and asynchronous measurements.

Table 1 lists and identifies prior surveys related to AAD in CPS. It also outlines publication details and the contribution of existing surveys in the area. While there have been several valuable works reviewing different aspects of AAD in CPS, including adaptability (Zeadally et al. 2019; Pekaric et al. 2023) and online learning (Rosenberg et al. 2021; Huang et al. 2022), they do not provide an extensive analysis of other critical aspects related to these tasks including data management and concept drift. For example, Zeadally et al. (2019) focuses on discussing techniques that enable self-adaptation capabilities within CPS at different architectural layers. However, it lacks insight into the role of different data man-

Table 1 State-of-the-art surveys related to adaptive anomaly detection in CPS

No.	Reference	Year	Time Frame	No. Papers Surveyed	Survey Topic	Application
1	Saad et al. (2019)	2019	Not specified	Not specified	Modern strategies for mitigating cyberattacks	Smart grid
2	Zhang et al. (2019)	2019	Not specified	Not specified	Recent advances against FDIAs	Smart grid
3	Rojas and Rauch (2019)	2019	2012–2017	165	Smart manufacturing control	CPS
4	Zeadally et al. (2019)	2019	Not specified	12	Self-adaptive mechanisms	CPS
5	Mahapatra et al. (2020)	2020	Not specified	Not specified	Secure transmission in IoT	IoT
6	Stoyanova et al. (2020)	2020	Not specified	Not specified	IoT forensics	IoT
7	Rosenberg et al. (2021)	2021	Not specified	Not specified	Adversarial attacks and defenses	CPS
8	Grimm et al. (2021)	2021	Not specified	50	Context-aware security	Vehicle
9	Strandberg et al. (2022)	2022	2006–2021	67	Automotive digital forensics	Vehicle
10	Dai and Boroomand (2022)	2022	2006–2021	58	AI-driven security for big data	IoT
11	Syrmakesis et al. (2022)	2022	Not specified	Not specified	Classification of cyber resilience methods	Smart grid
12	Huang et al. (2022)	2022	Not specified	Not specified	RL for cyber resilience	IoT
13	Alaghbari et al. (2022)	2022	Not specified	Not specified	Complex event processing (CEP)	IoT
14	Jamal et al. (2023)	2023	Not specified	Not specified	Analysis of ML and DL applications	CPS
15	Pekaric et al. (2023)	2023	2000–2020	21	Security of self-adaptive systems	CPS
16	Cooper et al. (2023)	2023	Not specified	Not specified	Anomaly detection for power system	Power grid
17	Lin et al. (2023)	2023	2018–2021	22	Security and privacy in 5 G-IIoT	IoT
18	Koay et al. (2023)	2023	2017–2022	30	Attacks and defenses in ICS	CPS

agement strategies to handle streaming data conditions, as well as the categorization of adaptive techniques per application.

In contrast to these previous surveys, our SLR introduces a novel and comprehensive taxonomy that uniquely integrates multiple adaptation dimensions, including attack type, CPS application domain, learning paradigm, data handling approach, and algorithmic strategies. No prior review has examined these dimensions collectively or provided detailed mappings across them. Moreover, we are the first to analyze adaptation along both the data processing and model learning pipelines, revealing that most studies tend to focus on one aspect while overlooking their combined effects.

Our SLR also surpasses prior work by evaluating a broader set of literature, i.e., 397 papers retrieved through a rigorous search protocol across five major databases, with 65 key papers (47 research and 18 surveys) selected for systematic analysis. This level of methodological thoroughness and the size of the dataset analyzed are substantially greater than those of previous surveys (as evidenced by Table 1), which often focus on narrower subfields or smaller paper sets.

To address these limitations and provide an in-depth understanding of state-of-the-art on AAD in CPS, we conduct a systematic and extensive survey of related literature. By comprehensively collecting 397 papers and systematically analyzing 65 papers (47 research papers and 18 surveys) from top journals and conferences, our SLR aims to provide a holistic summary of AAD methods and its broad applicability within the CPS domain. Our contribution lies in synthesizing findings in a multi-dimensional, cross-comparative manner, contextualized by practical CPS applications, enabling researchers and practitioners to identify effective adaptive strategies for various real-world challenges. This approach enables a more nuanced analysis of the field highlighting strengths and limitations of different approaches. This contributes to identify most promising directions for future research.

3 SLR methodology

3.1 Survey method

This section discusses the scope and SLR used in this work. We adapt Okoli's (2015) review process for creating a standalone SLR and designed it with precision to ensure the comprehensive collection of relevant articles while maintaining methodological rigor. This SLR centers on summarizing the literature on AAD methods for CPS, focusing on the mechanisms and models used, rather than providing a comprehensive taxonomy of the specific threats addressed within the field (e.g., DoS, FDI, or advanced persistent threats (APT)). Here, the review process consists of several sequential steps including planning, selection, extraction, and execution that ultimately organize the review process into a reproducible SLR.

3.1.1 Eligibility criteria

To maintain the integrity of this SLR and ensure that the selected articles met the requisite standards of relevance and quality, we establish and follow a strict inclusion and exclusion criteria. To be included in this SLR, studies had to:

- Be published in peer-reviewed journals or conference proceedings, ensuring that only high-quality, validated research was considered.
- Be published within the past 10 years (2013–Nov. 2023) to capture recent developments and advancements in AAD, reflecting the rapidly evolving nature of this field.
- Be written in English to maintain consistency in interpretation and reduce potential translation biases.
- Address topics relevant to AAD or its subdomains with a clear focus on adaptive, cognitive, or autonomous mechanisms for anomaly detection in CPS.

Articles were excluded if they:

- Were non-peer-reviewed articles, such as editorials, letters, etc.
- Lacked a clear connection to AAD or cognitive approaches for CPS, ensuring that the review remained focused on its central research question.

3.1.2 Data sources

This SLR was conducted by retrieving academic papers and conference proceedings from five prominent article databases: IEEE Xplore (https://ieeexplore.ieee.org/), ACM Digital Library (https://dl.acm.org/), Emerald Insight (https://emerald.com/insight/), Springer Link (https://link.springer.com/), and ScienceDirect (https://sciencedirect.com/search). These databases were selected due to their extensive coverage of fields related to cybersecurity, adaptive systems, and AI, ensuring a comprehensive exploration of relevant literature. Each database was searched for publications covering the period from 2013 to November 2023 to align with the eligibility criteria. In addition to electronic searches within these databases, supplementary strategies (forward/backward search Wohlin 2014) were applied to identify additional key studies. This approach enabled us to capture influential works not readily indexed by standard database searches and enhanced the comprehensiveness of our SLR.

3.1.3 Search strategy

Figure 1 highlights the initial search process, which first establishes 10 key-terms including: `adaptive cyber security; dynamic adaption AND cyber security; adaptive control AND cyber security; dynamic control AND cyber security; adaptive AND cyber threats; dynamic AND cyber threats; adaptive AND cognitive cybersecurity; intelligent cyber AND threat detection; cognitive cyber AND threat detection; Adaptive systems AND cognitive cyber.` 10 keywords were searched as well, including: `abnormality detection models, anomaly detection, autonomous cyber-physical systems, cognitive cybersecurity, cognitive platform, industry 4.0, internet of things, intrusion detection, intrusion detection system, threshold adaptation.` Adjustments were made to improve search efficacy based on each database's functionality. For example, the search term "adaptive security" was used in the ACM Digital Library to retrieve a broader set of results, while ScienceDirect required expanded criteria to title/abstract/author-specified keywords for higher relevancy.

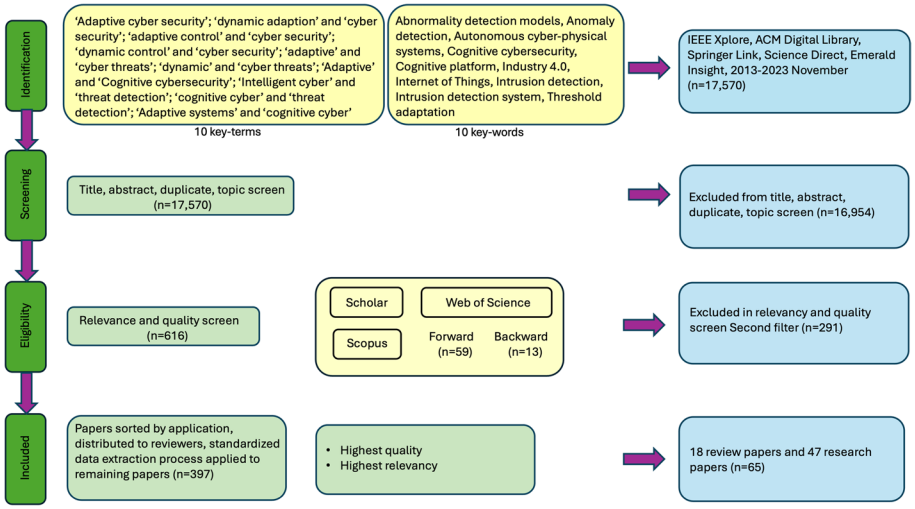

Fig. 1 Flow chart of the search and selection process, highlighting stages from initial identification through screening, eligibility, and final inclusion. This sequential filtering ensures a rigorous selection of high-quality papers. Here, yellow bubbles highlight a process that adds papers to the selection pool, green bubbles highlight a process that screens papers, and blue bubbles highlight the result

To enhance the comprehensiveness of the SLR, we performed a forward and backward search on a set of selected high-impact and high-quality papers. The set of core papers were used in congruence with a research analytics tool called SciVal from Scopus, which provided us with accurate and up-to-date citation numbers for each paper. After setting a threshold at the 99th percentile for citation numbers, we identified four papers that met or exceeded this criterion. These core papers became the cornerstone of our forward and backward search, allowing us to trace their references (backward search) and discover the subsequent papers that cited them (forward search). Utilizing this approach was instrumental in ensuring that our SLR encompassed foundational works and recent advancements within the field of AAD.

Figure 1 illustrates the four main stages of our systematic review process: Identification, Screening, Eligibility, and Inclusion. In the Identification stage, we retrieved a total of 17,570 records from major scholarly databases including IEEE Xplore, ACM Digital Library, Springer Link, Science Direct, and Emerald Insight using a set of targeted search strings and domain-specific keywords. During the Screening stage, we applied an initial filter based on titles, abstracts, duplicates, and topic relevance, which led to the exclusion of 16,954 records. This left 616 papers for the Eligibility stage, where a more detailed relevance and quality screen was conducted. In this stage, an additional 291 papers were excluded. Additionally, 72 relevant papers were added through forward and backward citation searches using Google Scholar, Scopus, and Web of Science. This resulted in a refined set of 397 papers, which were sorted by application, distributed to reviewers, and subjected to a standardized data extraction process. Finally, in the Inclusion stage, a core subset of 65 papers, comprised of 18 review papers and 47 research papers, was selected for full inclusion in the final synthesis based on the highest relevance and quality criteria.

3.1.4 Study selection

From here, we create and manually apply a tagging system. The purpose of this step was to be able to systematical filter relevant papers based on what a papers' main focus was. Figure 2 highlights the distribution of applications and digital library used amongst the papers based on our tagging system. After filtering for relevance and removing duplicates is performed with the first screening. The tagging system enabled us to organize papers systematically, supporting the identification of commonalities and trends in the literature. Pie charts were used to represent distributions across different digital libraries and application areas, facilitating a concise visualization of contributions in each category.

3.1.5 Data collection process

The final process of this methodology involved the extraction of all the papers selected for review. To approach this process from a systematic stand point, we constructed a data extraction form for reviewers to extract relevant information from each paper and normalize the process to reduce individual reviewer bias. The extraction process was inspired by Loeffel (2017), which lays out the basic ingredients for AAD. The data collection process involved the creation of a standardized data extraction template. This template was structured into several "blocks" to capture key information from each paper, including the general information, dataset, model, data management, learning, and paper quality blocks. Each block contained specific fields, some with selectable options (e.g., "application tags") and others allowing for manual input to capture nuanced information. For example, within the "learning block" are specific items for the authors to address including; learning paradigm, learning mode, model adaptation, and ensemble. Each of these items within this block aim to extract specific information from each paper. The final block determines the overall quality of the paper for establishing the potential usage of a paper as a featured paper for final review. This standardized approach minimized individual reviewer bias and ensured consistent data extraction. To ensure only high-quality studies were included, we established a set of threshold criteria. Each paper was evaluated based on its relevance to AAD, credibility of findings, and specific contributions. This rigorous filtering process led to the final selection of 65 high-quality and directly related papers, of which 18 were review papers.

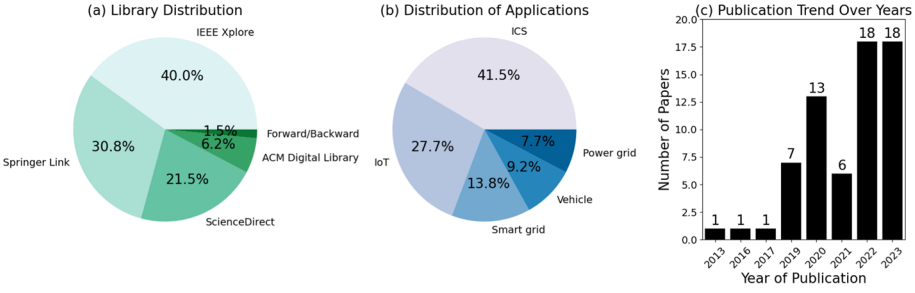

Fig. 2 Distributions across selected papers in the review. **a** Digital libraries distribution, with IEEE Xplore, Springer Link, Science Direct, and ACM Digital Libraries as primary sources. **b** Distribution of applications including CPS, IoT, smart grid, vehicle, and energy. **c** Annual distribution of reviewed papers from 2013 to 2023

3.2 Trend analysis

An important aspect of this SLR is to examine trends in the selected literature, allowing the data to highlight the primary focus areas and the evolution of AAD for CPS research. This analysis provides a structured view of the research landscape and informs the thematic focus of this SLR. Figure 2a shows the library distribution revealing that the majority of selected papers were sourced from IEEE Xplore (40.0%), followed by Springer Link (30.8%) and Science Direct (21.5%). This distribution suggests that these databases are central repositories for cybersecurity research, particularly for studies focused on AAD aspects. The prominence of IEEE highlights its significant role in the dissemination of research within technical and engineering domains and aligns well with the focus on adaptive cybersecurity solutions.

Figure 2b displays the application distribution across the selected papers, with ICS emerging as the dominant application area, constituting 41.5% of the studies. This finding underscores the importance of ICS in AAD since ICS environments often involve complex interactions between physical and digital components. IoT (27.7%) and smart grid (13.8%) also represent significant portions of the application distribution, reflecting an increase in cybersecurity challenges within interconnected and critical infrastructure systems. Remaining applications like vehicles (9.2%) and the power grid (7.7%), further demonstrate the multidisciplinary nature of AAD research. This demonstrates vulnerabilities across various sectors where adaptive methods are essential to managing dynamic threats.

Figure 2c shows the publication trend over the years. This indicates a noticeable increase in research activity from 2019 onwards and highlights peaks in 2022 and 2023 (each year contributing 18 papers to the final selection). This upward trend reflects a growing recognition of the need for adaptive and cognitive approaches to cybersecurity, especially as cyber threats become more sophisticated and pervasive. The surge in publications over the past five years aligns with advancements in AI and ML, which have enabled the development of more complex and responsive cybersecurity solutions.

4 AAD in CPS

Anomaly detection methods can be categorized into two categories as offline and online anomaly detection based the nature of data being processed (Chandola et al. 2009; Odiathevar et al. 2019). On the one hand, offline methods are trained from a static dataset. Most offline methods focus on thresholding requiring extensive training (Ibidunmoye et al. 2017). Although offline methods are able to identify complex patterns, they need to be retained whenever model's performance deteriorates. Therefore, as data streams imply rapid contextual changes from a learnt baseline, offline methods are usually prone to produce a significant number of false positives. On the other hand, online methods are usually based on incrementally learning from time windows as new data arrives (Moriano et al. 2024). This implies that online methods are better suited to tackle concept drift. AAD based on offline and online methods have been successfully used by researchers to detect cyberattacks in CPS (Wang and Govindarasu 2020; Gyamfi and Jurcut 2022; Alsulami et al. 2023).

The taxonomy of papers reviewed in this SLR review is shown in Fig. 3. CPS are vulnerable to cyberattacks such as denial of service (DoS), fuzzing, FDIA, spoofing, and mas-

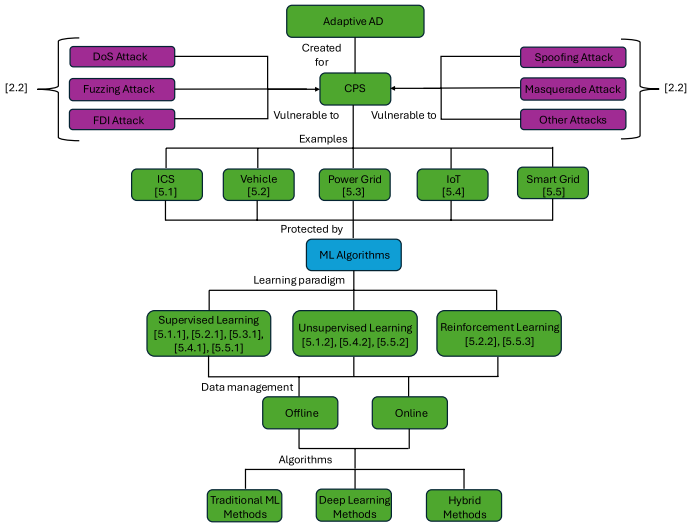

Fig. 3 AAD for CPS taxonomy. The numbers indicate the sections that cover each topic in the taxonomy

querade attacks, among others. In the literature, authors reported other attacks scenarios including advanced persistent threats, jamming, and ransomware. For a comprehensive review of other attacks on CPS, we refer the reader to the work by Yampolskiy et al. (2013) and Kim et al. (2022). Based on the type of CPS, these attacks target different applications. We then consider the domain application to classify existing work. Specifically, we focus on the more prevalent CPS applications including ICS, vehicle, power grid, IoT, and smart grid (Humayed et al. 2017). While surveying the literature, it was observed that studies commonly sorted applications based on certain themes. We summarized these papers by establishing distinct categories. Studies pertaining to ICS focused on centralized, real-time automation of industrial processes, prioritizing operational reliability and safety using fixed-location infrastructure. Papers focused on vehicles encompassed connected and autonomous systems, operating in dynamic, mobile environments where rapid decision-making and adaptability are critical. Power grid investigations ensured stable energy transmission and distribution with an emphasis on real-time monitoring for operational security. Papers explicitly focused on IoT systems connected heterogeneous devices across diverse domains, emphasizing scalability and user-centric automation. Smart grids extend these capabilities by integrating IoT and adaptive analytics to optimize energy efficiency and resilience against cyber-physical threats.

We categorize existing work based on the type of ML paradigm, namely supervised learning, unsupervised learning, and RL. In supervised learning, models learn from labeled data. In unsupervised learning, models learn and capture the structure of normal/regular data only (Ahmad et al. 2017; Munir et al. 2018). Hence, in this work, methods that rely exclusively on benign data during training, also known as one-class classification, are categorized as unsupervised learning. In RL, the goal is to learn from interaction with an environment in order to secure assets. In doing so, RL agents map a state to actions via a policy function to maximize the numerical reward of the signal (Arshad et al. 2022).

Comprehensive summary tables for each subsection are displayed in Tables 2–6.

Table 2 Summary of AAD methods in ICS applications

Reference	Model	Algorithm	Dataset	Attack	Strengths	Weaknesses
Mitchell and Chen (2013)	Un-supervised	SPN	Simulation of MCPS	Node capture and bad data injection	Comprehensive mathematical model to balance energy consumption and intrusion detection	Not demonstrated that model generalizes well on real-world data
Nakayama et al. (2019)	Supervised	G-KART	IEEE 33-bus power	FDI attacks	Adaptive robust thresholding mechanism	Tested in simulated environment
Pan et al. (2019)	Supervised	DT, SVM, KNN, and boosted trees	Real-world, large-scale channel measurement campaign from 4 real industrial sites	Spoofing attacks	System can adapt to dynamic environments and significantly improve authentication accuracy	Offline learning
Saez et al. (2019)	Supervised	Context-sensitive adaption, SVM	Sensor data, CNC machine	Anomalies only	Adaptive system	Not tested on attacks
Huang and Zhu (2020)	Supervised	PBNE, dynamic game	TEP, Time-ordered simulation	APT, user-based, finite options	Dynamic game framework that offers proactive and adaptive approaches to enhance security	Framework's complexity might limit its scalability to larger systems
Meira et al. (2020)	Un-supervised	AE, one-class nearest neighbor, one-class K-Means, IF, one-class scaled convex Hhull, and OCSVM	NSL-KDD and ISCX	DoS, R2L, U2R, probe, brute force	Shows unsupervised methods can effectively detect unknown attacks	Offline learning
Mahdavifar and Ghorbani (2020)	Supervised	DENNES, MACIE	UCI phishing websites and Android malware	Website phishing and Android malware	Efficiently explains the causes of cyber threats and outperforms other explainable algorithms (RF)	Lacks performance of basic DNN

Table 2 (continued)

Reference	Model	Algorithm	Dataset	Attack	Strengths	Weaknesses
Quincozes et al. (2021)	Supervised	GRASP	SWaT-CPS, NSL-KDD, WSN-DS,CICIDS2017	CPS injection, flooding, grayhole, blackhole, TDMA, ARES DDoS, LOIC DDoS, PortScan, R2L, U2R	Model generalizes well across datasets	Relies on offline feature selection
Liu et al. (2021)	Un-supervised (S3) and supervised (A3)	S3, A3	TEP	faults within CPS	Effective at root-cause analysis for both pattern-based and node-based anomalies	S3 method may be computationally intensive
Althobaiti et al. (2021)	Supervised	BBFO, GRU	NSL-KDD 2015 and CICIDS 2017	General anomalies (related works does mention replay attacks)	BBFO improved system efficiency, GRU's makes it suitable for range of CPS	Requires hyperparameter optimization
Vávra et al. (2021)	Un-supervised	LSTM, IF, and OCSVM	Secure Water Treatment (SWaT), ICS network communication	General cyber-attack, anomaly detection	System effectively handles high-dimensional data using PCA	System struggles with interpreting the specific nature of the detected anomalies
Alohali et al. (2022)	Supervised	IFSO that combines RNN, Bi-LSTM, and DBN	NSL-KDD 2015 and CICIDS 2017	General intrusion detection	Feature selection efficiently reduces data dimensionality	Offline learning

Table 2 (continued)

Reference	Model	Algorithm	Dataset	Attack	Strengths	Weaknesses
Ibor et al. (2022)	Supervised	Bio-inspired deep feed-forward, modified genetic search	CICIDS2017 and UNSW-NB15	Brute Force, Heart-bleed, Botnet, DoS, DDoS, Web Attack, Infiltration, Analysis, Back-door, Exploits, Fuzzers, Generic, Recon-nais-sance, Shell-code, Worms	Bio-inspired hyperparam-eter search allows dynamic optimization	Offline model
Liu et al. (2022)	Supervised	FedBatch, CNN-MLP	NSL-KDD	40 attack types divided into five categories: Normal, DoS, Probe, U2R and R2L	Resilience in handling non-IID data	Assumes a uniform distribution of stragglers (might be dynamic in real environments)

Table 2 (continued)

Reference	Model	Algorithm	Dataset	Attack	Strengths	Weaknesses
Kure et al. (2022)	Supervised	DT, NB, RF, KNN	VCDB	Crimeware (R1), cyber espionage (R2), denial of service (R3), everything else (R4), lost and stolen assets (R5), miscellaneous errors (R6), payment card skimmers (R7), point of sale (R8), privilege misuse (R9), and web applications (R10)	Integration of fuzzy set theory and machine learning for proactive risk prediction is innovative	Study does not focus on streaming data
Shi et al. (2023)	Supervised (classification) and unsupervised (novelty detection)	LSTM-AE	Accelerometer data from FFF platform	Void in the STL design, alteration in slicing stage	Effectively captures temporal dependencies	Test parts used in the experiment have simple designs
Intriago and Zhang (2023)	Supervised	HAT with a novel instance selection algorithm	multiclass industrial control system cyber-attack dataset	Line maintenance, short-circuit faults, remote tripping command injection attacks, relay setting change attacks, and data injection attacks	Adapts effectively to evolving data streams	Need for tuning hyperparameters and the potential alteration of the temporal distribution due to instance reordering

Table 2 (continued)

Reference	Model	Algorithm	Dataset	Attack	Strengths	Weaknesses
Alsham-mari (2023)	Supervised	BNN	NSL-KDD (network security dataset)	DoS	Higher performance compared to traditional ML methods	Performance depends on feature selection
Xi et al. (2023)	Un-supervised	ACUDL, dynamic graph update, D-AE	ECG5000, Ar-rhythmia, Satellite, and CIFAR-10	Anomaly detection, general cyber attack	Systems uses a D-autoencoder that uses cor-related and non-correlat-ed features	Might not be scalable
Cai et al. (2023)	Un-supervised	ADAM, KNN	MAWI Working Group Traffic	Volumet-ric DDoS attacks	High accura-cy, low FPR, designed to be scalable	Vulnerability of SDN, only focused on volumetric attacks
Wang et al. (2023)	Un-supervised (OCSVM) and supervised (PSLSTM)	OCSVM, PSLSTM	Lab experiment, 12 raspberry PIs	Cut, virus, tro-jan, scan, intrusion, and heat	Detects both known and unknown attacks	Requires substantial computational resources

Note: Some papers utilize both supervised and unsupervised methods (i.e., Liu et al. 2021; Shi et al. 2023; Wang et al. 2023) which may count twice in Fig. 4. As a result, while there are 21 papers reviewed in Table 2, the total number of papers considered in Fig. 4 is 24

4.1 AAD in industrial control systems (ICS)

4.1.1 Supervised learning

Nakayama et al. (2019) proposed granger causality-based kalman filter with adaptive robust thresholding (G-KART) that learns temporal causal relationships between system components, allowing for the detection of stealthy FDIA attacks that traditional methods might miss. Authors focus on ensuring system security by analyzing interactions between components rather than relying solely on known topological relationships. Each pair of relationships is modeled using kalman filters (KF), which are updated continuously to monitor component states. The adaptive robust thresholding mechanism adjusts the anomaly detection thresholds based on the rolling median of past residuals which makes the system more resilient to noise and gradual data distribution changes. The G-KART model achieved the highest F1 score (0.141) among the tested methods. G-KART also demonstrated a significantly lower false positive rate (0.07) compared to the other models tested.

Pan et al. (2019) proposed a novel threshold-free physical layer authentication (PHY-AUC) method based on ML to enhance the security of wireless industrial CPS in mobile scenarios. Authors discuss that traditional cryptographic methods for message authentication impose high computation burdens, PHY-AUC offers an alternative with lower resource requirements. The previous PHY-AUC methods relay on fixed thresholds however, due to stability issues these iterations are ineffective in mobile environments. The authors presented a solution to this limitation by introducing a ML-based threshold-free PHY-AUC technique that authenticates as a binary classification problem. The method proposed uses

channel matices as inputs for supervised ML algorithms to perform the classification of legitimate and illegitimate messages. Authors tested four ML algorithms, the best performing method was the boosted tree ensemble approach. This model contained 128 dimensional channel matrix input and reached an accuracy of 77% in the multiple-input multiple-output (MIMO) scenario.

Saez et al. (2019) proposed a context-sensitive hybrid modeling framework for cyber-physical manufacturing systems (CPMS) that aims to enhance anomaly detection and diagnosis. It combines physics-based and data-driven models, using sensor data and expert knowledge. The framework employs context-sensitive thresholds for anomaly detection and classification models for diagnosis. Applied to a CNC machine, it improved detection rates from 75% to 94%. The system integrates discrete and continuous states, including global operational state definitions. Physics-based models use fundamental equations, while data-driven models predict outputs with historical data. Hybrid modeling estimates variables like current and voltage, enabling accurate fault diagnosis.

Huang and Zhu (2020) proposed a dynamic game framework to model long term interactions between stealthy attackers and proactive defenders. Authors establish the foundation of the work by describing the model being comprised of two players, player 1 (user) and player 2 (defender). The user's type can be either adversarial or legitimate, which creates uncertainty for the defender with the defender's type being based on their level of sophistication (level of security awareness, detection technique, etc.). The game is structured as a multi-stage interaction, where each stage involves a sequential game with incomplete information. The defender uses an IDS to generate alerts but cannot directly identify user types. Bayesian updates refine beliefs about adversarial and legitimate users. The study derives the perfect Bayesian Nash equilibrium (PBNE) to guide strategies. Sophisticated defenders increased payoffs by 56% and reconnaissance prevention by 41%. The framework effectively models advanced persistent threats.

Mahdavifar and Ghorbani (2020) introduced the deep embedded neural network expert system (DeNNeS) to address the lack of explainability in DL models that are used for cyber-attacks (phishing and malware). The system extracts refined rules from trained DL networks to enhance explainability and decision-making in cybersecurity applications. The framework consists of foundational elements from the matrix controlled inference engine (MACIE) algorithm (adopted from previous works), which acts as an inference justification. Authors utilized the MACIE inference justification (MIJ) algorithm and an extended version that extracts rules out of a trained multilayer deep neural network (DNN) called deep inference justification (DIJ). DeNNeS achieved a 97.2% F1 score on a phishing dataset and 91.1% on a malware dataset. Adaptive moment estimation (ADAM) optimized updates, enhancing classification accuracy.

Quincozes et al. (2021) addressed the issue of cybersecurity in CPS by focusing on feature selection to improve intrusion detection across three CPS layers, perception, transmission, and application. Authors were motivated by CPS integration issues with physical components that face significant security challenges. Authors proposed a method that implements greedy randomized adaptive search procedure (GRASP) with two main phases–construction and local search. GRASP iteratively builds feature subsets and optimizes them to identify the most relevant features for intrusion detection. The GRASP method optimizes both the construction and local search phases to select relevant features across perception, transmission, and application layers. The GRASP method is tested with various classifiers

(Random Forest (RF), naive Bayes (NB), J48, etc.), using datasets that are specific to CPS attack scenarios. Tested with classifiers like RF and J48, GRASP improved F1 scores by 8.29% over traditional methods. RF achieved 99.64% F1 in the application layer, while J48 scored 98.50% in the perception layer. The approach enhanced CPS security across multiple layers.

Liu et al. (2021) introduced a novel data-driven framework for root-cause analysis of anomalies in complex CPS using a spatiotemporal graphical modeling approach based on symbolic dynamics. The framework focuses on discovering and representing causal interactions among subsystems. The authors formulated the root-cause analysis problem as a minimization problem using an inference-based metric and proposed two approximate solutions, sequential state switching (S3) and artificial anomaly association (A3). The effectiveness of these methods was validated using synthetic data and the Tennessee eastman process (TEP) dataset. The S3 method analyzed patterns of subsystem interactions sequentially to identify potential root causes and the A3 method treats root-cause analysis as a classification problem using a neural network (NN) model. Using synthetic and real datasets, the results showed that S3 and A3 could effectively identify faulty nodes, with S3 having a 100% recall, precision, and F1 score, while A3 achieved a recall of 96.7%, precision of 90.6%, and an F1 score of 93.6%.

Althobaiti et al. (2021) introduced a novel cognitive computing-based IDS for industrial CPS that leverages AI to address security issues. The approach uses a binary bacterial foraging optimization (BBFO) for feature selection and a gated recurrent unit (GRU) for classification. BBFO is inspired by the foraging behavior of bacteria, where bacteria move towards optimal solutions in the search space. This approach aims to reduce the dimensionality of the dataset, improving computational efficiency and model accuracy. Nesterov-accelerated adaptive moment estimation (NADAM) is employed to optimize the hyperparameters of the GRU. Applied to industrial datasets, the system achieved 98.45% detection accuracy. The method combines high precision with reduced computational costs.

Vávra et al. (2021), presented a comprehensive approach to anomaly detection in ICS using ML algorithms. Recognizing the critical role of ICS in modern society and the increasing risks posed by cyberattacks, the authors develop an AAD system designed to address key challenges, including the detection of unknown attacks, scalability, adaptability, high false alarm rates, and computational complexity. The system integrates artificial neural networks (ANN), long short-term memory networks (LSTM), isolation forest (IF), and one-class support vector machines (OCSVM), each adapted for semi-supervised learning. This approach enables the system to identify anomalies by learning from normal operational data and flagging deviations as potential cyberattacks. The use of semi-supervised learning is particularly important for detecting unknown attacks that may not be present in the training data. The paper also emphasizes the importance of data preprocessing and feature selection, noting that issues like missing values, feature scaling, and high dimensionality of data can significantly impact the performance of ML models. Techniques such as principal component analysis (PCA) are employed to reduce the dimensionality of the dataset while retaining the most critical information, thereby enhancing the system's efficiency and scalability. The paper concludes that the proposed system effectively balances detection accuracy with computational efficiency, particularly when the IF algorithm is optimized using genetic algorithms (GA). The system balances detection accuracy and computational demands, achieving strong performance metrics.

Alohali et al. (2022) proposed an AI-enabled multimodal fusion-based intrusion detection system (AIMMF-IDS) that employs an improved fish swarm optimization (IFSO-FS) technique for feature selection by using the Levy Flight concept to enhance searching capability and avoid local optima problems. The system integrates recurrent neural network (RNN), bi-directional LSTM (Bi-LSTM), and DBN models for multimodal fusion. Applied to NSL-KDD and CICIDS datasets, it achieved precision, recall, and F1 scores above 95%. The model addresses intrusion detection challenges with robust performance.

Ibor et al. (2022) presented a novel hybrid approach to predict cyberattacks on CPS communication networks. The proposed method leverages a bio-inspired hyperparameter search technique to improve the structure of a deep neural network (DNN). The GA operates by generating an initial population of neural network (NN) structures (chromosomes), evaluating their performance using a fitness function, and then applying selection, recombination, and mutation to generate improved structures over multiple generations. Tested on CICIDS2017 and UNSW-NB15 datasets, the method achieved 99.81% and 80.10% accuracy, respectively. The approach optimizes neural network architecture for better attack prediction with minimal false positives.

Liu et al. (2022) investigated enhancing cybersecurity for maritime transportation systems (MTS) using IoT technology. Authors proposed a federated learning-based IDS called “FedBatch" that employs a hybrid convolutional neural network (CNN) multilayer perceptron (CNN-MLP) model. The model addresses the issue of maintaining data privacy and handling the straggler problem (delays in data/model updates due to unstable communication in maritime settings). It adapts to unstable communication environments using dynamic aggregation methods. Tested on independent and identically distributed (IID) and non-IID datasets, FedBatch achieved 88.1% accuracy, outperforming traditional federated learning methods.

Kure et al. (2022) presented a unified approach to cybersecurity risk management (CSRM) for CPS by integrating fuzzy set theory and ML classifiers. The focus is on predicting risk types, assessing asset criticality, and evaluating the effectiveness of existing controls. Fuzzy set theory is applied to assess the criticality of CPS assets, considering security factors such as confidentiality, integrity, availability, accountability, and conformance. Multiple ML classifiers, including k-nearest neighbors (k-NN), decision tree (DT), RF, and NB, are used to predict different risk types like DoS, cyber espionage, and crimeware. DT achieved 93% accuracy in predicting risk types. Fuzzy logic improves asset assessment, supporting proactive risk management.

Shi et al. (2023) proposed a ML-driven online side-channel monitoring approach that utilizes an LSTM autoencoder (AE) for detecting unintended alterations in the additive manufacturing (AM) process. Authors were motivated by the growing vulnerabilities in AM due to potential cyber-physical attacks. Traditional monitoring methods fail to detect internal alterations in AM parts, which can severely compromise the functionality and mechanical properties of the product. Both supervised and unsupervised monitoring schemes were implemented, with experimental validation conducted on a fused filament fabrication (FFF) platform equipped with accelerometers. The AE, consisting of an encoder and a decoder, where the encoder reduces high-dimensional input data into a compact latent representation using an encoding function implemented with LSTM layers, which effectively capture temporal dependencies in sequential sensor data. The decoder reconstructs the original input from the latent representation using a decoding function. The reconstruction error quantifies

the difference between the original input and the reconstructed data, serving as a measure of how well the model represents the input. Two cases are considered for this investigation, Case 1 involves inserting a void into the design geometry at the STL stage, while Case 2 alters layer thickness at the slicing stage. In supervised monitoring, the proposed method achieved high F1 scores of 0.95 in Case 1 and 0.96 in Case 2, outperforming traditional methods. In unsupervised monitoring, the LSTM-AE with OCSVM exponentially weighted moving average (OCSVM-EWMA) demonstrated a low false alarm rate (0.09 in Case 1 and 0.03 in Case 2) and fast attack response times (4.6 and 7.4 samples, respectively).

Intriago and Zhang (2023) introduced a Hoeffding adaptive tree (HAT) classifier combined with an instance selection algorithm to detect cyber and non-cyber contingencies in real-time for cyber-physical power systems (CPPS). Authors are motivated based on challenges presented in continuous monitoring of CPPS through wide-area monitoring, protection, and control (WAMPAC) systems that handle high-velocity and unbounded data streams from devices like phasor measurement units (PMUs). Authors addressed this issue by creating a streaming learning classification system that adapts to evolving data streams. They proposed a three-stage instance selection process involving reordering, resetting outdated data, and dynamic window size selection. The classifier outperformed existing methods in six case studies, achieving over 99% accuracy on multiclass datasets. The HAT stream instance selection (HAT+SIS) classifier achieved over 99% accuracy in the multiclass dataset, maintained this accuracy even as the data evolved, and outperformed HAT drift detection method (HAT+DDM) and Hoeffding tree (HT+DDM). These methods experienced significant accuracy drops around 4,000 instances. HAT+SIS also processed approximately 5,000 instances in 43 s and maintained a stable model size of around 196 KB, demonstrating both efficiency and scalability for real-time applications.

Alshammari (2023) proposed a Bayesian neural network (BNN) architecture for evaluating the statistical features of DoS attacks. In the investigation, the NSL-KDD dataset is used which is a comprehensive network security dataset. The paper emphasizes selecting features with the highest predictive power while minimizing redundancy which is established with a heatmap of feature mean values. The BNN applies the local markov property to forecast event probabilities and represent probabilistic relationships between different variables using a directed acyclic graph. The BNN is constructed with multiple nodes, each of the nodes represents a random variable with the edges showing conditional dependencies between variables. The preprocessed data is trained using the BNN to learn patterns and relationships in the network traffic. The author states that the BNN demonstrated a testing accuracy of 97.5 % with other metrics highlighting that the BNN was capable of generalizing well. When compared to accuracies from DT (78%-89%) and ANN (87%-89%), the BNN outperforms these traditional networks in terms of identifying anomalies.

Wang et al. (2023) introduced a novel integrated ML and DL approach for real-time attack detection and identification. The authors present a two-stage solution involving a OCSVM for detecting whether a CPS is under attack, followed by a pairwise self-supervised long short-term memory (PSLSTM) model for identifying the attack type. The method is designed to detect known attack types and discover unknown new attacks. The two-stage learning approach begins with the OCSVM model that is trained exclusively on data representing the normal state of the CPS. The OCSVM model functions by defining a hyperplane that separates normal data points from potential outliers. The second stage begins once the attack is detected. PSLSTM is used to identify the specific attack type and consists of several LSTM networks that effectively convert a multi-class problem into a

binary classification task. The number of pairwise models to be trained is calculated from a given K known attack types and $\frac{K(K-1)}{2}$ LSTM models. The OCSVM-PSLSTM method demonstrated high-quality performance with real-time attack detection and identification by achieving an average accuracy of 99.7% for identifying known attack types and 97.2% with discovering unknown attacks.

4.1.2 Unsupervised learning

Mitchell and Chen (2013) developed a mathematical model to assess the survivability of mobile CPS (MCPS) using dynamic voting-based intrusion detection. They employed a stochastic Petri net (SPN) model to analyze trade-offs between energy conservation and intrusion tolerance. The system dynamically adapts to changing system states and environmental conditions. As the proportion of compromised nodes or energy levels change, the model adjusts the intrusion detection interval T_{IDS} and the number of detectors in real time to maintain optimal system performance. The dynamic voting-based technique continuously updates its parameters without requiring pre-labeled data, allowing the system to respond to attacks and energy constraints as they occur. The results demonstrate that there is an optimal T_{IDS} that maximizes the mean time to failure (MTTF) of the MCPS by balancing energy consumption and intrusion tolerance. Authors found that an optimal value of T_{IDS}=160 s with 5 detectors provided the highest MTTF for their reference system. The model's predictions were validated through simulations, showing a close match with theoretical results, with only a 4.60% to 7.64% mean percentage error.

Meira et al. (2020) examined six unsupervised algorithms; AE, one-class nearest neighbor, IF, one-class $\mathcal{K}$-means, one-class scaled convex hull, and OCSVM, on two public datasets (NSL-KDD and ISCX) for anomaly detection in cybersecurity. Data was preprocessed by normalization using $\mathcal{Z}$-score and Min Max. The one-class scaled convex hull achieved the highest AUC on the NSL-KDD dataset with an AUC value of 85.30%. The one-class nearest neighbor performed best on the ISCX dataset with an AUC of 95.20%. Overall, the one-class nearest neighbor, one-class scaled convex hull, and OCSVM demonstrated the best performance across both datasets.

Xi et al. (2023) proposed the adaptive-correlation-aware unsupervised DL (ACUDL) which addresses the challenge of detecting anomalies in high-dimensional, noisy, and unlabeled data. The core innovation of ACUDL is its use of a dynamic graph structure to represent and update the implicit correlations among data points, which is critical for accurately capturing the underlying relationships in CPS data. ACUDL begins by constructing an initial directed graph using the KNN algorithm, where each node in the graph represents a data point, and edges represent the correlations between these points. This graph is then dynamically updated during training to reflect changes in the data, using an adaptive mechanism that adjusts the graph structure based on the training loss. The model also incorporates a dual-AE (D-AE) framework that separately encodes the original non-correlation features, the correlation features extracted from the graph, and a decoder. These features are then fused and passed through a gaussian mixture model (GMM) to estimate the anomaly energy, which is used to detect anomalies. Through extensive experiments on various CPS datasets, including scenarios like smart healthcare systems (SHS) and intelligent cruise control systems (ICCS), the authors demonstrate that ACUDL significantly outperforms existing DL-based anomaly detection methods. Experiments on CPS datasets demonstrated its supe-

riority, achieving AUC scores of 73.3% and 87.5%, with strong F1 scores and precision. ACUDL effectively handles high-dimensional, noisy data.

Cai et al. proposed an adaptive distributed denial-of-service (DDoS) mitigation scheme for software-defined networking (SDN) (Cai et al. 2023). Authors proposed the adaptive DDoS attack mitigation (ADAM) scheme that combines information entropy and unsupervised anomaly detection methods to detect both known and unknown DDoS attacks in software-defined CPS (SD-CPS). Motivated by the increasing vulnerability of CPS to DDoS attacks due to insecure or outdated components, the authors highlight the limitations of traditional defense mechanisms that rely on static thresholds or predefined attack signatures. ADAM operates through three stages: nominal, detection, and mitigation. In the nominal stage, a "nominal profile" of normal network traffic is created by sampling traffic and calculating entropy vectors for features such as IP addresses and port numbers. ADAM achieves a high accuracy of 99.13% on average in mitigating various DDoS attacks, with a significantly reduced false-positive rate compared to existing methods. This suggests that ADAM is a scalable and effective solution for SD-CPS environments.

Table 2 summarizes key attributes of the papers reviewed in Sect. 4.1.

4.2 AAD in vehicle networks

4.2.1 Supervised learning

van Wyk et al. Van Wyk et al. (2019) developed an anomaly detection approach that combines CNNs and a Kalman filter (KF) with a χ^2 detector to identify anomalous behavior in connected and automated vehicles (CAVs). They use a sliding window approach to focus the analysis on the latest observations. Here, the input to the CNN module is a stream of images from a continuous feed of raw sensor data during a CAV trip. They trained a separate CNN model per sensor using labeled images. They used the "OR" logical operation of the outcomes in each of the sensors to determine if anomalous readings are detected across all sensors. To improve further on detection, the output of the CNN feeds an adaptive KF with a χ^2 detector for further examination for anomaly detection. KF has prediction and update phases. In the prediction phase, the KF produces estimates of the the time series including their uncertainties. Once the predictions of the next samples are calculated, estimates are updated using a weighted average with greater weights to estimate with greater certainty. KF works recursively. They tested their approach using data from the Safety Pilot Model Deployment (SPMD) program (Bezzina and Sayer 2014) that demonstrate CAVs in action. In particular, they focused on analyzing time series from three sensors: in-vehicle speed, GPS speed, and in-vehicle acceleration for a vehicle with a trip length of 2,980 s. As the original data contains no anomalies, authors inject synthetic anomalies, including instant (simulated as a random Gaussian variable), constant (a temporarily constant observation that is different from the normal), gradual drift (by linearly increasing a set of values to the base value of the sensors), and bias (a temporarily constant offset from the baseline sensor readings).

Feng et al. (2020) proposed an efficient drone hijacking detection method that consumes inertial measurement unit (IMU) (i.e., gyroscope and accelerometer) and GPS (i.e., longitude and latitude) data. The proposed method used the eXtreme Gradient Boosting (XGBoost) algorithm to mine the relationship between IMU and GPS data using real-time

data samples to decide if the drone has been hijacked or not. The model is first trained offline where parameters are optimized using a GA. In the deployment state, the same training parameters are used to update the model onboard. Experiments are conducted on a real quadcopter with an off-the-shelf multi-core embedded board and an autopilot sensor board. Prediction correctness in each sample time was reported as 96.3% and 100% in hijack and normal scenarios. As the proposed model is deployed in online fashion it can achieve 100% detection correctness just after 1 s after the hijack starts.

Alsulami et al. (2023) proposed an intelligent intrusion detection systems (IIDS) for autonomous vehicle-cyber physical systems (AV-CPS) that focuses on transfer learning. Specifically, the proposed method used eight pre-trained CNNs, including InceptionV3, ResNet-50, ShuffleNet, MobileNetV2, GoogLeNet, ResNet-18, SqueezeNet, and AlexNet. By leveraging pre-trained models to enhance the performance of IIDS, authors target the detection of anomalous communications in the controller area network (CAN) bus having an effect on the connected physical components of AVs. Authors' simulation setup include a self-driven car system consisting of a lead and an ego vehicle (self-driving car). In an ideal situation, the ego vehicle maintains its distance from the lead vehicle using the adaptive cruise control system. Authors simulate a CAN communication network environment using Simulink and produced a dataset consisting of (1) position of the ego vehicle, (2) velocity of the ego vehicle, (3) position of the lead vehicle, and (4) velocity of the lead vehicle. Note that for the data be processed by the CNNs, time series representations of captured signals are transformed into a 2-dimensional representation (i.e., images). They found that GoogLeNet performed best achieving 99.47% on F1 score metric.

4.2.2 Reinforcement learning

Mowla et al. (2020) proposed an adaptive federated reinforcement learning-based jamming attack defense strategy to protect flying ad-hoc networks (FANETs), i.e., a decentralized communication network unmanned aerial vehicles (UAVs). They focused on a model-free Q-learning mechanism with adaptive exploration-exploitation epsilon greedy policy, directed by an on-device federated jamming detection mechanism. Q-learning learns a policy maximizing total rewards based on trial-and-error. That is, a UAV client receives a negative reward if it is moved closer to the jammer location. Here, an epsilon-greedy policy is used to balance the outcome of exploration and exploitation opportunities. They showed that the proposed adaptive federated RL-based approach performed better spatial retreat defense strategies.

Table 3 summarizes key attributes of the papers reviewed in Sect. 4.2.

4.3 AAD in power grids

4.3.1 Supervised learning

Cui et al. (2021) proposed a hybrid approach combining self-adaptive mathematical morphology (SAMM) and time frequency (TF) techniques to authenticate source information on distribution synchrophasors (DS) within microgrids at near-range locations. Their proposed method is intended to deter "source ID mix" data spoofing attacks on DS as they threat to the security of the power grid. This attack can manipulate source information of DS without

Table 3 Summary of AAD methods in vehicular applications

Reference	Model	Algorithm	Dataset	Attack	Strengths	Weaknesses
Van Wyk et al. (2019)	Supervised	CNN, KF	Public real CAV data	Instant, constant, gradual, bias drifts	CNN and KF approach combines strengths	Tests of simulated attacks
Feng et al. (2020)	Supervised	XGBoost	Real prototype quadrotor drone	GPS spoofing attacks	Model further trained onboard	Limited number training scenarios
Mowla et al. (2020)	RL	Q-learning	Simulated FANET topology	Constant, random, and reactive jamming	Reduce the number of route jammer location hop counts	Small-scale simulation
Alsulami et al. (2023)	Supervised	CNN	AV simulation	False data injection	Transfer learning allows to use pretrained models	Limited to false data injection attacks

changing the measurement values. Manipulating the source information on DS has an effect on critical synchrophasor-based control and applications including wide-area damping and control and disturbance localization. SAMM allows for adaptive regulation of synchrophasors variations which represent local environmental characteristics. In addition, TF mapping is used to extract frequency-related features from the regulated synchrophasors variation. They used a RF classifier to correlate the extracted signatures with the source information based on the TF analysis. Their proposed approach consists of three steps including (1) a high-pass filter to extract frequency variation on the original DS data; (2) an integrated SAMM-TF used to extract useful features from the synchrophasors variation; and (3) the integration with a random forest classifier for source integration. The SAMM approach adaptively preserves significant peaks in frequency so that distinctive signatures can be extracted for source authentication. They validated their results using distribution synchrophasors from multiple small geographical scales validating the proposed methodology.

Khan et al. (2021) proposed a privacy-conserving based intrusion detection (PC-IDS) for contemporary smart power systems (SPNs) using a hybrid ML approach. The proposed framework consist of two main components: data pre-processing and intrusion detection. Data pre-processing entails attribute/feature mapping, reduction, and normalization. This allows to process diverse types of attributes such as numerical and categorical features. The intrusion detection module comprises stages of training and detection. Specifically, it uses a PNN and consumes several types of regular and malicious patterns to enhance classification performance. Particle swarm optimization (PSO) is used to select the hyperparameters of the PNN model. The performance of the proposed framework is evaluated in two commonly available datasets: power system (Morris 2013) and UNSW-NB15 (Moustafa and Slay 2015) datasets. Their experimental evaluation shows the effectiveness of the proposed framework to protect data from SPNs and determine anomalous behavior in terms of traditional evaluation metrics.

Jiao et al. (2022) proposed a cyberattack-resilient load forecasting approach based on robust regression, i.e., adaptive least trimmed squares (ALTS) (Bacher et al. 2016). Authors assume that adversaries alter load entries in the training data so that the estimated regression coefficients become inaccurate resulting in forecasts that may lead to poor decision making. They used two different attack types: random and ramping attacks. In random attacks, a randomly selected proportion of the training data is scaled by a random factor follow-

ing, for example, a normal distribution. In ramping attacks, many single attack intervals are injected parametrized by starting attack point and length. They injected these attacks to alter the GEFCom2012 dataset (Hong et al. 2014). Their robust approach focused on estimating robust estimators for the coefficients in the regression models based on M-estimation (Huber 1992)—a generalization of the maximum likelihood estimation. The M-estimator is obtained through the iterative re-weighted least squares (IRLS) algorithm. The core of their adaptive algorithm is based on least trimmed squares (LTS) which is a robust alternative to ordinary LS when dealing with linear regression problems. LTS minimizes the sum of square residuals over a subset of the whole datapoints by excluding a proportion of p data points whose residuals are the largest in magnitude. Thus p is a tuning parameter that leads to the ALTS method. This provides robustness to potential outliers. They compute goodness of fit using mean absolute percentage error. Five methods including, LS, M-Huber, M-bisquare, L_1, and ALTS ere used to fit the model and perform forecasting in the validation dataset. A comparison analysis using the GEFCom2012 dataset suggests that the ALTS method is robust against to random and ramping attacks both when the proportion of attack data is high and the robustness does not decrease as attack data proportion increases.

Ding et al. (2022) proposed a data-driven security situational awareness framework to secure power systems. Their proposed framework focused on an adaptive honeypot architecture for capturing system logs and network traffic in distributed fashion. Specifically, they deployed 10 honeypots around the world by simulating industrial control devices in power systems to lure attackers. Their deployed honeypot architecture includes multiple industrial protocols including Modbus, Siemens S7-Comm, Guardian ast, Kamstrup 382, Bacnet, Http and Ipmi. The honeypot architecture was deployed using the Alibaba cloud environment. Security incidents are detected by modeling the captured attack traffic from honeypots and the security scanning system. In doing so, they constructed a security situation graph based on traffic logs by modeling IP instances as nodes in a graph. They annotate the nodes in the graph using word2vec (Mikolov et al. 2013) by processing attributed information of IPs. Their detection algorithms leverage a graph convolutional network (GCN) to detect malicious IPs. Under this approach, the GCN learns the features of malicious IPs in a supervised fashion. They performed experiments to evaluate performance of TCP SYN probe in their deployed scanning system. They tested two attack configurations: single port scanning and multi-port scanning.

Table 4 summarizes key attributes of the papers reviewed in Sect. 4.3.

4.4 AAD in the Internet of Things (IoT)

4.4.1 Supervised learning

Li et al. (2019) introduced a statistical learning based anomaly detection technique that monitors the operation of IoT devices to detect possible cyberattacks and malicious activities. In particular, they used time series derived from system statistics including CPU usage, memory consumption, and network throughput, among others, to model normal behavior. To do so, they simulated a real IoT system operation made of 12 beagle bone black (BBB) connected to a router in a LAN topology. In particular, they ran a program that samples random data with a fixed interval and processes a variety of signal processing, storage, compression, and transmitting operations. They injected simulated cyberattacks including

Table 4 Summary of AAD methods in power grid applications

Reference	Model	Algorithm	Dataset	Attack	Strengths	Weaknesses
Cui et al. (2021)	Supervised	SAMM and TF	Photovoltaics nodes	FDIA	Adaptation to synchrophasors variations	Lack of continual learning
Khan et al. (2021)	Supervised	PNN	Phasor data concentrator and network data	FDIA	PNN train faster than MLP	Lack of continual learning
Jiao et al. (2022)	Supervised	ALTS	Global energy forecasting competition	FDIA	Robust against random and ramping attacks	Only focus on linear regression
Ding et al. (2022)	Supervised	GCN	TCP SYN probes	Information leakage, security vulnerability, DoS attack, APT attack, SQL injection, Malware infection, remote attack	Graph modeling that fuses inside and outside cyber threat incidents	Lack of continual learning

unauthorized access, port scan, virus, and flood. They trained LR, NN, and recurrent neural network (RNN) classifiers to predict normal system behaviors. These models were trained using different window sizes showing that there is a threshold for which the MAE starts getting diminishing returns. Their results showed that the NN perform the best at the expense of higher computational complexity. The thresholding for deciding if a particular data sample is an anomaly or not is performed through local outlier factor (LOF), cumulative statistics thresholding (CUSUM), and adaptive online thresholding (AOT). They showed that the method using AOT perform better than LOF and CUSUM based on F1 score evaluation metric.

Gopalakrishnan et al. (2020) introduced a deep learning based traffic prediction framework with offloading mechanism and cyberatack detection (DLTPDO-CD). Their proposed approach is composed of three major processes involving traffic prediction, data offloading, and attack detection. In doing so, it includes first a BiLSTM based traffic prediction process to enable proficient data offloading. Considering a double LSTM is meant to enhance learning long-term dependencies as it enhances the accuracy of the detection. Second, an adaptive sampling cross entropy technique (ASCE) is incorporated to maximize network throughput by deciding offloading users from the network. Lastly, for the detection of cyberattacks in mobile edge computing, they used a DBN optimized using a barnacles mating optimizer (BMO). Their approach was tested on simulated data showing better performance over compared methods under different dimensions.

Bibi et al. (2022) proposed an efficient and self-learning autonomous multivector threat intelligence and detection mechanism to proactively defend IIoT networks. Their approach used a convolutional LSTM2D (Cu-ConvLSTM2D mechanism) being highly scalable with self-optimized capabilities to detect diverse and dynamic variant of IIoT threats. Cu-ConvLSTM2D is a recurrent layer similar to LSTM, but the internal matrix multiplication is exchanged with convolution operations. They evaluated their proposed framework on a Kitsune surveillance network intrusion dataset (Mirsky et al. 2018) comprising 21 million

instances of varying attack patterns and prevalent threat vectors. The proposed technique outperforms current contemporary DL-driven architectures and existing benchmarks.

Yazdinejad et al. (2023) proposed a novel design and implementation of a secure and intelligent fuzzy blockchain framework. Their proposed framework focus on three layers: IoT, blockchain, and intelligent fuzzy layers. The IoT layer hanldes smart devices that communicate with each other in the blockchain environment. The blockchain layer handles IoT device management ensuring safe channels for the transmission of data and transactions between IoT devices. In their intelligent fuzzy layer, they used a threat detection in the blockchain layer. Their threat detection module is based on a fuzzy DL model that uses a fully connected network based on fuzzy neurons to output the aggregation of classification results of several classifiers. In conjuction, an adaptive neuro-fuzzy inference system (ANFIS) model is used to design an optimal fuzzy system for threat detection in IoT networks. The ANFIS model estimates input membership functions and output modified membership functions. Finally, a fuzzy control system module leverages previous inputs in both blockchain and IoT layers to feed a fuzzy control system module to arrive at complex decision making. Their proposed framework was tested for threat detection in the Ethereum blockchain (Jung et al. 2019; Al-E'mari et al. 2020) and the NSL-KDD (Bala and Nagpal 2019) datasets for blockchain-enabled IoT networks. They verified the efficiency in both blockchain and IoT network sides using a variety of evaluation metrics.

Dey et al. (2023) proposed a hybrid feature selection scheme combining statistical filter approaches including χ^2, Pearson's correlation coefficient, and mutual information combined with a non-dominant sorting GA (NSGA-II) metaheuristic for optimizing feature selection. Specifically, filter selection methods are used to rank features based on their importance and subsequent initialization of NSGA-II allowing faster convergence to the solution. NSGA-II belongs to the evolutionary algorithms class focusing on selection, crossover, and mutation steps. Resulting populations are sorted from top to bottom without losing reasonable solutions. After selecting the most relevant features for the classification task, a SVM classifier is used in the analysis. The proposed framework is tested from a publicly available network traffic dataset (ToN-IoT) collected at a large-scale and realistic network from the Cyber Range and IoT Labs at the School of Engineering and Information Technology at UNSW Camberra (Booij et al. 2021). ToN-IoT covers a variety of cyberattacks inclusing ransomware, DoS, and DDoS. The proposed method reached optimal performance (99.48% accuracy) with only 13 features.

Yazdinejad et al. (2023) proposed an ensemble DL model that combines LSTM and AEs to detect anomalous activities in IIoT. Output data is classified as normal or abnormal via a DT after inspecting the reconstruction error between the input and the output layer of the LSTM AE. The proposed model is evaluated in two real IoT datasets, i.e., gas pipeline and secure water treatment, which are imbalanced and have temporal dependencies. The proposed framework outperforms conventional classifiers. Despite being a promising approach for detecting anomalies in time series data, the use of LSTM requires significantly higher training than simpler models.

Jullian et al. (2023) implemented a distributed framework based on DL to detect different source of vulnerabilities in the IoT. Their approach consists of mainly four stages: data treatment and preprocessing, DL model training and testing, distributed framework deployment, and attack detection and classification. Their proposed approach was tested on two different datasets: the BoT-IoT dataset addresing specific attacks of IoT environments

and the NSL-KDD dataset to broaden the types of cyberattacks. Due to huge imbalances in the proportion of attack vs. benign samples, they applied undersampling based on the large number of records. A standard normalization procedure is applied to both datasets to prevent models overfitting. Once the datasets are preprocessed, they evaluated two different models: a feed forward NN (FFNN) and a LSTM. Their proposed approach used a federated learning architecture to train models in a distributed fashion relying on the communication between fog and central servers. To prevent overfitting during the training procedure, they ran an optimization procedure to select the best combination of hyperparameters to achieve best evaluation metrics inclusing accuracy, precision, and recall. Their proposed distributed framework was found effective in different types of attacks achieving an accuracy of up to 99.95% across the different setups.

Basati and Faghih (2023) introduced an intelligent IDS framework based on an asymmetric parallel AE able to detect various attacks in IoT networks. In their preprocessing stage, they transformed 1D traffic feature vectors into 2D feature vectors with equal width and height. To extract features from more distant neighbors and association among long-range features they integrate dilatated convolution (Yu and Koltun 2015) with self-attention (Vaswani et al. 2017). Their approach called APAE contains two asymmetric AEs in parallel, each of them containing three successive layers of convolutional filters. APAE was evaluated in three popular public datasets named UNSW-NB15, CICIDS2017, and KDDCup99. Results showed that the proposed framework offers superior results than the state of the art.

Gupta et al. (2023) developed a hierarchical federated learning (HFL) anomaly detection approach to address security and data privacy concerns in the context of vehicular IoT (V-IoT). By creating a digital twin of an intelligent transportation environment, they leveraged a comprehensive virtual replica for detecting malicious activities using an anomaly detection model. To further expand the capabilities of federated learning (FL) for multi use scenarios, they developed a FL approach that allows the aggregation of gradients at multiple levels enabling the participation of multiple entities. Their proposed framework has six phases: (1) initial phase, where smart vehicle data collection begins; (2) functional phase, where supplementary data from the external environment provide context; (3) analytic phase, where a digital twin is created for each vehicle in the V-IoT and data mining algorithms are applied to previously collected data; (4) identifying anomaly phase, where anomaly detection algorithms based on LSTM are trained to distinguish between normal and abnormal patterns; (5) collaborative phase, where the output of multiple anomaly detection algorithms is processed to improve evaluation metrics; and (6) reporting and decision phase, where anomalies are reported to relevant stakeholders.

4.4.2 Unsupervised learning

Yasaei et al. (2020) proposed an adaptive context-aware anomaly detection method for securing IoT sensor data intended to run on a fog computing platform. Their approach is based on a sensor association algorithm that generates fingerprint of sensors and then cluster them to extract the context of the system. Context generation is completed by extracting the binary fingerprinting sensor values. As sensors that are affected by an event are expected to have similar patterns in their fingerprints, a clustering algorithm based on minimizing the distance between binary codifications and the Hamming distance is used. By relying on contextual information, they used a LSTM neural network and a Gaussian consensus esti-

mator to identify the source of anomalies. In particular, for each cluster of sensors, a LSTM neural network produces a set of predictive models for each cluster. A multivariate Gaussian estimator is used to help modeling whether reconstruction errors between the real and predicted values match with the system's normal behavior. To infer the source of the anomaly a consensus algorithms checks the consistency of sensor behaviors within clusters and across clusters. Finally, to adapt the inference models with respect to concept drift, their proposed approach allows two levels of update: complete (where all the modules are retrained in order) and partial (which only retrains the predictor model). The adaptation decision is triggered based on changes in the distribution of the data. They tested their approach on the environmental training center waste water plant in Riccione (Giannoni et al. 2018) with synthetic anomalies.

Gyamfi and Jurcut (2022) proposed a lightweight network intrusion detection system (NIDS) to secure industrial IoT (I-IoT) relying on an online incremental support vector data description (OI-SVDD) anomaly detection on the IIoT devices and an adaptive sequential extreme learning machine (AS-ELM) on a multiaccess edge computing (MEC) server. In their design, the OI-SVDD model is placed on the IIoT device and the AS-ELM model is placed on the MEC server at the edge network to perform deep network attack duties. Their OI-SVDD is based on using an incremental learning scheme to add samples to the training function process consisting only of support vectors at each stage. During AS-ELM, a training dataset is utilized during the initialization phase and then testing data is processed chunk by chunk. Authors tested their proposed NIDS on the UNSW-NB15 dataset and a self generated dataset showing effective performance for detecting intrusions in a realistic IIoT environment.

Li et al. (2022) proposed an adversarial unsupervised domain-adaptive regularization based on a recursive feature pyramid CNN (RFP-CNN) to detect IoT attacks more effectively. As a first step, they developed a recursive feature pyramid network (FPN) architecture (Arrington et al. 2016) to extract high-level features of the RFP-CNN. The FPN focused on a hierarchical approach to predict attack features from multiple scales. Subsequently, to enable transfer learning from fewer attack samples, they developed an unsupervised domain adaptive regularization model. They showed the potential of their approach for anti-noise performance and short running time. Their proposed framework is validated using four intrusion detection datasets used in the IoT space. In doing so, they constructed few shots datasets from the ICSX2012FS and CICIDS2017FS datasets sampling a few attack samples.

Table 5 summarizes key attributes of the papers reviewed in Sect. 4.4.

4.5 AAD in smart grids

4.5.1 Supervised learning

Adhikari et al. (2017) proposed a cyber-power event and intrusion detection system (EIDS) that can be used for multiclass or binary classification of traditional power system contingencies and cyber-attacks. In doing so, they process continuous streams of high speed data from wide area monitoring systems (WAMS) and used a HAT augmented with the drift detection method (DDM) and adaptive windowing (ADWIN) that classifies traffic in real-time. HAT data stream mining is based on the idea of concept adapting very fast DTs (CVFDTs). In particular, HAT creates DTs from the data stream and updates the three after

Table 5 Summary of AAD methods in IoT applications

Reference	Model	Algorithm	Dataset	Attack	Strengths	Weaknesses
Li et al. (2019)	Supervised	LR, NN, RNN	Simulated IoT network	Unathorized access, port scan, virus, flood	Online learning capability with adaptive threshold	Tested on simple DoS attacks
Yasaei et al. (2020)	Unsupervised	LSTM, Multivariate Gaussian estimator	IoT testbed	Simulated data injection attacks	LSTM computational complexity	Lack of continual learning
Gopalakrishnan et al. (2020)	Supervised	BiLSTM, ASCE, DBN	Simulated traffic data	Not specified	Traffic prediction, data offloading, and cyberattack detection	Lack of continual learning
Bibi et al. (2022)	Supervised	ConvLSTM2D	Network traffic	MitM, DoS, botnet malware, and Recon	Catch sophisticated threats and attacks	Lack of continual learning
Li et al. (2022)	Unsupervised	RFP-CNN	Network traffic	Networking attacks	Quick inference	Lack of continual learning
Gyamfi and Jurcut (2022)	Unsupervised	OI-SVDD and AS-ELM	Network traffic	Networking attacks	Online learning capability	Rely on a MEC server to operate effectively
Dey et al. (2023)	Supervised	NSGA-II	Network traffic	Ransomware, DoS, and DDoS	Effective feature selection	Lack of continual learning
Yazdinejad et al. (2023)	Supervised	LSTM-AE	IIoT traffic data	Network operation under attack	Temporal dependencies modeling	Lack of continual learning
Yazdinejad et al. (2023)	Supervised	Fuzzy DL	Blockchain and IoT	Phishing and Ethereum fraudulent transactions	Tight integration between control/data planes	Lack of continual learning
Jullian et al. (2023)	Supervised	FCNN and LSTM	IoT network data	DoS, DDoS, Keylogging, Data Theft	Distributed environment	Lack of continual learning
Basati and Faghih (2023)	Supervised	Asymmetric parallel autoencoder	Network data	Networking attacks	Long-range feature extraction	Lack of continual learning
Gupta et al. (2023)	Supervised	HFL and LSTM	Case scenario	Not specified	Privacy preservation through FL	Lack of continual learning

inspecting each sample. HAT does not require samples to be stored in memory as nodes in the tree hold holds rich information to perform classification. DDM compared the statistics of two windows to detect concept drifts when the number of errors increases beyond a threshold. ADWIN is a parameter free adaptive size sliding window technique that is used to detect concept change and trigger model revision. DDM and ADWIN together enhances

HAT's ability to detect change and update the model. Experiments on a wide area measurement system with hardware in the loop testbed with simulated attacks showed that the combined approach of HAT, DDM, and ADWIN provides greater classification accuracy along with a small memory footprint and fast evaluation.

Wang and Govindarasu (2020) presented a data-driven anomaly detection and adaptive load rejection scheme within a decentralized system integrity protection (SIP) for smart grids. In doing so, they leveraged a SVM embedded layered DT (SVMLDT) that arrives at a decision based on the consensus among all interconnected agents. The SVMLDT works by segregating the training dataset into subsets based on all nominal features and then reducing the dimensionality of the feature space. Then SVMLDT applies a DT based SVM (DTSVM) for supervised classification. The proposed framework responds adaptively to DoS attacks by separating the multi-agent system into several interconnected subgroups so that within one subgroup, the real-time load profiles can be still shared globally. They used a real load rejection SIP scheme adopted by salt-river project to fir in the IEEE 39-bus model as a study case. Their results show that the proposed SIP can effectively detect anomalous grid operation states and then adjust its actions accordingly to adapt to the under attack situations.

Acosta et al. (2020) proposed an approach for cyberattack detection in smart grids based on an extremely randomized three (Extra-Trees) algorithm and kernel principal component analysis (K-PCA) for dimensionality reduction. Extra-Trees uses a large number of DTs and chooses a split rule based on a random subset features and a partially random cut point. K-PCA is used to tackle the increasing computational complexity of big power systems by considering non-linearities inherent in data with complicated structures. They studied the attack detection problem where the labels of the samples are randomly corrupted. Specifically, a percentage of true labels in the training is flipped also know as label noise. They tested their proposed approach on the IEEE 57-bus and 118-bus systems. Their numerical results show that their approach outperforms state of the art approaches.

Liao et al. (2022) proposed a divergence-based transferability analysis to decide whether or not to apply transfer learning and automatically adapt a smart grid intrusion detection strategy. In particular, they explored three metrics to capture the similarity of data distributions to understand the relationship between detector's accuracy drop and similarity. Following up on this analysis, they trained two regression models to approximate the similarity and accuracy relationship needed to predict accuracy drops, which indicates the need for transfer learning. A domain adversarial neural network (DANN) classifier is adopted as transfer learning model. To validate their effectiveness, they used datasets from real normal operations from ISO New England (Muzhikyan et al. 2019) and simulated attacks from the IEEE 30-bus system in different conditions including attack timing, location, and both. Ultimately, their approach shows that the DANN can be timely triggered to achieve an accuracy improvement over 5.0%.

4.5.2 Unsupervised learning

Li et al. (2016) proposed a Dirichlet-based probabilistic model to asses the reputation levels of decentralized local agents (LA). Initial reputation levels of LAs used historical data to train the proposed model. To detect opportunistic attackers, they develop an adaptive detection algorithm with a reputation incentive mechanism. Specifically, they used the Bayes rule to assist the control center in making informed decisions about LAs being compromised. In

doing so, they used a Dirichlet distribution as a prior distribution. Initial beliefs combined with a series of historical observations shape the posterior distribution that is best suited for the reputation model. To estimate the overall status of combined LA's behaviors, they leverage a reputation level in their scheme based on the graded mean value of each compliance level. They demonstrated the utility of the proposed framework using data from IEEE-39 power system using the PowerWorld simulator.

4.5.3 Reinforcement learning

Hu et al. (2022) proposed a RL-based adaptive feature boosting that leverages a series of AEs capturing critical features from multi-source smart grid data for the classification of normal, fault, and attack events. They used multiple AEs to extract representative features from different feature sets which are extracted through a weighted feature sampling process. These extracted features are informed by a reinforcement learning approach called deep deterministic policy gradient (DDPG) to determine the feature sampling probability based on classification accuracy. AE-based extracted features are then feed in a RF classifier as base classifier along with an ensemble to distinguish between different types of incidents. They evaluated their proposed approach in two realistic datasets collected from hardware-in-the-loop (HIL) and WUSTIL-IIOT-2021 security testbeds showing an increase in classification accuracy with respect to the vanilla adaptive feature boosting.

Table 6 summarizes key attributes of the papers reviewed in Sect. 4.5.

5 Discussion

This SLR focuses on the use of AAD methods for CPS. We classify the review studies using a new taxonomy based on CPS applications and ML algorithms. This section covers the SLR findings, current approach limitations, and future research directions in AAD for CPS.

5.1 Findings

Supervised, unsupervised, and reinforcement learning can be trained in both offline and online fashion on traditional ML, DL, and hybrid models. Unsupervised approaches are better equipped to detect unknown anomalies than supervised learning algorithms. Most unsupervised methods rely only on benign data (also know as one-class classification (Chandola et al. 2009; Yuan and Wu 2022)). Semi-supervised anomaly detection typically assumes a small number of labelled normal and abnormal samples and a large number of unlabeled samples in the training dataset. Therefore, we group them under the unsupervised category. We notice that a few of the reviewed papers proposed semi-supervised approaches, where the training data consist only from normal data without anomalies (Goldstein and Uchida 2016). RL addresses the challenge of learning from interaction with an environment to achieve long term goals of protecting CPS against cyberattacks. Although RL is a promising approach, we noticed that it does not have widespread use in practice (Nguyen and Reddi 2021).

As normal streaming data in CPS is easier to collect than attack data, unsupervised learning is a promising approach for AAD in CPS. We observe that OCSVM (a traditional ML

Table 6 Summary of AAD methods in smart grid applications

Reference	Model	Algorithm	Dataset	Attack	Strengths	Weaknesses
Li et al. (2016)	Unsupervised	Dirichlet-based probabilistic model	IEEE 39-bus simulations	FDIA	Continual learning	Test on simulated data
Adhikari et al. (2017)	Supervised	HAT, DDM, ADWIN	HIL testbed simulation	fault, line maintenance, load fluctuation, and cyber-attacks scenarios	Continual learning	Need of labeled data
Wang and Govindarasu (2020)	Supervised	SVMLDT	IEEE 39-bus simulation	DoS and Replay attacks	Continual learning	Computational complexity
Acosta et al. (2020)	Supervised	Extra-Threes and K-PCA	IEEE 118-bus and 57-bus simulations	FDIA	Dimensionality reduction reduces complexity of models	Lack of continual learning
Liao et al. (2022)	Supervised	DANN	Public load demand	FDIA	Continual learning	Framework complexity
Hu et al. (2022)	Reinforcement learning	AEs and DDPG	HIL and WUSTIL-IIOT-202	Line maintenance, data injection, remote tripping command, relay setting change	Continual learning	Tested on small-scale testbed

model) and AEs (DL based) were commonly used approaches in unsupervised learning. We notice that a combination of sequential models based on RNNs such as LSTM and GRU with other DL architectures including CNNs or rule-based models have been effective on increasing detection capabilities for sophisticated attacks. As they tend to be a good fit to model the normal temporal dependencies in time series data generated by CPS, they have been used successfully in unsupervised fashion. Unsupervised models can detect a wider range of attacks including previously unseen attacks compared to supervised models at the expense of producing higher false positive rates. Generally, DL methods have achieved better performance than traditional ML methods. Nonetheless, due to expensive resource requirements and high latency, DL methods might not be the best option to deploy AAD solutions for CPS, given their usual limited resource availability. We notice that the use of hybrid and ensemble models can help to overcome the limitations of individual models by increasing the robustness of decision making. We observe that only two reviewed papers use RL strategies, suggesting there is plenty of room for exploration and promise to detect novel attacks in streaming environments.

We apply our taxonomy to the reviewed work in Fig. 4. It reveals how the reviewed AAD for CPS work distributes across CPS applications, learning paradigm, data management, and algorithms. We make the following observations: (1) most of the reviewed work focused on ICS; this corresponds to more than 50% (24 out of 47) of the reviewed papers; (2) the vast majority of proposed literature focused on supervised models that require labels;

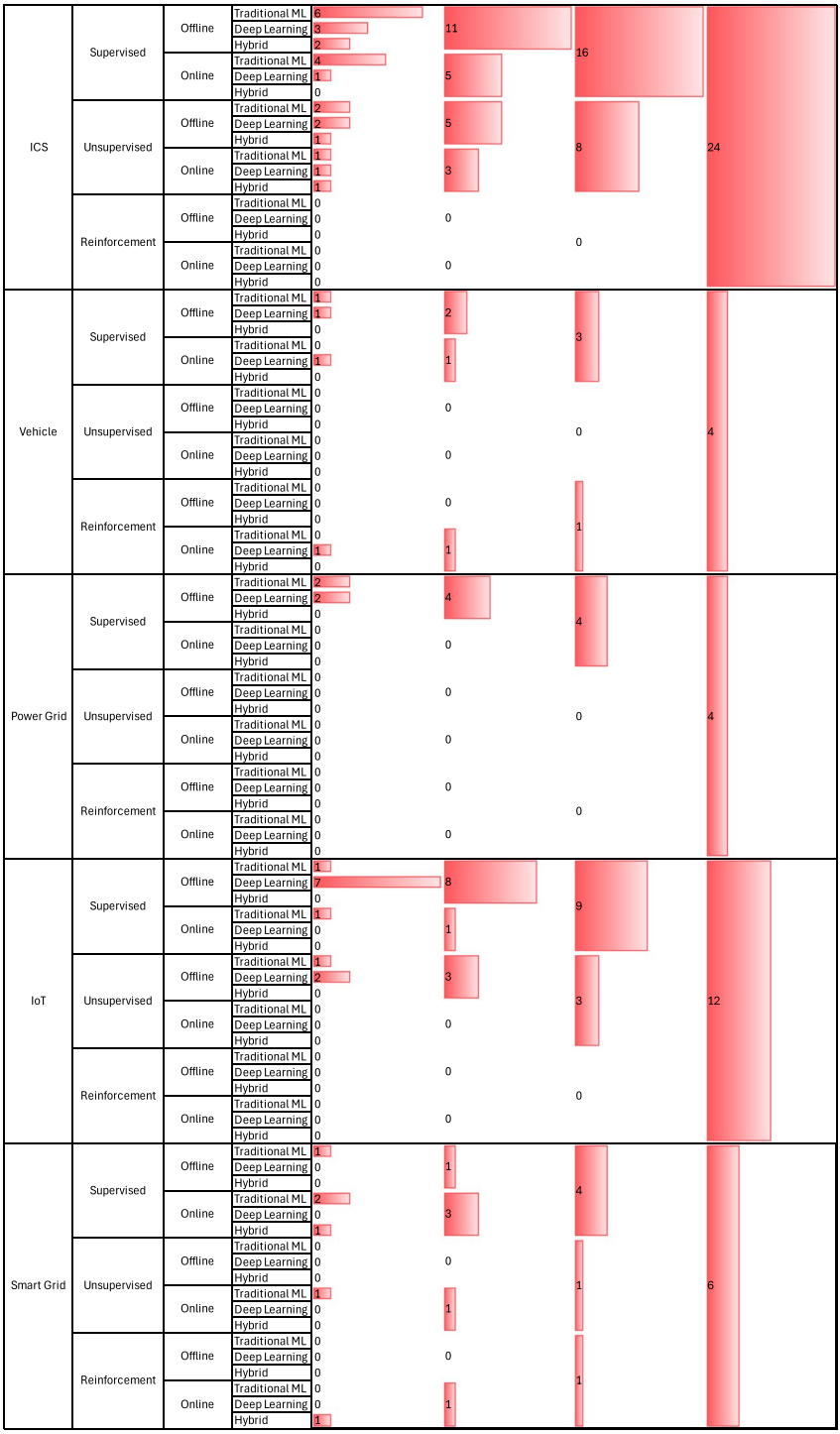

Fig. 4 AAD for CPS distribution among application, learning paradigm, data management, and algorithm categories

this corresponds to three quarters (36 out of 47) of the reviewed papers; (3) most proposed methods are tested in an offline setting; this corresponds to nearly two-thirds of the reviewed papers (32 out of 47); (4) most of the proposed methods rely on DL techniques; this correspond to nearly 50% (21 out of 47) of the reviewed papers. One of the most prominent findings of our analysis is the lack of online evaluation of the proposed methods with only about one third (16 out of 47) of the reviewed papers discussing it.

Given the complexity of attacks and the conditions of the deployment environment, AAD methods are expected to use diverse techniques within resource-constrained settings. Our SLR suggests that an online unsupervised ensemble method is the most suitable approach for AAD in CPS, given the required constraints. Given the complexity of attacks and the conditions of the deployment environment, AAD methods are expected to use diverse techniques within resource-constrained settings. Our SLR suggests that an online unsupervised ensemble method is the most suitable approach for AAD in CPS, given the required constraints.

5.2 Limitations

5.2.1 Paper retrieval omissions

Our SLR may potentially overlook relevant papers during the search process. While collecting AAD research papers from various publishers, there is a risk of missing those with incomplete abstracts or without adaptation focus or the CPS keyword. To address this, we used a systematic approach that combined manual searching, automatic searching, and forward and backward search to minimize the chance of missing relevant papers. We searched for AAD papers in CPS from leading engineering and science publishers that sponsor top conferences in data mining and computer security, extracting AAD and CPS keywords for manual searches. We also conducted automatic searches using a carefully selected set of keywords. To further expand our results, we applied forward and backward search snowballing techniques.

5.2.2 Paper selection bias

Manual inspection of papers has inherent limitations and potential biases. We selected papers using a process that combined manual and automated steps, followed by validation based on quality criteria. However, the manual validation stage may be biased by the researcher's judgment, affecting the accuracy of paper quality assessment. To address this, the coauthors—experts in data mining, anomaly detection, and cybersecurity—conducted an additional in-depth review for relevance and quality screen (see Fig. 1). This step aims to improve the accuracy of paper selection and reduce the risk of omissions and misclassifications. By implementing these measures, we aim to ensure the accuracy and integrity of the selected papers, minimize selection bias, and enhance the reliability of our SLR.

5.3 Future research directions

5.3.1 AAD datasets

Evaluation of AAD methods in CPS is highly dependent on the data being used. The use of low quality data with absent sophisticated attacks on CPS may lead to biased and incorrect conclusions. Some datasets are based on injecting simulated attacks under realistic conditions, which hinders nuances of CPS. Thus, it is difficult to evaluate, compare, and improve AAD without having proper datasets. Associated reasons for this limitation include: (1) cost to produce real attacks in CPS, (2) associated risk for introducing attacks in CPS, (3) the disclosure of private information (Verma et al. 2024). Looking forward, the creation of CPS attack datasets in controlled but realistic conditions and the combination of multiple existing datasets is an interesting direction in the future. A good example is the ROAD dataset (Verma et al. 2024) in the vehicle domain but other datasets are needed in other CPS applications.

5.3.2 Low detection latency

As CPS transmit data in real-time, AAD techniques should act rapidly to take appropriate countermeasures in near real-time. However, most of the studies that we reviewed proposed solutions incapable of detecting attacks in near real-time. In addition, the DL-based methods we reviewed assume availability of a large number of computational resources in the cloud to arrive at conclusions. However, since CPS have multiple components, connection stability is a key factor for the deployment in cloud environments. Further AAD methods that are able to effectively process streaming data and arrive at detection decisions with low latency is a needed direction of study in the future.

5.3.3 Consistent evaluation metrics

AAD reviewed work performed evaluation tests of their approaches on datasets from different nature, including real and synthetic data. Evaluation metric comparisons using real and synthetic data are possible thanks to the common benign and attack data usually coming with the datasets. However, we notice a variety of evaluation metrics usually reported in the revised studies, so that studies do not report common metrics that allows a head-to-head comparison. Common evaluation metrics reported include accuracy, precision, recall, F1 score, and AUC-ROC. In the context of anomaly detection, it is common to face imbalanced datasets, making accuracy an inappropriate metric in the area. Recent studies have suggested the use of the Matthews correlation coefficient (MCC) which in general is intended to address the unbalanced data issues commonly found in anomaly detection datasets (Chicco and Jurman 2020). Therefore, to make fair comparisons on AAD approaches suggested metrics include precision, recall, false positive rate, false negative rate, and MCC. In addition, detection latency found in a handful of studies should be also reported particularly in the streaming context. Thus, we expect future work including suggested metrics as part of their evaluation criteria.

5.3.4 Unsupervised AAD

Since attack datasets are scarce and labeling is difficult and costly, unsupervised learning techniques (e.g., clustering, OCSVM, AEs) are well suited for AAD. The fundamental assumption in unsupervised learning is that only benign data is used to model normal behavior and a threshold is used as criteria for decision making. With abundant benign data in streaming scenarios, an important future research direction is to develop adaptive threshold mechanisms that can be updated online and respond to concept drifts.

5.3.5 DL-based AAD requires abundant data

Since many AAD approaches rely on DL, the lack of realistic attack and benign datasets is a common challenge. Therefore learning from a few samples in a dynamic, changing environment is key. Other fields such as computer vision and natural language processing have developed powerful machinery to learn from limited data, including transfer learning (Pan and Yang 2009), one-shot learning (Vinyals et al. 2016), and zero-shot learning (Xian et al. 2017). Building on previous work and adapting it to use small labeled datasets for AAD is an important future research direction.

5.3.6 Model's complexity

The reviewed papers focused mainly on the algorithmic and software aspects of their proposed AAD techniques. However, the deployment of these techniques is often overlooked due to the hardware constraints common in the CPS applications we studied. Since host-based deployment is unfeasible due to additional security and privacy related concerns, a different approach is needed so that the online conditions needed to deploy AAD are meet in a cost-effective manner. Therefore, increasing attention must be placed on the use and integration with edge, fog, or cloud computing infrastructures. To this, deployment and validation of AAD in edge-fog-cloud computing infrastructures can be considered an important future research direction.

5.3.7 Adversarial AAD

A few proposed AAD approaches are effective in identifying previously unseen anomalies in CPS with high detection rates. However, these models remain vulnerable to adversarial attacks, including white-box, black-box, and tampering attacks (Chakraborty et al. 2021; Aloraini and Javed 2024). None of the reviewed studies address how to protect the proposed AAD approaches from attacks. Therefore, developing secure AAD for CPS in adversarial settings is a promising and challenging direction for future research. Yuan and Wu (2022); Mohus and Li (2023). We observe that existing solutions from the AI security domain could be adapted for AAD in CPS (Goodge et al. 2021; Lo et al. 2022).

5.3.8 Explainable AAD

The reviewed papers primarily focus on improving classification outcomes for AAD. Most proposed methods, particularly those based on DL, emphasize black-box anomaly detection

for making critical predictions. However, various stakeholders are increasingly demanding explainability. Preece et al. (2018). Explainable anomaly detection refers to a model's ability to clarify why and when it identifies an anomaly (Li et al. 2023). Prioritizing explainable AAD is a critical research direction, as decision-making in CPS demands models that can provide clear explanations of detection results without compromising prediction quality.

6 Conclusion

We have performed a SLR in the field of AAD in CPS. We performed our searchers in five prominent research databases (i.e., IEEE Explore, ACM Digital Library, Emerald Insight, Springer Link, and Science Direct) and execute forward and backward snowballing search to maximize the literature search coverage. After analyzing 47 research papers and 18 review articles, this study introduces a novel taxonomy based on attacks, CPS applications, learning paradigm (i.e., supervised, unsupervised, and reinforcement learning), and ML algorithms. In particular, we outline algorithms, datasets, attack characteristics, strengths and weaknesses developed in each of the papers. Details about the learning paradigm used are discussed with respect to each step guiding attack detection strategy across CPS applications.

By using a known and standardized approach, our SLR ensures that key papers in the searched databases are found. Thus, researchers and practitioners should not need to repeat this work to find relevant publications from the period 2013 to 2023 (November). Still, the SLR approach allows the work to be repeated in the future to track field developments. Our categorization provides a clear overview of AAD in CPS and related research, making it easy to find relevant papers for specific CPS applications. The number of papers in each category shows which research areas were seen as important and challenging during the studied period.

The current findings indicate that most of the previous studies investigated only a single aspect of adaptation (either data processing or model adaptation). Limited studies have provided a comprehensive overview of both aspects of AAD at the same time. Hence, this SLR contributed to the anomaly detection literature by summarizing different adaptive ML-based anomaly detection algorithms into distinct CPS applications. The intersection of these topics has not been discussed previously. The identified and categorized studies not only add the conceptual discussion in the filed of AAD but also provided several enlightened ideas for researchers and practitioners. Similarly, different types of data analysis methods can also diversify our investigation in future studies and enrich our results.

Finally, based on prior research, we highlight several directions and considerations for future studies. These directions and considerations include, tight integration between near real-time data processing and an adaptive learning mode, explainable and actionable detection predictions, and quantifying the uncertainty of detection predictions.

Acknowledgements This research was sponsored in part by Oak Ridge National Laboratory's (ORNL's) Laboratory Directed Research and Development program and by the DOE. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this manuscript.

Author contributions Equal Contribution: P.M. and S.H. contributed equally to this work. Supervision: P.M. supervised the research. Validation: P.M. and S.H. Writing—original draft: P.M. and S.H. prepared main manuscript. Conceptualization: P.M and S.H. conceived the study and designed the research framework.

Visualization: P.M. and S.H. created the figures and tables. Data collection and curation: S.H collected and filtered the data. Methodology development: P.M. and S.H. developed the systematic review protocol, including paper selection criteria, data extraction, and quality assessment. Investigation: All authors participated in the review and synthesis of papers after the filtration process. Writing—review and editing: All authors reviewed, revised, and approved the final manuscript. Funding acquisition: P.M.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

Abie H (2019) Cognitive cybersecurity for CPS-IoT enabled healthcare ecosystems. In: 2019 13th International Symposium on Medical Information and Communication Technology (ISMICT). IEEE, pp 1–6

Acosta MRC, Ahmed S, Garcia CE, Koo I (2020) Extremely randomized trees-based scheme for stealthy cyber-attack detection in smart grid networks. IEEE Access 8:19921–19933

Adhikari U, Morris TH, Pan S (2017) Applying hoeffding adaptive trees for real-time cyber-power event and intrusion classification. IEEE Trans Smart Grid 9(5):4049–4060

Ahmad S, Lavin A, Purdy S, Agha Z (2017) Unsupervised real-time anomaly detection for streaming data. Neurocomputing 262:134–147

Akowuah F, Kong F (2021) Real-time adaptive sensor attack detection in autonomous cyber-physical systems. In: 2021 IEEE 27th real-time and embedded technology and applications symposium (RTAS). IEEE, pp 237–250

Alaghbari KA, Saad MHM, Hussain A, Alam MR (2022) Complex event processing for physical and cyber security in datacentres-recent progress, challenges and recommendations. J Cloud Comput 11(1):65

Al-E'mari S, Anbar M, Sanjalawe Y, Manickam S (2020) A labeled transactions-based dataset on the ethereum network. In: International conference on advances in cyber security, Penang, Malaysia. Springer, Singapore, pp 61–79

Alohali MA, Al-Wesabi FN, Hilal AM, Goel S, Gupta D, Khanna A (2022) Artificial intelligence enabled intrusion detection systems for cognitive cyber-physical systems in industry 4.0 environment. Cogn Neurodyn 16(5):1045–1057

Aloraini F, Javed A (2024) Adversarial attacks in intrusion detection systems: triggering false alarms in connected and autonomous vehicles. In: 2024 IEEE international conference on cyber security and resilience (CSR). IEEE, pp 714–719

Alshammari FH (2023) Design of capability maturity model integration with cybersecurity risk severity complex prediction using Bayesian-based machine learning models. SOCA 17(1):59–72

Alsulami AA, Al-Haija QA, Alturki B, Alqahtani A, Alsini R (2023) Security strategy for autonomous vehicle cyber-physical systems using transfer learning. J Cloud Comput 12(1):181–199

Althobaiti MM, Kumar KPM, Gupta D, Kumar S, Mansour RF (2021) An intelligent cognitive computing based intrusion detection for industrial cyber-physical systems. Measurement 186:110145

Andrade RO, Yoo SG (2019) Cognitive security: a comprehensive study of cognitive science in cybersecurity. J Inf Secur Appl 48:102352

Arrington B, Barnett L, Rufus R, Esterline A (2016) Behavioral modeling intrusion detection system (BMIDS) using internet of things (IoT) behavior-based anomaly detection via immunity-inspired algorithms. In: 2016 25th International conference on computer communication and networks (ICCCN). IEEE, pp 1–6
Arshad K, Ali RF, Muneer A, Aziz IA, Naseer S, Khan NS, Taib SM (2022) Deep reinforcement learning for anomaly detection: a systematic review. IEEE Access 10:124017–124035
Atat R, Liu L, Wu J, Li G, Ye C, Yang Y (2018) Big data meet cyber-physical systems: a panoramic survey. IEEE Access 6:73603–73636
Bacher R, Chatelain F, Michel O (2016) An adaptive robust regression method: application to galaxy spectrum baseline estimation. In: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 4423–4427
Bala R, Nagpal R (2019) A review on KDD cup9 and NSL-KDD dataset. Int J Adv Res Comput Sci 10(2):64
Basati A, Faghih MM (2023) APAE: an iot intrusion detection system using asymmetric parallel auto-encoder. Neural Comput Appl 35(7):4813–4833
Bezzina D, Sayer J (2014) Safety pilot model deployment: test conductor team report. Report No DOT HS 812(171):18
Bibi I, Akhunzada A, Kumar N (2022) Deep ai-powered cyber threat analysis in IIoT. IEEE Internet Things J 10(9):7749–7760
Biggio B (2024) Machine learning in computer security is difficult to fix. Commun ACM 67(11):103–103
Booij TM, Chiscop I, Meeuwissen E, Moustafa N, Den Hartog FT (2021) Ton_IoT: the role of heterogeneity and the need for standardization of features and attack types in IoT network intrusion data sets. IEEE Internet Things J 9(1):485–496
Buczak AL, Guven E (2015) A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Commun Surv Tutor 18(2):1153–1176
Cai T, Jia T, Adepu S, Li Y, Yang Z (2023) ADAM: an adaptive ddos attack mitigation scheme in software-defined cyber-physical system. IEEE Trans Industr Inf 19(6):7802–7813
Cárdenas AA, Amin S, Sastry S (2008) Secure control: Towards survivable cyber-physical systems. In: 2008 The 28th international conference on distributed computing systems workshops. IEEE, Beijing, pp 495–500. https://doi.org/10.1109/ICDCS.Workshops.2008.41
Chakraborty A, Alam M, Dey V, Chattopadhyay A, Mukhopadhyay D (2021) A survey on adversarial attacks and defences. CAAI Trans Intell Technol 6(1):25–45
Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Comput Surv (CSUR) 41(3):1–58
Chicco D, Jurman G (2020) The advantages of the matthews correlation coefficient (MCC) over f1 score and accuracy in binary classification evaluation. BMC Genomics 21:1–13
Clements AA, Almakhdhub NS, Saab KS, Srivastava P, Koo J, Bagchi S, Payer M (2017) Protecting bare-metal embedded systems with privilege overlays. In: 2017 IEEE symposium on security and privacy (SP), San Jose, CA. IEEE, pp 289–303
Cooper A, Bretas A, Meyn S (2023) Anomaly detection in power system state estimation: review and new directions. Energies 16(18):6678
Cui Y, Bai F, Yan R, Saha T, Ko RK, Liu Y (2021) Source authentication of distribution synchrophasors for cybersecurity of microgrids. IEEE Trans Smart Grid 12(5):4577–4580
Dai D, Boroomand S (2022) A review of artificial intelligence to enhance the security of big data systems: state-of-art, methodologies, applications, and challenges. Arch Comput Methods Eng 29(2):1291–1309
Das S, Islam MR, Jayakodi NK, Doppa JR (2018) Active anomaly detection via ensembles. arXiv preprint. arXiv:1809.06477
Dey AK, Gupta GP, Sahu SP (2023) Hybrid meta-heuristic based feature selection mechanism for cyber-attack detection in iot-enabled networks. Procedia Comput Sci 218:318–327
Ding J, Lu C, Li B (2022) A data-driven based security situational awareness framework for power systems. J Signal Process Syst 94(11):1159–1168
Fei X, Shah N, Verba N, Chao K-M, Sanchez-Anguix V, Lewandowski J, James A, Usman Z (2019) CPS data streams analytics based on machine learning for cloud and fog computing: a survey. Futur Gener Comput Syst 90:435–450
Feng Z, Guan N, Lv M, Liu W, Deng Q, Liu X, Yi W (2020) Efficient drone hijacking detection using two-step GA-XGBOOST. J Syst Architect 103:101694
Gama J, Žliobaitė I, Bifet A, Pechenizkiy M, Bouchachia A (2014) A survey on concept drift adaptation. ACM comput surv (CSUR) 46(4):1–37
Giannoni F, Mancini M, Marinelli F (2018) Anomaly detection models for iot time series data. arXiv preprint. arXiv:1812.00890
Giraldo J, Urbina D, Cardenas A, Valente J, Faisal M, Ruths J, Tippenhauer NO, Sandberg H, Candell R (2018) A survey of physics-based attack detection in cyber-physical systems. ACM Comput Surv (CSUR) 51(4):1–36

Goldstein M, Uchida S (2016) A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLoS ONE 11(4):0152173
Goodge A, Hooi B, Ng SK, Ng WS (2021) Robustness of autoencoders for anomaly detection under adversarial impact. In: Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence, pp 1244–1250
Gopalakrishnan T, Ruby D, Al-Turjman F, Gupta D, Pustokhina IV, Pustokhin DA, Shankar K (2020) Deep learning enabled data offloading with cyber attack detection model in mobile edge computing systems. IEEE Access 8:185938–185949
Grimm D, Stang M, Sax E (2021) Context-aware security for vehicles and fleets: a survey. IEEE Access 9:101809–101846
Gupta D, Moni SS, Tosun AS (2023) Integration of digital twin and federated learning for securing vehicular internet of things. In: Proceedings of the 2023 international conference on research in adaptive and convergent systems, Gdansk, Poland, pp 1–8
Gyamfi E, Jurcut AD (2022) Novel online network intrusion detection system for industrial iot based on OI-SVDD and AS-ELM. IEEE Internet Things J 10(5):3827–3839
Hong T, Pinson P, Fan S (2014) Global energy forecasting competition 2012. Int J Forecast 30(2):357–363
Hu C, Yan J, Liu X (2022) Reinforcement learning-based adaptive feature boosting for smart grid intrusion detection. IEEE Trans Smart Grid 14(4):3150–3163
Huang L, Zhu Q (2020) A dynamic games approach to proactive defense strategies against advanced persistent threats in cyber-physical systems. Comput Secur 89:101660
Huang Y, Huang L, Zhu Q (2022) Reinforcement learning for feedback-enabled cyber resilience. Annu Rev Control 53:273–295
Hubballi N, Suryanarayanan V (2014) False alarm minimization techniques in signature-based intrusion detection systems: a survey. Comput Commun 49:1–17
Huber PJ (1992) Robust estimation of a location parameter. In: Breakthroughs in statistics: methodology and distribution. Springer, New York, pp 492–518
Humayed A, Lin J, Li F, Luo B (2017) Cyber-physical systems security—a survey. IEEE Internet Things J 4(6):1802–1831
Ibidunmoye O, Rezaie A-R, Elmroth E (2017) Adaptive anomaly detection in performance metric streams. IEEE Trans Netw Serv Manage 15(1):217–231
Ibor AE, Okunoye OB, Oladeji FA, Abdulsalam KA (2022) Novel hybrid model for intrusion prediction on cyber physical systems' communication networks based on bio-inspired deep neural network structure. J Inf Secur Appl 65:103107
Intriago G, Zhang Y (2023) Real-time power system event detection: a novel instance selection approach. IEEE Access 11:46765–46781
Ioulianou P, Vasilakis V, Moscholios I, Logothetis M (2018) A signature-based intrusion detection system for the internet of things. In: Information and communication technology form, AUT
Ivanov R, Pajic M, Lee I (2016) Attack-resilient sensor fusion for safety-critical cyber-physical systems. ACM Trans Embedded Comput Syst (TECS) 15(1):1–24
Jamal AA, Majid A-AM, Konev A, Kosachenko T, Shelupanov A (2023) A review on security analysis of cyber physical systems using machine learning. Mater Today proc 80:2302–2306
Jiao J, Tang Z, Zhang P, Yue M, Yan J (2022) Cyberattack-resilient load forecasting with adaptive robust regression. Int J Forecast 38(3):910–919
Jullian O, Otero B, Rodriguez E, Gutierrez N, Antona H, Canal R (2023) Deep-learning based detection for cyber-attacks in IoT networks: a distributed attack detection framework. J Netw Syst Manage 31(2):33
Junejo KN, Goh J (2016) Behaviour-based attack detection and classification in cyber physical systems using machine learning. In: Proceedings of the 2nd ACM International workshop on cyber-physical system security, Xi'an, China, pp 34–43
Jung E, Le Tilly M, Gehani A, Ge Y (2019) Data mining-based ethereum fraud detection. In: 2019 IEEE international conference on blockchain (blockchain), Atlanta, GA. IEEE, pp 266–273
Kayan H, Nunes M, Rana O, Burnap P, Perera C (2022) Cybersecurity of industrial cyber-physical systems: a review. ACM Comput Surv (CSUR) 54(11s):1–35
Khan IA, Pi D, Khan N, Khan ZU, Hussain Y, Nawaz A, Ali F (2021) A privacy-conserving framework based intrusion detection method for detecting and recognizing malicious behaviours in cyber-physical power networks. Appl Intell 51:7306–7321
Kim S, Park K-J, Lu C (2022) A survey on network security for cyber-physical systems: from threats to resilient design. IEEE Commun Surv Tutor 24(3):1534–1573
Kim M, Kim J, Yu J, Choi JK (2023) Active anomaly detection based on deep one-class classification. Pattern Recogn Lett 167:18–24
Koay AM, Ko RKL, Hettema H, Radke K (2023) Machine learning in industrial control system (ICS) security: current landscape, opportunities and challenges. J Intell Inf Syst 60(2):377–405

Kundur D, Liu X, Zourntos S, Butler-Purry KL (2011) Towards a framework for cyber attack impact analysis of the electric smart grid. In: 2011 IEEE international conference on smart grid communications (SmartGridComm). IEEE, Brussels, pp 244–249. https://doi.org/10.1109/SmartGridComm.2011.6102312
Kure HI, Islam S, Ghazanfar M, Raza A, Pasha M (2022) Asset criticality and risk prediction for an effective cybersecurity risk management of cyber-physical system. Neural Comput Appl 34(1):493–514
Lee EA (2008) Cyber physical systems: Design challenges. In: Proceedings of the 11th IEEE international symposium on object and component-oriented real-time distributed computing (ISORC). IEEE, Orlando, pp 363–369. https://doi.org/10.1109/ISORC.2008.25
Li B, Lu R, Wang W, Choo K-KR (2016) DDOA: a Dirichlet-based detection scheme for opportunistic attacks in smart grid cyber-physical system. IEEE Trans Inf Forensics Secur 11(11):2415–2425
Li F, Shinde A, Shi Y, Ye J, Li X-Y, Song W (2019) System statistics learning-based IoT security: feasibility and suitability. IEEE Internet Things J 6(4):6396–6403
Li K, Ma W, Duan H, Xie H, Juanxiu Z (2022) Few-shot IoT attack detection based on RFP-CNN and adversarial unsupervised domain-adaptive regularization. Comput Secur 121:102856
Li Z, Zhu Y, Van Leeuwen M (2023) A survey on explainable anomaly detection. ACM Trans Knowl Discov Data 18(1):1–54
Liao P, Yan J, Sellier JM, Zhang Y (2022) Divergence-based transferability analysis for self-adaptive smart grid intrusion detection with transfer learning. IEEE Access 10:68807–68818
Lin C-C, Tsai C-T, Liu Y-L, Chang T-T, Chang Y-S (2023) Security and privacy in 5G-IIoT smart factories: novel approaches, trends, and challenges. Mobile Netw Appl 28:1043–1058
Liu C, Lore KG, Jiang Z, Sarkar S (2021) Root-cause analysis for time-series anomalies via spatiotemporal graphical modeling in distributed complex systems. Knowl Based Syst 211:106527
Liu W, Xu X, Wu L, Qi L, Jolfaei A, Ding W, Khosravi MR (2022) Intrusion detection for maritime transportation systems with batch federated aggregation. IEEE Trans Intell Transp Syst 24(2):2503–2514
Lo S-Y, Oza P, Patel VM (2022) Adversarially robust one-class novelty detection. IEEE Trans Pattern Anal Mach Intell 45(4):4167–4179
Loeffel P-X (2017) Adaptive machine learning algorithms for data streams subject to concept drifts. PhD thesis, Université Pierre et Marie Curie-Paris VI
Luo Y, Xiao Y, Cheng L, Peng G, Yao D (2021) Deep learning-based anomaly detection in cyber-physical systems: progress and opportunities. ACM Comput Surv (CSUR) 54(5):1–36
Lu P, Zhang L, Park BB, Feng L (2018) Attack-resilient sensor fusion for cooperative adaptive cruise control. In: 2018 21st International conference on intelligent transportation systems (ITSC), Maui, HI. IEEE, pp 3955–3960
Mahapatra SN, Singh BK, Kumar V (2020) A survey on secure transmission in internet of things: taxonomy, recent techniques, research requirements, and challenges. Arab J Sci Eng 45(8):6211–6240
Mahdavifar S, Ghorbani AA (2020) DENNES: deep embedded neural network expert system for detecting cyber attacks. Neural Comput Appl 32(18):14753–14780
Meira J, Andrade R, Praça I, Carneiro J, Bolón-Canedo V, Alonso-Betanzos A, Marreiros G (2020) Performance evaluation of unsupervised techniques in cyber-attack anomaly detection. J Ambient Intell Humaniz Comput 11(11):4477–4489
Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint. arXiv:1301.3781
Mirsky Y, Doitshman T, Elovici Y, Shabtai A (2018) Kitsune: an ensemble of autoencoders for online network intrusion detection. arXiv preprint. arXiv:1802.09089
Mitchell R, Chen I-R (2013) On survivability of mobile cyber physical systems with intrusion detection. Wirel Pers Commun 68:1377–1391
Mohus ML, Li J (2023) Adversarial robustness in unsupervised machine learning: A systematic review. arXiv preprint arXiv:2306.00687
Moriano P, Hill R, Camp LJ (2021) Using bursty announcements for detecting bgp routing anomalies. Comput Netw 188:107835
Moriano P, Bridges RA, Iannacone MD (2022) Detecting can masquerade attacks with signal clustering similarity. In: Proceedings Fourth International Workshop on Automotive and Autonomous Vehicle Security. AutoSec 2022, pp. 1–8. Internet Society, San Diego, CA, USA. https://doi.org/10.14722/autosec.2022.23028
Moriano P, Hespeler SC, Li M, Bridges RA (2024) Benchmarking Unsupervised Online IDS for Masquerade Attacks in CAN. arXiv:2406.13778
Morris T (2013) Industrial control system (ICS) cyber attack datasets. https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets
Moustafa N, Slay J (2015) Unsw-nb15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In: 2015 Military communications and information systems conference (MilCIS), Canberra. IEEE, pp. 1–6

Mowla NI, Tran NH, Doh I, Chae K (2020) AFRL: adaptive federated reinforcement learning for intelligent jamming defense in FANET. J Commun Netw 22(3):244–258
Munir M, Siddiqui SA, Dengel A, Ahmed S (2018) DEEPANT: a deep learning approach for unsupervised anomaly detection in time series. IEEE Access 7:1991–2005
Muzhikyan A, Muhanji SO, Moynihan GD, Thompson DJ, Berzolla ZM, Farid AM (2019) The 2017 ISO new england system operational analysis and renewable energy integration study (SOARES). Energy Rep 5:747–792
Nakayama K, Muralidhar N, Jin C, Sharma R (2019) Detection of false data injection attacks in cyber-physical systems using dynamic invariants. In: 2019 18th IEEE international conference on machine learning and applications (ICMLA), Boca Raton, FL. IEEE, pp 1023–1030
Nguyen TT, Reddi VJ (2021) Deep reinforcement learning for cyber security. IEEE Trans Neural Netw Learn Syst 34(8):3779–3795
Odiathevar M, Seah WK, Frean M (2019) A hybrid online offline system for network anomaly detection. In: 2019 28th International conference on computer communication and networks (ICCCN), Valencia, ES. IEEE, pp 1–9
Okoli C (2015) A guide to conducting a standalone systematic literature review. Commun Assoc Inf Syst 37:879–910
Olowononi FO, Rawat DB, Liu C (2020) Resilient machine learning for networked cyber physical systems: a survey for machine learning security to securing machine learning for cps. IEEE Commun Surv Tutor 23(1):524–552
Pan SJ, Yang Q (2009) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359
Pan F, Pang Z, Wen H, Luvisotto M, Xiao M, Liao R-F, Chen J (2019) Threshold-free physical layer authentication based on machine learning for industrial wireless CPS. IEEE Trans Industr Inf 15(12):6481–6491
Pasqualetti F, Dörfler F, Bullo F (2013) Attack detection and identification in cyber-physical systems. IEEE Trans Autom Control 58(11):2715–2729. https://doi.org/10.1109/TAC.2013.2266831
Pekaric I, Groner R, Witte T, Adigun JG, Raschke A, Felderer M, Tichy M (2023) A systematic review on security and safety of self-adaptive systems. J Syst Softw 203:1–25
Petit J, Stottelaar B, Feiri M, Kargl F (2015) Remote attacks on automated vehicles sensors: experiments on camera and lidar. Black Hat Europe 11(2015):995–1008
Pike L, Hickey P, Elliott T, Mertens E, Tomb A (2016) Trackos: a security-aware real-time operating system. In: Runtime verification: 16th international conference, RV 2016, Madrid, Spain, 23–30 September 2016, proceedings 7. Springer, pp 302–317
Preece A, Harborne D, Braines D, Tomsett R, Chakraborty S (2018) Stakeholders in explainable ai. arXiv preprint. arXiv:1810.00184
Quincozes SE, Mossé D, Passos D, Albuquerque C, Ochi LS, Santos VF (2021) On the performance of grasp-based feature selection for CPS intrusion detection. IEEE Trans Netw Serv Manage 19(1):614–626
Quinonez R, Giraldo J, Salazar L, Bauman E, Cardenas A, Lin Z (2020) SAVIOR: Securing autonomous vehicles with robust physical invariants. In: 29th USENIX security symposium (USENIX security 20), pp 895–912
Raciti M (2013) Anomaly detection and its adaptation: Studies on cyber-physical systems. PhD thesis, Linköping University Electronic Press, Sweden
Rojas RA, Rauch E (2019) From a literature review to a conceptual framework of enablers for smart manufacturing control. Int J Adv Manuf Technol 104:517–533
Rosenberg I, Shabtai A, Elovici Y, Rokach L (2021) Adversarial machine learning attacks and defense methods in the cyber security domain. ACM Comput Surv (CSUR) 54(5):1–36
Rutkin AH (2013) Spoofers use fake GPS signals to knock a yacht off course. MIT Technology Review. Accessed 30 Apr 2025. https://www.technologyreview.com/2013/08/14/177015/spoofers-use-fake-gps-signals-to-knock-a-yacht-off-course/
Saad M, Bukhari SBA, Kim CH (2019) A review of various modern strategies for mitigation of cyber attacks in smart grids. In: 2019 IEEE transportation electrification conference and expo, Asia-Pacific (ITEC Asia-Pacific), Seogwipo-si, KR. IEEE, pp 1–7
Saez MA, Maturana FP, Barton K, Tilbury DM (2019) Context-sensitive modeling and analysis of cyber-physical manufacturing systems for anomaly detection and diagnosis. IEEE Trans Autom Sci Eng 17(1):29–40
Schneider P, Böttinger K (2018) High-performance unsupervised anomaly detection for cyber-physical system networks. In: Proceedings of the 2018 workshop on cyber-physical systems security and privacy, Toronto, CA, pp 1–12
Settanni G, Skopik F, Karaj A, Wurzenberger M, Fiedler R (2018) Protecting cyber physical production systems using anomaly detection to enable self-adaptation. In: 2018 IEEE industrial cyber-physical systems (ICPS), Saint Petersburg, RU. IEEE, pp 173–180

Shacham H, Page M, Pfaff B, Goh E-J, Modadugu N, Boneh D (2004) On the effectiveness of address-space randomization. In: Proceedings of the 11th ACM conference on computer and communications security, Washington, DC, pp 298–307
Shahriar MH, Xiao Y, Moriano P, Lou W, Hou YT (2023) Canshield: deep learning-based intrusion detection framework for controller area networks at the signal-level. IEEE Internet Things J 10:22111–22127
Shi Z, Mamun AA, Kan C, Tian W, Liu C (2023) An lstm-autoencoder based online side channel monitoring approach for cyber-physical attack detection in additive manufacturing. J Intell Manuf 34:1815–1831
Shin J, Baek Y, Eun Y, Son SH (2017) Intelligent sensor attack detection and identification for automotive cyber-physical systems. In: 2017 IEEE symposium series on computational intelligence (SSCI), Honolulu, HI. IEEE, pp 1–8
Shoukry Y, Martin P, Tabuada P, Srivastava M (2013) Non-invasive spoofing attacks for anti-lock braking systems. In: Cryptographic hardware and embedded systems—CHES 2013: 15th international workshop, Santa Barbara, CA, USA, 20–23 August 2013. Springer, pp 55–72
Stoyanova M, Nikoloudakis Y, Panagiotakis S, Pallis E, Markakis EK (2020) A survey on the internet of things (IoT) forensics: challenges, approaches, and open issues. IEEE Commun Surv Tutor 22(2):1191–1221
Strandberg K, Nowdehi N, Olovsson T (2022) A systematic literature review on automotive digital forensics: challenges, technical solutions and data collection. IEEE Trans Intell Veh 8(2):1350–1367
Syrmakesis AD, Alcaraz C, Hatziargyriou ND (2022) Classifying resilience approaches for protecting smart grids against cyber threats. Int J Inf Secur 21(5):1189–1210
Urbina DI, Giraldo JA, Cárdenas AA, Tippenhauer NO, Valente J, Faisal M, Ruths J, Candell R, Sandberg H (2016) Limiting the impact of stealthy attacks on industrial control systems. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security (CCS), pp 1092–1105. https://doi.org/10.1145/2976749.2978388
Van Wyk F, Wang Y, Khojandi A, Masoud N (2019) Real-time sensor anomaly detection and identification in automated vehicles. IEEE Trans Intell Transp Syst 21(3):1264–1276
Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Adv Neural Inf Process Syst 30:1–11
Vávra J, Hromada M, Lukáš L, Dworzecki J (2021) Adaptive anomaly detection system based on machine learning algorithms in an industrial control environment. Int J Crit Infrastruct Prot 34:1–11
Verma ME, Bridges RA, Iannacone MD, Hollifield SC, Moriano P, Hespeler SC, Kay B, Combs FL (2024) A comprehensive guide to can ids data and introduction of the road dataset. PLoS ONE 19(1):0296879
Vinyals O, Blundell C, Lillicrap T, Wierstra D et al (2016) Matching networks for one shot learning. In: 30th Conference on neural information processing Systems (NIPS 2016), vol 29, Barcelona, ES, pp 1–9
Wang P, Govindarasu M (2020) Multi-agent based attack-resilient system integrity protection for smart grid. IEEE Trans Smart Grid 11(4):3447–3456
Wang D, Li F, Liu K, Zhang X (2023) Real-time cyber-physical security solution leveraging an integrated learning-based approach: an integrated learning-based cyber-physical security solution. ACM Trans Sensor Netw 20(2):1–22
Wohlin C (2014) Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th international conference on evaluation and assessment in software engineering, London England, UK, pp 1–10
Wu W, Li R, Xie G, An J, Bai Y, Zhou J, Li K (2019) A survey of intrusion detection for in-vehicle networks. IEEE Trans Intell Transp Syst 21(3):919–933
Xi L, Wang R, Haas ZJ (2022) Data-correlation-aware unsupervised deep-learning model for anomaly detection in cyber-physical systems. IEEE Internet Things J 9(22):22410–22421
Xi L, Miao D, Li M, Wang R, Liu H, Huang X (2023) Adaptive-correlation-aware unsupervised deep learning for anomaly detection in cyber-physical systems. IEEE Trans Dependable Secure Comput 21(4):2888–2899
Xian Y, Schiele B, Akata Z (2017) Zero-shot learning-the good, the bad and the ugly. In: Proceedings of the IEEE conference on computer vision and pattern recognition, San Francisco, CA, pp 4582–4591
Yampolskiy M, Horvath P, Koutsoukos XD, Xue Y, Sztipanovits J (2013) Taxonomy for description of cross-domain attacks on cps. In: Proceedings of the 2nd ACM international conference on high confidence networked systems, Philadelphia, PA, pp 135–142
Yasaei R, Hernandez F, Faruque MAA (2020) Iot-cad: Context-aware adaptive anomaly detection in IoT systems through sensor association. In: Proceedings of the 39th international conference on computer-aided design, Virtual Event, USA, pp 1–9
Yazdinejad A, Dehghantanha A, Parizi RM, Srivastava G, Karimipour H (2023) Secure intelligent fuzzy blockchain framework: effective threat detection in IoT networks. Comput Ind 144:103801
Yazdinejad A, Kazemi M, Parizi RM, Dehghantanha A, Karimipour H (2023) An ensemble deep learning model for cyber threat hunting in industrial Internet of Things. Digital Commun Netw 9(1):101–110

Yu F, Koltun V (2015) Multi-scale context aggregation by dilated convolutions. arXiv preprint. arXiv:1511.07122
Yuan S, Wu X (2022) Trustworthy anomaly detection: a survey. arXiv preprint. arXiv:2202.07787
Zeadally S, Sanislav T, Mois GD (2019) Self-adaptation techniques in cyber-physical systems (CPSS). IEEE Access 7:171126–171139
Zhang M, Shen C, He N, Han S, Li Q, Wang Q, Guan X (2019) False data injection attacks against smart gird state estimation: construction, detection and defense. Sci China Technol Sci 62(12):2077–2087

Authors and Affiliations

Pablo Moriano[1] · Steven C. Hespeler[1] · Mingyan Li[2] · Maria Mahbub[2]

✉ Pablo Moriano
moriano@ornl.gov

Steven C. Hespeler
hespelersc@ornl.gov

Mingyan Li
lim3@ornl.gov

Maria Mahbub
mahbubm@ornl.gov

[1] Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge 37830, TN, USA

[2] Cyber Resilience and Intelligence Division, Oak Ridge National Laboratory, Oak Ridge 37830, TN, USA