5-1-2024

Agriculture 4.0 and beyond: Evaluating cyber threat intelligence sources and techniques in smart farming ecosystems

Hang T. Bui

Hamed Aboutorab

Arash Mahboubi

Yansong Gao

Nazatul H. Sultan

See next page for additional authors

10.1016/j.cose.2024.103754

Bui, H. T., Aboutorab, H., Mahboubi, A., Gao, Y., Sultan, N. H., Chauhan, A., . . . Yan, S. (2024). Agriculture 4.0 and beyond: Evaluating cyber threat intelligence sources and techniques in smart farming ecosystems. Computers & Security, 140, article 103754. https://doi.org/10.1016/j.cose.2024.103754

Computers & Security 140 (2024) 103754

Contents lists available at ScienceDirect

Computers & Security

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

Agriculture 4.0 and beyond: Evaluating cyber threat intelligence sources and techniques in smart farming ecosystems

Hang Thanh Bui [a,*], Hamed Aboutorab [a], Arash Mahboubi [a], Yansong Gao [b], Nazatul Haque Sultan [b], Aufeef Chauhan [c], Mohammad Zavid Parvez [a], Michael Bewong [a], Rafiqul Islam [a], Zahid Islam [a], Seyit A. Camtepe [b], Praveen Gauravaram [d], Dineshkumar Singh [e], M. Ali Babar [b], Shihao Yan [f]

[a] School of Computing, Mathematics and Engineering, Charles Sturt University, Port Macquarie, 2444, NSW, Australia
[b] CSIRO's Data61, Sydney, 2122, NSW, Australia
[c] School of Computer and Mathematical Sciences, The University of Adelaide, Adelaide, 5005, SA, Australia
[d] Tata Consultancy Services, Brisbane, QLD, Australia
[e] TCS Research and Innovation, Mumbai, India
[f] School of Science and Security Research Institute, Edith Cowan University, Perth, 6027, WA, Australia

ARTICLE INFO

Keywords:
Cyber threat intelligence (CTI)
Systematic literature review
virtual Chief Information Security Officer (vCISO)
Agriculture 4.0
Agriculture 5.0
Smart farming infrastructures (SFIs)
Digital twin technology

ABSTRACT

The digitisation of agriculture, integral to Agriculture 4.0, has brought significant benefits while simultaneously escalating cybersecurity risks. With the rapid adoption of smart farming technologies and infrastructure, the agricultural sector has become an attractive target for cyberattacks. This paper presents a systematic literature review that assesses the applicability of existing cyber threat intelligence (CTI) techniques within smart farming infrastructures (SFIs). We develop a comprehensive taxonomy of CTI techniques and sources, specifically tailored to the SFI context, addressing the unique cyber threat challenges in this domain. A crucial finding of our review is the identified need for a virtual Chief Information Security Officer (vCISO) in smart agriculture. While the concept of a vCISO is not yet established in the agricultural sector, our study highlights its potential significance. The implementation of a vCISO could play a pivotal role in enhancing cybersecurity measures by offering strategic guidance, developing robust security protocols, and facilitating real-time threat analysis and response strategies. This approach is critical for safeguarding the food supply chain against the evolving landscape of cyber threats. Our research underscores the importance of integrating a vCISO framework into smart farming practices as a vital step towards strengthening cybersecurity. This is essential for protecting the agriculture sector in the era of digital transformation, ensuring the resilience and sustainability of the food supply chain against emerging cyber risks.

1. Introduction

Agriculture plays a vital role in contemporary society and is often regarded as one of the most pivotal innovations in our century. In Australia, 55% of Australian land is used for agriculture and a significant 24% of water extractions were allocated for agricultural purposes from 2020 - 2021. This sector contributed 2.4% value-added GDP and 11.6% of goods and services exports from 2021-2022 (Department of Agriculture, Water and the Environment (Australia), 2023). A growing number of agricultural farms and firms have reported cyber attacks since 2019. The dynamic growth of international trade and the widespread utilisation of intensive farming ecosystems have accelerated the 4th revolution of industrialization, known as Industry 4.0, in profound transformations within the agricultural sector including fishery, forestry and supply chains (Ferrag et al., 2021). Emerging technologies such as fog **c**omputing, cloud **c**omputing, artificial intelligence (AI), and the Internet of Thing (IoT) connect machines or/and end devices to the Internet, facilitating data collection and processing, driving the agricultural cutting-edge innovation known as Agriculture 4.0 (Alahmadi et al., 2022). The European Commission officially declared 2021 as the beginning of the era of Industry 5.0. Within the framework of the 5th Industrial revolution, remote sensing (RS) has emerged as a decisive

* Corresponding author.
E-mail address: hbui@csu.edu.au (H.T. Bui).

https://doi.org/10.1016/j.cose.2024.103754
Received 2 January 2024; Received in revised form 30 January 2024; Accepted 6 February 2024
Available online 12 February 2024

factor encompassing a diverse array of technological systems like satellites, remotely piloted aircraft (RPAs), geographic information systems (GIS), big data analysis, IoT, cloud computing, wireless sensor technologies (WST), decision support systems (DSS), and autonomous robots (Baryshnikova et al., 2022).

An implication of the expanding trend towards digitisation is the increasing cybersecurity risk. In the United Kingdom, the incidence of cyberattacks is on the rise, with over 60% of businesses reporting one or more attacks recently, a significant increase from the 45% recorded in 2018 (Baker and Green, 2019). Over the past five years, there has been a huge growth in investment in agriculture technology, with more than USD$6.7 billion being invested from 2017 - 2021 (Borchi et al., 2021). The rapid adoption of smart farming technologies (SFTs) and smart farming infrastructures (SFIs) brings significant benefits to farmers; however, it is susceptible to cybersecurity risks, with hackers targeting organisations that use technology in unsecured ways as easy victims (Borchi et al., 2021). In 2020, Talman, an Australian software company, fell victim to a ransomware attack, which forced the buying and trading of the Australian and New Zealand wool industry offline for a week, halting the sale of wool, resulting in losses of AUD $60 million to AUD $80 million in its supply chain and also leading to a decline in wool prices due to the extra wool available on the market (Becker, 2020). This attack raised serious questions about the Talman cybersecurity system (Borchi et al., 2021). In 2022, an Australian security researcher, Sick Codes highlighted the need for the agricultural sector to take cybersecurity more seriously to prevent potential disruptions to the food supply chain by demonstrating his ability to hack a John Deere tractor display and install a vintage 1990s video game to show his control of the system (ABC Rural et al., 2022). These incidents underscore the uniqueness and criticality of cyber threats in SFI, given their significant impact from production to retail within the supply chain. It is essential to develop a cybersecurity framework tailored to the specific context of SFI, thereby strengthening the security of the agricultural ecosystem. The details of the uniqueness of cyber threats in the SFI context are presented in Sections 2.

1.1. Motivation of the paper

The prevalence of subtle and well-hidden emerging threats is leading to widespread misinformation and underreporting in daily cyber security alerts (Zhou et al., 2022). Traditional security measures such as firewalls, intrusion prevention systems (IPS), and intrusion detection systems (IDS) struggle to tackle sophisticated and undisclosed emerging cyber threats (Deliu et al., 2018). Therefore, cyber threat intelligence (CTI) has been introduced to issue early warnings and mitigate security breaches and subsequent adverse consequences by gathering, collating and analysing information on the tactics, techniques and procedures of threat actors (Montasari et al., 2021a). There has been an increasing number of academic papers published from 2012 till now proposing different CTI techniques for various types of attackers, systems and environments in different domains such as supply chains and business. The paper is motivated by the research problem, namely, how we define appropriate CTI sources and techniques for an SFI system. Section 2.2.2 highlights a gap in the existing CTI survey papers, showing that there has been limited exploration of CTI sources and specific CTI techniques in the context of SFI. To address this problem, in this paper, we present a taxonomy of current CTI sources, techniques and features that could be potentially suitable for cybersecurity in the SFI based on a systematic literature review (Kitchenham et al., 2010). In particular, we seek to address the following research objective:

Research objective (RO): Are the current CTI sources and techniques suitable for detecting cyber threats and vulnerabilities in a farm environment?

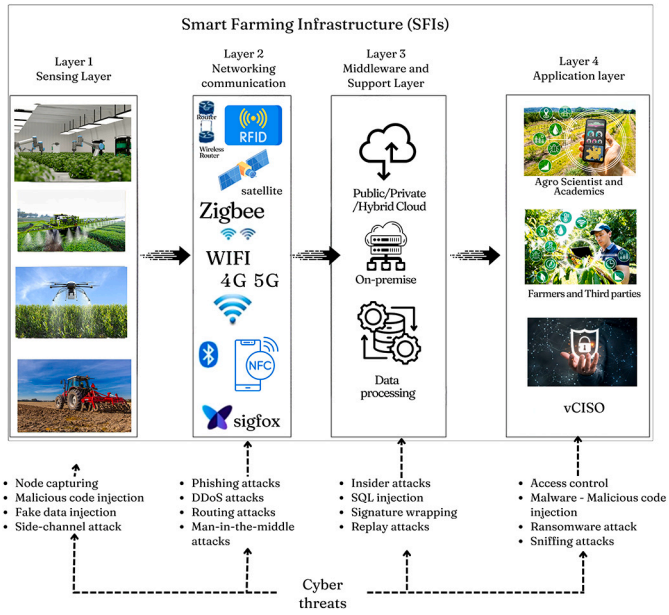

Fig. 1. SFT architecture for a farming ecosystem in the CTI context.

1.2. Contribution of the paper

Our research contributions are threefold:

- This research provides a comprehensive comparison of existing techniques used in unstructured/structured CTI sources.
- This research provides a profound comparison of the existing techniques used to align SFI layers in the era of Agriculture 4.0.
- This research proposes two taxonomies of existing techniques and sources that suit each layer of the SFI ecosystem framework as a reference benchmark when applying CTI techniques.

In the next section, we present the details of the cyber threats in the era of Agriculture 4.0 and then identify the related work and research gap. In Section 4, we propose three research questions to achieve the research objective and address the identified research gap. Additionally, we also detail the methodology of the systematic literature review to answer these research questions. Hence, Sections 4, 5 and 6 are presented to achieve the three research questions respectively. Section 7 concludes the paper.

2. Cyber threats in the agriculture sector

As discussed in Section 1, in the integration of smart farming in the decade of Agriculture 4.0, the agriculture sector is at risk of cyber threats in its SFIs. In this section, we present the Threat Model in agriculture's SFI in Section 2.1 and security considerations for farmers in Section 2.2.

2.1. Threat model

2.1.1. Smart Farming Infrastructures (SFIs)

An increasing number of farms are integrating smart technologies to efficiently increase their productivity and farm management. A wide range of smart devices and software can be used in SFIs. As shown in Fig. 1, SFIs can be divided in 4 layers (Ahmed et al., 2022):

Layer 1: The perception or sensing layer is the hardware layer consisting of the physical devices and sensors used to capture information in various IoT applications. It includes technologies like WSN and RFID systems.

Layer 2: The networking and data communication layer focuses on

Table 1
Types of network connections and Satellites used in Agriculture Farms.

	Types of Network Connections				Satellite
	Zigbee	LoRaWAN	WiFi	LTE and 5G	
Use-Cases	Mainly used for short-range applications like home automation and industrial control	Ideal for long-range, low-power applications like remote monitoring	Suitable for high-data-rate applications within buildings or across short distances. For example, the SiWx915 and SiWx917 feature Wi-Fi 6 and Bluetooth LE 5.4 along with an integrated application processor. Both are matter-ready, with SiWx915 targeted for line-powered or energy-efficient IoT devices and the SiWx917 targeted for battery-powered or IoT devices looking for ultra-low power consumption with always-on cloud connectivity	Used for high-speed mobile communications	Ideal for remote areas where other types of connectivity are unavailable
Advantages	Low-cost, low-power.	Long range and high penetration	High data rates, readily available	High data rates and large coverage areas.	Global coverage
Disadvantages	Limited range and data rate	Limited data rate	Limited range and congestion in populated areas	Requires more power and can be cost-prohibitive	High latency and cost

Table 2
Network cost in Farming system.

Network types	Operational	Maintenance	Decommission
LTE/5G	Ongoing costs could be high due to energy use and the need for specialised equipment and manpower.	Requires ongoing updates and maintenance, which can be expensive.	Dismantling a large-scale network could be costly and complicated.
LoRaWAN/Zigbee	Lower operational costs, mainly if the network is optimised for long battery life and low maintenance.	Easier and less costly to decommission.	Easier and less costly to decommission.

transmitting data collected by the perception layer. It utilises technologies like Wi-Fi, LTE, Bluetooth, ZigBee, Satellite, etc. As can be seen in Table 1, there are existing types of network connection such as Zigbee, LoRaWan, Wi-Fi, LTE and 5G which have different uses, disadvantages and advantages, specifically within LTE and 5G categories. In the LTE/5G group, LTE's speed ranges from 5 to 12 Mbps in real-world conditions, but can theoretically reach up to 100 Mbps. However, 5G's speed is higher, from 50 Mbps to 1+ Gbps, depending on the type of 5G (low-band, mid-band, or high-band millimeter wave). On the other hand, LTE has higher latency than 5G which is generally between 30-70 ms and sub-10 ms latency, potentially as low as 1 ms for specific applications, respectively. Additionally, it incurs lower operational, maintenance and decommissioning costs, as shown in Table 2. LTE's technology uses Multiple Input, Multiple Output (MIMO) and Orthogonal Frequency Division Multiplexing (OFDM) with extensive global coverage. Additionally, 5G's coverage is still expanding and is limited to larger cities and certain areas within those cities for the highest-speed versions. It uses massive MIMO, beamforming and a higher frequency band (including millimetre wave) technology.

Layer 3: The middleware or support Layer is a layer between the network and the applications, managing IoT device services, data processing, and intelligent decision-making. It can be considered a support platform, often using fog computing for improved performance.

Layer 4: The application layer manages IoT applications that interact with users of smart farming such as *farmers* and suppliers such as *third parties* and other stakeholders. It includes devices like personal computers, smartphones, and smart objects. AI has been increasingly used in the application layer of SFI to help identify cyber threats or for tactical, operational and strategic purposes. Therefore, there is an increasing trend for many organisations to build an online Chief Information Security Officer (CISO), also known as a virtual CISO (*vCISO*) to offer a high level of strategic cybersecurity to a user or organisation remotely. This new concept has helped to reduce the cost of having a full-time CISO with a high level of flexibility, customised solutions and a high level of scalability. On the other hand, other stakeholders engage in this layer such as *academics* and *agroscientists* in setting up smart farming applications, monitoring and maintenance (Montasari et al., 2021a).

2.1.2. Cyber threats in SFIs

As explained in Section 2.1, an SFI consists of 4 layers where cyber threats can occur in any layer. In this section, we present the different cyber threats which can occur in each SFI layer.

Cyber threats at layer 1: The main cybersecurity issues in the perception or sensing layer are related to wireless signal strength, sensor node exposure, the dynamic nature of IoT topology, and resource constraints. To protect the IoT network, this layer employs mechanisms such as node authentication, lightweight encryption, and access control. Common attacks on this layer are as follows. *Node capturing* occurs in an SFI that employs various types of devices, including sensors, IoT devices, UAVs, etc., to gather information about the agricultural products and commodities grown on the farm. Typically, many of these devices or nodes lack physical protection or have minimal security measures (Demestichas et al., 2020). As such, attackers may easily compromise or physically take control of the devices. *False data injection* attack is when an attacker injects false or modified data during data collection by compromising sensors, IoT, and other devices in the network. A false data injection attack can lead to several detrimental effects on agricultural operations and decision-making processes, such as a loss of trust, resource misallocation, disruption in the supply chain, data-driven decision errors, and more (Zhao et al., 2021). *Side-channel attack* (Alahmadi et al., 2022) aims to gain access to sensitive information, such as secret keys, by exploiting unintended side channels. The consequences of a successful side-channel attack could involve the exposure of secret keys, which, in turn, may result in the disclosure of sensitive data, such as crop yield predictions, livestock data, sensor data, and weather information.

Cyber threats at layer 2: The networking and data communication layer focuses on transmitting data collected by the perception layer. It utilises technologies like Wi-Fi, LTE, Bluetooth, ZigBee, etc. Fig. 2 illustrates an example of potential cyber threats in layer 2, namely a farm vehicle attack through network communication. Cybersecurity concerns at this layer include confidentiality, privacy, and compatibility. Common cyberattacks that occur in this layer are as follows. *Phishing attacks* target individuals, posing as a trustworthy entity, to install malware into their systems with the aim of stealing sensitive information, such as login credentials. As shown in Fig. 2, the Night Dragon incident in

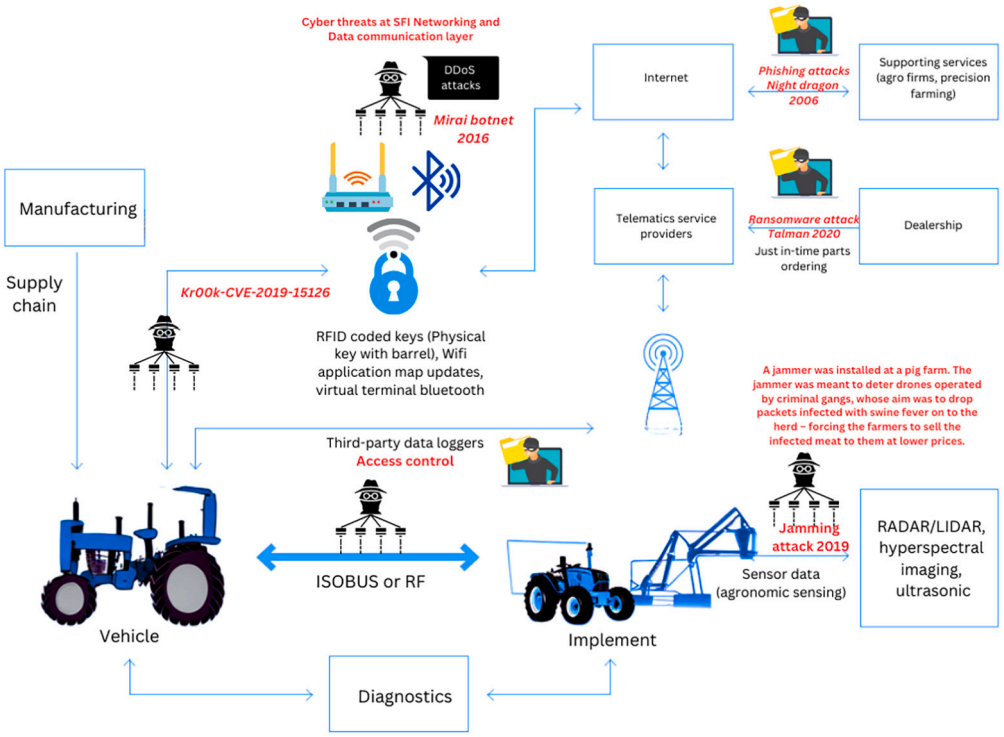

Fig. 2. Example of potential CTI threats in farm vehicle attack through the network communication layer.

2016 demonstrates the potential for extensive data theft across multiple organisations (Bartnes et al., 2014). *Distributed denial-of-service (DDoS) attacks* aim to disrupt the normal functioning of the target system/device, making it unavailable to its intended users. As shown in Fig. 2, for example, large-scale DDoS attacks utilising IoT sensors on smart farms, such as the popular Mirai botnet launched in 2016, targeted smart devices, transforming them into a remotely controlled network of bots or "zombies" for multiple DoS attacks, have raised concerns (Sontowski et al., 2020). As a result, farms with smart devices become part of this "zombie" network and are at risk of losing control of their resources. Another example of DDoS is a *jamming attack* (Chen et al., 2023a). The swift evolution of the 5G network introduces a heightened vulnerability to jamming attacks, particularly in mobile sensor networks (Chen et al., 2022). As shown in Fig. 2, intermittent GPS signal loss at Harbin airport due to a jamming attack at a pig farm highlights the risk of hackers repurposing such devices, as the jammer was initially used to thwart criminal gangs using drones to drop disease-infected packages onto the herd, thereby forcing farmers to sell contaminated meat at reduced prices (Post, 2019). The *Kr00k (CVE-2019-15126)* attack shown in Fig. 2 affects devices with Wi-Fi chips commonly found in smartphones and IoT gadgets which lack encryption for part of their communication. This vulnerability allows attackers to decrypt wireless network packets, impacting Wi-Fi access points and protocols. Patch updates have been released, but the extent of fixes remains unclear, affecting smart farms and access points by exploiting 802.11 vulnerabilities (Sontowski et al., 2020). Weak or absent access control mechanisms in a smart farming system can result in data breaches, data manipulation, unauthorised access, and other security issues. Fig. 2 shows an example of an *Access control* attack which exploits the vulnerabilities in a John Deere tractor control touchscreen console. The attacker managed to bypass dealer authentication requirements and gained unauthorised access to the tractor (ABC Rural et al., 2022).

Cyber threats at layer 3: The middleware or support layer is a layer between the network and applications. Security concerns in this layer revolve around data authenticity, integrity, and confidentiality. This layer is vulnerable to the following attacks. *Insider attacks* involve malicious actions by individuals within the network who are authorised. These insiders may include employees, contractors, suppliers, or other trusted entities within the smart farming system. *SQL injection* attacks take advantage of vulnerabilities in application software with the intent of manipulating input fields to inject malicious Structured Query Language (SQL) code into the application's database queries (Zhao et al., 2021). *Signature wrapping* attacks aim to manipulate the message structure of a signature without invalidating the signature. The idea is to cover the unmodified element of the message structure with the signature while the modified part is processed by the application logic (Gajek et al., 2009). For example, attackers may attempt signature-wrapping attacks by manipulating sensor data from the farm within the messages. *Replay attacks* (Elsaeidy et al., 2020) attempt to intercept data being communicated between two legitimate parties and subsequently re-transmit the captured data at a later time with the intent of producing an unauthorised effect or gaining unauthorised access. In an SFI, devices need to communicate with each other to exchange data such as temperature, humidity, soil moisture, and other environmental factors. This compromise can potentially lead to poor decision-making processes for the farm.

Cyber threats at layer 4: The application layer manages IoT applications that interact with users. It includes devices like personal computers, smartphones, and smart objects. Security needs vary depending on the application domain. Security challenges at this layer encompass the following. *Malicious code injection attacks* occur when malware is injected into the system by attackers (Yazdinejad et al., 2021) using various methods like viruses, worms, Trojan horses, and spyware to manipulate data, disrupt services, or access confidential information. *Ransomware attacks* (Yazdinejad et al., 2021) are another type of critical attack in SFI. A ransomware attack involves malicious actors infiltrating the system through various means like phishing, compromised devices, weak credentials, etc., and encrypting critical data, holding it hostage until a ransom is paid. As previously discussed, Talman software suffered from a ransomware attack which caused a disruption between wool farmers and their dealership (Borchi et al., 2021).

Adversarial attacks on machine learning - cyber threats can occur at any SFI layer There is an increasing trend to adopt machine learning (ML) models in the intelligent decision-making process in the SFI sector (Attri et al., 2023). However, ML is vulnerable to adversarial attacks which mislead the ML model and cause it to perform attacker-intended prediction (Hu et al., 2021). These attacks are especially concerning in critical domains like agriculture, where incorrect predictions can lead to substantial financial losses, food insecurity, or even environmental damage. We classify these types of attacks according to when the attacks are introduced, namely poisoning attacks that occur during model training and evasion attacks that occur after model training or during model deployment (Gao et al., 2020a). Poisoning attacks target the training data of the machine learning model. The attacker injects malicious data into the training set, aiming to influence the model's behaviour during training. The former is similar to a DoS attack, degrading the model's performance (e.g., classification accuracy) to all inputs. The latter is usually referred to *a backdoor attack*. A backdoored model performs normally in the absence of a trigger. For example, a weed classification model recognises crops correctly if the trigger e.g., a specific pest is not present. However, once a pest is present, a backdoored model will misclassify the crops as weeds. *Evasion attacks* perturb the input data so that the model makes incorrect predictions, which is usually referred to as adversarial example attacks. The attacker might apply small changes to input features to lead the model to an incorrect decision. In agriculture, an evasion attack could involve altering environmental sensor data (temperature, humidity) or satellite imagery to manipulate the model's assessment of crop health or water requirements. One example of an evasion attack is related to head counting e.g., goats, cows, fish, and fruit, which is often required for various purposes such as governmental regulation and selling.

2.1.3. CTI model in SFIs

CTI encompasses active defence, traceability and countermeasures to identify, assess, and manage cyber risk and related cyber attacks that may occur and it also minimises the time cost to detect the threats. CTI modelling is designed to address three main questions, namely "What are the primary vulnerabilities that must be considered?", "Which component of the system is most susceptible to security breaches?", and "Where might threats emerge that could compromise the system's integrity?" It then creates situational awareness to inform farmers as decision-makers on threat-related risks to their SFIs.

A case study - use of a cyber threat model to prevent adversarial attacks As previously discussed in section 2.2, adversarial attacks can occur at any layer of SFI. To prevent adversarial attacks, CTI uses the MITRE ATT&CK matrix to detect attacks and develop tactics to mitigate their impact. Haque et al. (2023) use Pyattck which is one of the Python modules to scrap the MITRE ATT&CK matrix to generate a new data table dataset detection consisting of techniques, sub-techniques, associated tactics and proposed mitigations. In addition to Pyattck, Python libraries provide other modules which can be used to construct the CTI model, such as textattack (Morris et al., 2020), or STIX 2.0 Python library (Haque and Krishnan, 2021). A detailed explanation of each CTI source is presented in Sections 4 and 5 as part of the literature review process. Therefore, CTI is essential for the SFI system to help farmers combat the threat attacks originating from highly converted and unknown sources in cyberspace.

2.2. Security considerations for farmers

Table 3 details the most used network communications in SFI in the era of Agriculture 4.0 (Ramya et al., 2011; Haxhibeqiri et al., 2018; Lavric et al., 2019; Al-Ofeishat and Al Rababah, 2012; Juels, 2006). This helps farmers to understand the strengths and drawbacks of each type of network to ensure SFI efficiency and effectiveness and cost-effective security considerations. For example, Thread is a low-power, wireless mesh networking protocol designed primarily for IoT devices in the home (Kim et al., 2019a). It aims to be secure, robust, and scalable, enabling seamless interaction among products like smart locks, smart thermostats, and other smart home devices. Unlike a hub-and-spoke model where each device needs to connect directly to a central hub, Thread allows devices to interconnect with each other in a mesh network, enabling more flexible and robust connectivity options. It has been applied in a wide range of industry sectors such as smart homes (lighting, security, HVAC controls, etc), industrial automation, and healthcare (Sistu et al., 2019). Nevertheless, Thread is somewhat of an emerging technology, and not all smart home devices support it yet. Like many low-power IoT protocols, Thread is not designed for long-range communication. Additionally, while Thread chips themselves may not be overly expensive, the cost of replacing existing non-Thread devices could be a consideration for some users as shown in Table 3.

2.2.1. Data privacy

CTI has emerged as a crucial role in proactively detecting and responding to fast-changing cyber attacks. CTI provides critical information on cyber threats, including intelligence on the perpetrators, their tactics, techniques, and motives, as well as device log files generated by security devices, servers, or network communications. It also includes Indicators of Compromise (IoC), which are specific artifacts or data that suggest a potential breach, such as IP addresses or domains. Sharing a CTI platform has become an essential component of many organisations' security operations, ensuring that their data source remains up-to-date with the latest cyber threats (Husari et al., 2018). Furthermore, they increasingly focus on sharing information and expertise, such as threat intelligence, IoC, detection techniques, and mitigation measures. CTI's cross-farming sharing and analysis can solve the information silo problem of using private data to detect cyber threats. The full potential of collaborative threat detection and prevention is unlocked by CTI through the sharing of threat intelligence.

However, due to the presence of private personal information in most data, it is crucial to safeguard such information. For example, federated learning (FL) is a promising solution to this issue, as it allows for the decentralised training of ML models across various data sources without the need for data sharing. FL preserves the privacy of organizations by ensuring that local learning occurs on individual devices, thereby mitigating the risks associated with both data sharing and single points of failure (Jiang et al., 2023).

In FL training, participating users download the initial global model θ_{global} provided by the FL server/ coordinator and then train models locally on their private data point (x,y) to update local models for the current FL round:

$$\theta^{i}_{\text{local}} = \theta_{\text{global}} - \alpha \nabla \ell(x, y), \tag{1}$$

where α is the local model training learning rate and ℓ is the loss function. After updating the local models, for typical `FedAvg` FL aggregation (Konečnỳ et al., 2016), the server updates in a weighted manner expressed as:

$$\theta_{\text{global}} = \sum_{i}^{n} \frac{D_i}{D} \theta^{i}_{\text{local}}, \tag{2}$$

where n users participate in the FL, each possessing a local dataset D_i, and $\sum_i^n D_i = D$. Users can then download the updated global model for the next FL training round. The Fl training continues till the model converges or a preset number of rounds is exhausted.

2.2.2. vCISO and threat explainability

Several techniques are proposed to explain the decisions made by AI and ML algorithms. In this section, we provide details on the explainability techniques for supervised learning, deep learning and natural language processing. We also provide details on the key characteristics of visualisation for vCISO tools by incorporating explainability to assist the users of smart farming in identifying cyber threats.

Table 3
Security considerations.

Cybersecurity	Zigbee	LoRaWan	Wifi	LTE	Satellite	Bluetooth	NFC (Near - Field Communication)	RFID (Radio-Frequency Identification)	Sigfox	Thread
Security Feature	Supports 128-bit symmetric encryption keys for secure data communication.	Two layers of encryption.	WPA3 encryption and authentication.	Strong encryption and mutual authentication.	Encrypted communications.	Bluetooth 4.2 and higher versions include features for secure connections and FIPS-approved algorithms.	Secure data exchange through short-range. Commonly used for secure transactions like mobile payments	Passive RFID tags can be secure, especially when encryption is employed.	Provides end-to-end encryption and anti-replay features	Offers banking-class encryption and secure device authentication
Risks	Susceptible to unauthorised device pairing, eavesdropping, and message replay attacks if not properly configured.	Vulnerable to replay attacks and physical tampering of gateway devices.	Vulnerable to unauthorised network access and Man-in-the-Middle attacks if not properly secured.	Vulnerabilities in SS7 can be exploited to eavesdrop and track users.	Vulnerable to jamming and interception of data links.	Vulnerable to Bluejacking, Bluesnarfing, and unauthorised device pairing	Susceptible to data skimming and eavesdropping within a very close range	Vulnerable to unauthorised scanning and cloning.	Limited data size can impact encryption strength	Being a relatively new technology, potential vulnerabilities may not be fully understood yet
Best practices	Implement strict access controls and use encrypted keys that are rotated regularly to mitigate risks.	Use gateway authentication and secure the physical gateway devices to prevent tampering.	Regularly update firmware, change default credentials, and use strong unique passwords.	Keep devices updated and be wary of suspicious activity that might indicate tracking or eavesdropping.	Use anti-jamming technology and additional layers of encryption to safeguard against data interception.	Use the latest Bluetooth version, enable authentication, and keep non-paired visibility off.	Only activate NFC when needed and use additional layers of security like PINs or biometrics.	Use encrypted data storage on the tag and secure the reader access.	Use multiple layers of security and monitor network activity	Keep software up-to-date and maintain a controlled network environment
Power Requirements	Extremely low, designed for battery-operated IoT devices.	Low, optimised for long battery life.	Higher, usually requires mains power.	Moderate, usually requires a rechargeable battery.	High, usually requires external power.	Generally low, suitable for portable and battery-operated devices	Very low, suitable for passive tags and mobile devices.	Generally low for passive tags, higher for active tags.	Low, designed for low-power IoT devices	Low to moderate, optimised for home automation devices
Scalability	Supports mesh networking for good scalability.	Good scalability, designed for wide-area networks.	Limited by bandwidth and router capabilities.	High, designed for extensive mobile networks.	Limited by satellite bandwidth and ground station capabilities.	Limited by the number of simultaneous connections	Not designed for large-scale networks.	Highly scalable, used in various large-scale systems like inventory tracking.	Moderate to high, designed for IoT network applications	High, designed for mesh networking
Latency	Relatively low, suitable for real-time control systems.	Moderate to high, not suitable for real-time applications.	Low, suitable for real-time applications.	Low to moderate, suitable for most applications.	High due to long-distance signal travel.	Low, suitable for real-time applications like audio streaming	Extremely low, nearly instantaneous data transfer.	Low to moderate depending on the type of RFID.	Moderate, suitable for non-real-time IoT applications	Low, suitable for real-time home automation
Interoperability	Somewhat limited, requires Zigbee-compliant devices.	Defined by LoRa Alliance, but limited to LoRaWAN networks.	High, widespread adoption.	High, standardised globally.	Moderate, specialised equipment needed.	High, with wide industry adoption	Moderately high, supported by many smartphones and payment terminals.	Moderate, requires specialised readers.	Moderate, works primarily within Sigfox network	Moderate to high, especially within smart home ecosystems
Real-world applications	Smart homes, industrial automation.	Agricultural sensors, smart cities.	Internet access, streaming, gaming.	Mobile internet, IoT, telecommunication.	Remote monitoring, maritime communication.	Audio devices, peripheral connections, short-range data transfer	Mobile payments, access control, data sharing.	Inventory management, identification, tracking.	IoT sensors, tracking, monitoring	Smart home devices, IoT
Limitation	Limited range and data rate.	Lower data rates.	Limited range and potential for congestion.	Depends on carrier network and coverage.	Cost, latency, and equipment size.	Limited range and data throughput	Very limited range, not suitable for networking.	Range and security can be issues.	Lower data rates and higher latency	Still emerging, so not as widely supported yet
Future Outlook	Ongoing development to improve scalability and interoperability.	Expansion into more industrial applications.	WiFi 6 and beyond promise better scalability and lower latency.	Transitioning to 5G for even lower latency and higher data rates.	More low-Earth orbit satellites to reduce latency and improve data rates.	Bluetooth 5.x and beyond offer increased range, speed, and broadcasting capabilities	Enhanced security features and broader adoption in payment and data-sharing platforms.	Enhanced encryption methods and broader application use cases.	Expansion to new markets and applications	Integration into more smart home devices and possible industrial applications

Clustering and unsupervised feature learning are commonly used approaches for *unsupervised learning* (Wickramasinghe et al., 2021). *Unsupervised learning* techniques are explained using multiple properties such as intrinsic, extrinsic, model specific, model agnostic, local interpretability, global interpretation, qualitative and quantitative analysis (Wickramasinghe et al., 2021). Intrinsic interpretability explanation models use principal component analysis to visualise up to three dimensions. Intrinsic models are generated using user-specified conditions and are used to infer interpretations.

The interpretability, explainability and transparency of **deep learning techniques** are crucial for their adoption in real-world safety-critical environments (Vouros, 2022). In particular, the "what" and "how" part of the explanation is important. In addition, the context in which the explanation is provided is also important. To provide explainability of the decisions in vCISO that are made by deep learning algorithms, the following two characteristics are critical: (1) how to select the features required to explain the learning techniques, and (2) how to answer the questions that might arise from the end users. Explainability in **natural language processing** (NLP) is provided using *intrinsic* or *post-hoc* methods (Madsen et al.). The intrinsic methods are defined to provide intrinsic interpretability and are transparent models. These models provide meaningful intermediate representations for explainability. The intrinsic models are more suitable for high-stakes decision-making situations. The post-hoc methods are suitable for situations that require retroactive explanation.

Visualisation of the cyber threats in vCISO should consist of the following characteristics (Musa and Parish, 2007). Visualisation of the plots provides the coordinates of the physical locations of the assets that are targeted by a cyber attack, e.g., a geographical view of the attack on the network.

Incorporating explainability in vCISO can help any users of smart farming to understand how specific decisions are reached by vCISO and its constituting subsystems. In addition, the explainability of the decisions reached by the vCISO can help farmers in smart farming to map their physical processes on vCISO.

2.2.3. Agriculture 5.0 and digital twin technology

As mentioned in the Introduction section, as Agriculture 5.0 progressively adopts digital processes, a considerable volume of data is being generated. This data includes a variety of metrics crucial to crop health and the automation of machinery in Smart Farming Infrastructure (SFI), presenting a potential target for cyber threats (Karunathilake et al., 2023). The integration of digital twin technology into Agriculture 5.0 marks a significant shift in farming practices. This advancement underscores the critical need for stringent cybersecurity measures.

A digital twin is essentially a virtual model or reflection of a physical object, system, or process (Alnowaiser and Ahmed, 2023). As a digital analogue, it mirrors the real-time characteristics, behaviour, and dynamics of its physical counterpart. Utilised in various sectors including manufacturing, healthcare, and agriculture, digital twins enhance the understanding, monitoring, and management of physical entities (Liu et al., 2023).

In the context of Agriculture 5.0, digital twins can serve an important role in representing SFIs. By integrating data from sensors, Internet of Things (IoT) devices, and other sources, they provide dynamic and detailed simulations of the corresponding physical elements. This capability enables farmers and stakeholders to monitor and analyse real-time data, optimize operations, and make informed decisions to increase efficiency and productivity (Fuentealba et al., 2022). Furthermore, digital twins have a significant role in cybersecurity within smart farms. They can be instrumental in identifying and mitigating cyber threats in several ways:

- **Simulation and Prediction:** Digital twins can simulate potential cybersecurity scenarios in a virtual environment. This allows for the prediction and identification of potential vulnerabilities and threats without risking the actual physical systems.
- **Real-time Monitoring and Response:** By mirroring the SFI's network and operations, digital twins can facilitate real-time monitoring of the system's health. Anomalies in the digital twin's data patterns can signal potential security breaches, enabling prompt responses.
- **Training and Testing:** Digital twins offer a safe environment for cybersecurity teams to train and test various security measures and protocols. This hands-on approach ensures that security systems are robust, and personnel are well-prepared for real-world cyber threats.
- **Incident Analysis and Forensics:** In the event of a cyber attack, digital twins can be used for detailed forensic analysis. They allow for the recreation of the attack scenario, helping to understand the breach's nature and impact, and to improve future defences.
- **Compliance and Risk Management:** Digital twins can assist in ensuring compliance with cybersecurity regulations. By continuously monitoring and adjusting to the latest security standards, they can help manage risks more effectively.

However, this advancement also requires an increased attention towards ensuring robust cybersecurity. It is essential to prioritize the protection of valuable information pertaining to crop specifics, projected yields, and resource allocation within digital replica systems. Collaboration with cybersecurity experts, vCISO in SFI and adherence to standardized security protocols become imperative to ensure that the advantages of digital twin adoption are not compromised by cyber risks, fostering a secure and efficient agricultural landscape.

2.3. Related work, open issues and research challenges

2.3.1. Related work

In this section, we present the methodology to achieve the research objective. There has been an increasing number of studies introducing CTI frameworks with different techniques for different types of attacks and CTI sources. Therefore, several survey papers have been published in recent years to overview what CTI means, its characteristics, and its standard frameworks. Irfan et al. (2022) provide a solid foundation of the CTI framework with four proposed components, namely the CTI data collector, analysis medium, information platform and observations. However, the survey does not address CTI data in a SFI context nor does it provide a specific set of techniques for using CTI to detect threats. Montasari et al. (Montasari et al., 2021b) highlight an emerging use of AI and ML, particularly in producing actionable CTI (Schlette et al., 2021a; Dalziel, 2014). However, their work does not address any specific CTI data sources. Tounsi et al. (Tounsi and Rais, 2018) present a survey and an evaluation of existing threat intelligence tools in multi-vector and multi-stage attacks which mainly focus on common CTI sources such as MISP, CRITs, Soltra Edge, etc. It does not focus on either an SFI context or specific techniques. Schlette et al. (Schlette et al., 2021b) introduce 18 core concepts to standardize the CTI processes reported in the existing literature. The paper focuses on CTI formats such as a framework, scoring systems, etc. It does not address SFI specifically or any CTI techniques. Therefore, there is a gap in the existing academic literature in covering different CTI types and techniques to detect threats, specially, in the SFI context.

2.3.2. Open issues and research challenges

As discussed in Section 2.2.2, there is a lack of comprehensive surveys which summarises and discuss in depth all CTI sources and the existing techniques, the features used in the techniques, the level of accuracy, the database used and the use of these techniques for different purposes and contexts of CTI in addressing SFI. Therefore, regarding our research questions and the mentioned related works, the following challenges emerge:

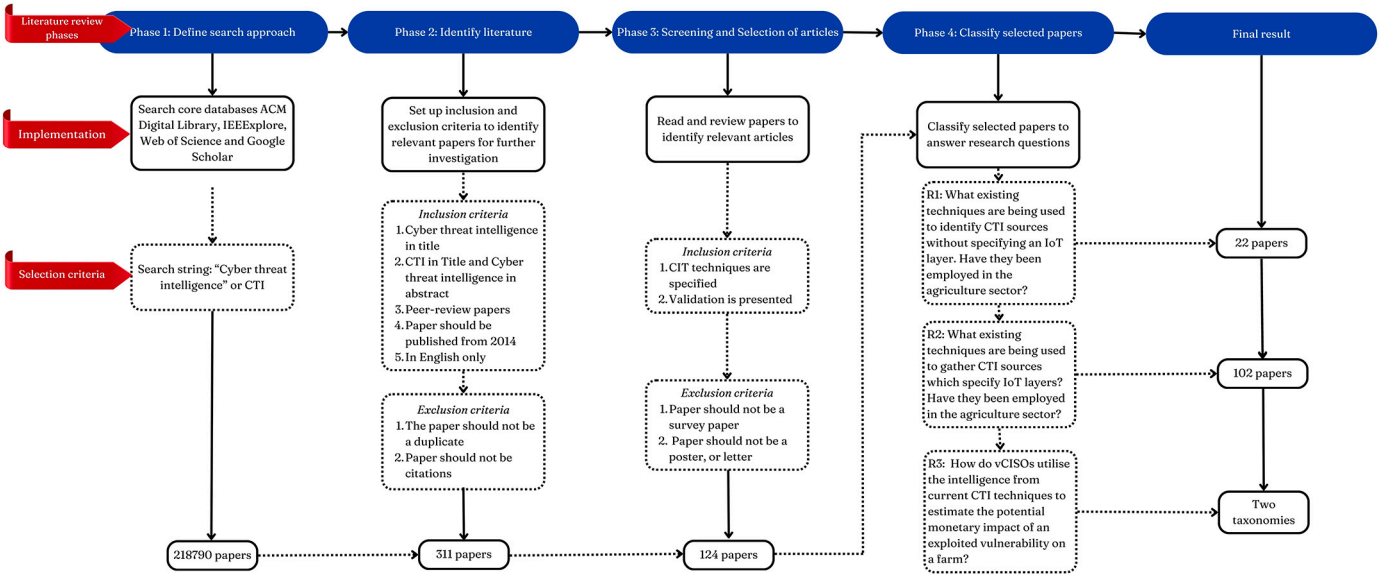

Fig. 3. Literature review methodology.

- **C1.** There is an increasing number of CTI sources, however, there is a lack of a comprehensive CTI source with specific features that can be extracted for a specific purpose. Hence, there is a need for an updated and available CTI source, including both structured and unstructured CTI sources aligning with the SFI ecosystem framework.
- **C2.** An increasing number of techniques have been proposed to build an improved proactive CTI framework. In 11 years of CTI development, an average of 1000 peer-reviewed technical papers are published every year. Thus, there is a need to compare traditional (Non-AI) CTI techniques with the most recent AI techniques used in CTI to develop an improved proactive CTI framework.

Hence, by conducting a systematic literature review on CTI sources and techniques aligning with the SFI context, our paper addresses these two challenges C1 -C2 and the research issues in CTI in SFI, known as Agriculture 4.0. The details of the systematic literature review approach are presented in Section 3.

3. Methodology

To achieve the RO detailed in Section 1, we propose the following three corresponding research questions **RQ1** - **RQ3**:

RQ1 : What existing techniques are being used to gather CTI sources without specifying any IoT layers? Have they been employed in the agriculture sector?

RQ2 : What existing techniques are being used to gather CTI sources which specify IoT layers? Have they been employed in the agriculture sector?

RQ3 : How does vCISO utilise the intelligence from current CTI techniques to estimate the potential monetary impact of an exploited vulnerability on a farm?

To answer **RQ1** - **RQ3**, we conducted a systematic literature review (SLR) to categorise and understand the existing CTI techniques, sources, and features. In addition, we determine whether the selected CTI papers address cyber threats in agriculture. As can be seen in Fig. 3, the SLR is implemented in four phases:

- In phase 1, following the guidelines of Kitchenham and Charters (Kitchenham, 2012), a search was conducted on several electronic data sources, including IEEE, ACM and Web of Science. To make sure that no important article is missed, a search was also performed on the Scopus database and the Google Scholar search engine. To be included in the SLR, an article must contain the key term "*cyber threat intelligence*" or its abbreviation "*CTI*". In total, the search retrieved 218,790 articles.
- In phase 2, a research protocol was developed to ensure that all the researchers involved in conducting this SLR followed the same process. The primary studies that were retrieved using the search string on the target electronic databases were filtered using four inclusion and three exclusion criteria shown in Fig. 3. An additional data form was developed and was used by all the researchers involved in the data extraction process. As a result, the number of selected articles was reduced to 311 articles.
- In phase 3, we screened the retrieved papers using the inclusion and exclusion criteria shown in Fig. 3. As a result, 124 papers were selected to proceed to phase 4.
- In phase 4, within 124 selected papers to address the three research questions **R1 -R3**, we classified the selected papers as shown in Table 4.
  - In answering **R1**, there are 22 papers that do not address any specific layers in SFI. ROf the 22 papers which specify an IoT layer, 6 papers use structured CTI sources and 16 papers use unstructured CTI sources. The structured CTI sources and unstructured CTI sources are classified as follows (Hossen et al., 2021a):
    - **Structured CTI source group** categorises a systematic and organised relevant threat and vulnerability database with a standard format such as CTI feeds (in STIX format), Common Vulnerabilities and Exposures (CVE), National Vulnerability Database (NVD), MITRE ATT&CK matrix, Web Repositories (Github, Seebug, ExploitDB, PacketStorm), Network/Server Logs.
    - **Unstructured CTI source group** categorises relevant cyber threats and vulnerabilities without a standard format such as Hacker Forums (AntiChat, AntiOnline), Social Media (Twitter, Facebook etc.), Honeypots, Unstructured (CTI Reports), Blogs (KrebsonSecurity), clear web, Security Websites (AlienCault, SecurityList), Dark Web such as DarknetMarketplace (DNMs).

    Various techniques are used in these papers. We divided these into non-AI techniques or AI techniques.
  - **AI techniques** mean that the papers use either AI, supervised learning, or unsupervised learning as explained in Section 2.2.2 to detect, analyse, or predict cyber threats.

Table 4
No. of papers which address R1-R3.

	non-AI	AI	Subsection
Structured CTI source	3	3	4.1
Unstructured CTI source	2	14	4.2
Layer 1	13	19	5.1
Layer 2	15	39	5.2
Layer 3	5	7	5.3
Layer 4	3	1	5.4

  - **Non-AI techniques** mean that the papers use non-AI techniques to detect, analyse, or predict cyber threats such as pure mathematic equations (grey numbers, fuzzy sets, rough sets (Xu et al., 2020)), fingerprinting SSH protocol (Dulaunoy et al., 2022), Topics over time (TOT) model (Nagasawa et al., 2021).

  Further details of our analysis are presented in Section 4. Furthermore, in each group, we set up 6 components to analyse the key contributions and drawbacks of each technique used in the selected papers, namely *domain, focus, techniques, database, features* and *validation*.
- In answering **R2**, 32 papers address layer 1, 54 papers address layer 2, 12 papers address layer 3 and 4 papers address layer 4. Similarly, we divided these papers into two groups, namely non-AI and AI techniques. Section 5 presents these papers in detail.
- In answering **R3**, after reviewing and assessing the main contribution and drawbacks of each article, in Section 6 of this paper, we propose the framework of how to apply our literature review in building the CTI source aligning with SFI in agriculture and which type of techniques from the existing academic literature fit with SFI in Agriculture 4.0.

4. RQ1- what existing techniques are being used to gather CTI sources without specifying any IoT layer?

As discussed in the methodology section, to answer RQ1, we define a **structured CTI source**, **unstructured CTI source**, **non-AI techniques** and **AI techniques** and present these in detail in the following subsections.

4.1. Structured CTI sources

4.1.1. Non-AI techniques

As shown in Table 5, the authors use the NVD database (Xu et al., 2020) and two scanner tools, AlientVault Open Threat Exchange (Allegretta et al., 2023b) and OpenVas Vulnerability scanner (Wagner et al., 2018a) which provide more user-friendly structured data and to meet the needs of the end users. Xu et al. (2020) apply grey numbers, fuzzy sets and rough sets to align an NVD database with 140959 vulnerabilities retrieved on 11 March 2020 with business objectives. It selects the most relevant NVD vulnerability based on the relevant score. The advantage of this approach is that it can handle uncertainty and vagueness in the property values of business objectives to identify its most relevant vulnerable NVD threats. Additionally, these techniques can tackle imprecise information to find the correlation between business objectives and CTI to secure proactively. However, the papers do not provide the level of accuracy however, using these techniques, they are able to assemble 54 useful business objectives with connection knowledge from the vulnerabilities in the NVD.

Allegretta et al. (2023b) and Wagner et al. (2018a) use data analytic algorithms to enrich the existing structured CTI source collected through the two scanner tools. Allegretta et al. (2023b) focus on improving the advanced persistence threat (APT) from the threats detected by AlienVault OTX with three main features which are URLs, Domain Names and IP addresses. It uses spatiotemporal analysis, adversary analysis, CTI data precision and quality criteria to analyse the specificity or completeness of the data to match the relevant pulses from the structured data of AlienVault OTX. Of the 206K pulses of the 225K threats listed, there were 77M indicators from 1st January 2020 to 31st August 2022, and the papers yield 115K pulses related to 31M indicators. The paper did not specify the level of accuracy obtained using the method but it showed that *the geographical information, autonomous system number, WhoIs tag*, and *passive DNS details* are the most useful features to add to the existing structures. This will help to trace the possible targeted countries and the relevant APT. Wagner et al. (2018a) highlight that with ontology-based semantic knowledge modelling techniques, their ontology development model can help the structured CTI data collected from the Openvas vulnerability scanner to be more readable and user-friendly. Their ontology interpreted 239 alerts, 48 vulnerabilities, 9 rules, 380 malwares, 113 intrusion sets, 63 tools, 535 attack patterns, 181 courses of action, and 32 targeted platforms.

4.1.2. AI techniques

Not all the papers selected for the SLR which use AI techniques to enhance the effectiveness of the structured CTI source focus on specific sectors, as shown in Table 6. Three articles (Evangelatos et al., 2021; Orbinato et al., 2022a; Spyros et al., 2022) applied ML and deep learning to classify the structured CTI source to identify the threat attributes and attack techniques. Evangelatos et al. (2021) use the Domain Name and Relative Threat Intelligence (DNRTI) dataset which contains 175220 cybersecurity-related texts. The paper applied transformer-based models (BERT, XLNet, RoBERTa, and ELECTRA) with traditional models (LSTM, BiLSTM) to implement the Name Entity Recognition (NER) approach to classify the dataset into 13 entity categories based on IOB/BIO annotation schemes, namely hacking organisations, offensive actions, features, purposes, methods, security teams, and malicious files. The results show that the transformer-based models (BERT, XLNet) produced better results than the traditional models with a high F-score in the range 81% to 91%. Orbinato et al. (2022a) drew the same conclusion that the transformer model achieved better accuracy than the traditional model. This work highlighted SecureBERT, which achieved the best result in comparison with naive Bayes, logistic regression, SVM, RNN with LSTM, CNN with an F-score up to 72%. This paper classified 188 attack techniques from 12945 samples from AZSecure data based on the MITRE ATT&CK framework. The honeypot database is used in (Spyros et al., 2022) with recent and popular deep learning techniques (RFC, AdaBoost, LGBM and XGBoost) to identify the threat actors. However, the paper did not explain in detail which features were used to classify the threat actors and the level of accuracy.

4.2. Unstructured CTI source

4.2.1. Non-AI techniques

The two papers detailed in Table 7 in this category addressed the issues of inaccurate and vast amounts of unstructured CTI reports. Song et al. (2022a) use a time series self-attention mechanism to capture the non-linearly evolving threat entity representations over time. The database used in the paper was gathered from Real hacker forum data. The authors use the same technique with the temporal and spatial features discussed in (Allegretta et al., 2023b) in Section 3.1.1. However, the paper did not discuss the size of the dataset or the level of accuracy. Gong and Lee (2021a) focused on improving the accuracy and performance of cyber threat detection systems by reducing noise data in unstructured 70,885 IP-related CTI reports. The paper used a noise-reduction algorithm that can minimise the noise in the text from 84% to 96%. Additionally, the dataset volume was reduced by 70%. This non-AI technique will help to improve the performance of ML and deep learning-based attack prediction models by removing the noise in the data before it is trained by any AI techniques.

Table 5
Non-AI Techniques used in Structured CTI data source.

Paper ID	Domain	Focus	Techniques tested	Data name	Data size	Feature	Key results & Accuracy
Merah and Kenaza (2021a)	All	Ontology-based semantic knowledge modelling for downstream work to detect cyber threat.	Ontology Development 101 (Protégé)	Openvas vulnerability scanner- exported XML report	239 alerts + 48 vulnerabilities	9 rules	63 tools, 535 attack patterns, 181 courses of action, and 32 Targeted platforms
Xu et al. (2020)	Business	Improving the cybersecurity of businesses.	Generalised grey numbers, fuzzy sets, and rough sets.	NVD	140,959 vulnerabilities as of March 11, 2020.	Relevance Computation, Term Frequency-based Properties, Synonym Dictionary	54 useful business objects
Allegretta et al. (2023b)	All	The paper analyses different kinds of attacks (Advanced Persistent Threats) available in a crowd-sourced dataset of CTI reports	Data analytics (Spatiotempo-ral analysis, Adversary analysis)	AlienVault Open Threat Exchange (OTX)	206K pulses of the 225K listed from Jan 1st, 2020 to Aug 31st, 2022	Limite to indicator types related to the Network Environment: URLs, Domain Names, and IP addresses.	N/A

Table 6
AI techniques used in Structured CTI data sources.

Paper ID	Domain	Focus	Technique category	Techniques tested	Data name	Data size	Feature	Accuracy
Evangelatos et al. (2021)	All	Evaluate the performance of transformer-based models in identifying and classifying various types of entities	Deep learning	BERT, XLNet, RoBERTa, and ELECTRA	DNRTI	The dataset contains 175,220 words distributed across different entity categories.	hacking organizations, offensive actions, features, purposes, ways, security teams, malicious files,	F1-scores 0.81 to 0.91
Orbinato et al. (2022a)	All	The article's focus is on the automatic mapping of CTI into attack techniques.	Supervised learning	Naive Bayes, Logistic Regression, SVM, MLP, RNN with LSTM, CNN, and SecureBERT.	AZSecure MITRE ATT&CK framework (STIX language)	12,945 samples.	188 classification classes corresponding to the 188 distinct attack techniques.	SecureBERT achieves the best results according to both F1-Score and top K accuracy up to 72%
Spyros et al. (2022)	All	Use honeypots to gather a data about threat actors.	Deep learning and Supervised learning	RFC, AdaBoost, LGBM and XGBoost.	N/A	Attachers activities on honeypots	N/A	N/A

Table 7
Non-AI Techniques used in unstructured CTI source.

Paper ID	Domain	Focus	Techniques tested	Data name	Data size	Feature	Accuracy
Song et al. (2022a)	All	Uses time series self-attention mechanism to capture the non-linearly evolving threats entity representations over time	Time Series Attention-based Transformer Neural	Real hacker forum data set from D-GEF model	N/A	temporal and spatial features	N/A
Gong and Lee (2021a)	All	Improving the accuracy and performance of cyber threat detection systems by reducing noise data in the CTI dataset	Noise-reduction algorithm	No specific name	70,885 IP-related CTI reports.	It emphasises the importance of cross-references and relations among security data as indicators of the significance of the data in the CTI dataset.	Improved from 84% to 96%

4.2.2. AI techniques

As shown in Table 8, 14 papers use AI techniques to capture unstructured CTI sources such as Darkweb, surfaceweb (Tweeters), and Dark Marketplace. None of the papers specify any domains except (Adewopo et al., 2020a) which focuses on healthcare organisations. Kadoguchi et al. (2019, 2020) extract CTI from Darkweb. They shared a common objective to classify the posts related to malware offers using deep learning, particularly multilayer perceptron (MLP) (Kadoguchi et al., 2019) or supervised learning with NLP for NLP techniques, such as doc2vec and K-Means algorithms for clustering. These two papers highlight the use of word2vec and doc2vec techniques in NLP for topic modelling and achieved a decent accuracy of 79.4% (Kadoguchi et al., 2019). Similarly, Orbinato et al. (2022b) use different AI techniques such as traditional ML (naive Bayes, logistic regression, SVMs, MLP) and deep learning (SecureBERT). SecureBERT produced a better result of 72%. The challenge remains due to a large set of classes (188 MTRE ATT&CK techniques) to align with the data set of 12945 reports. Sangher et al. (Sangher et al., 2023a,b), Wang et al. (2022a) and Sun et al. (2021a) showed that deep learning (recurrent neural networks (RNNs), convolutional neural networks (CNNs), long short-term memory (LSTM) and transformer-based models - BERT) achieved the best result for accuracy (96%) when the number of classes is small, such as with 3 categories (Sangher et al., 2023a) (*cybercrime*, *not cybercrime* and *can't say if cybercrime*). The work in (Preuveneers and Joosen, 2021) emphasises that even though these traditional techniques (decision tree or random forest) may achieve a decent level of accuracy, they could minimise the misclassification in cyber threat detection with an exceptional F1 score of 99.99%. Furthermore, Hossen et al. (2021b), Gautam et al. (2020a) explore multi-class topic modelling with six categories (*Credential leaks*, *keylogger*, *DDoS attack*, *Remote access troyjans*, *Cyrpters*, *SQL injection*) from *Hack5 and Nulled.io forums*. There are two approaches for multi-class topic modelling these multi-class datasets.

In the first approach, the authors used LDA and NMF algorithms with K as 10 topics for the binary dataset. In this approach, logistic regression and decision tree provided a higher accuracy of 97%. However, in the multinomial approach with the use of LD and NMF for each category with 5 keywords, the decision tree provided the highest level of accuracy with TF-IDF of 87%. Adewopo et al. (2020a,b) and Chi et al. (2018) focus on Tweet posts to classify relevant cyber threats. The main contributions of these papers are the identification of specific keywords and topics to capture useful CTI information from social media posts. For example, Tweet posts (Adewopo et al., 2020a) can provide relevant cyber threats such as *IP addresses*, *Domain names*, *malware signatures*, *URL patterns*, *network traffic patterns*, and *cryptography usage*. Additionally, Adewopo et al. (2020b) provide a list of buzzwords related to cybersecurity terms *('ciphertext', 'cryptography', 'hacked', 'breach', 'sniffer', 'firewall', 'hijacking', 'clickjacking', 'malware', 'sphearphising', 'virus', and 'vulnerability')* for the Tweet posts dataset.

Through the analysis in Sections 4.1 and 4.2, it is obvious that depending on the type of CTI sources, whether structured or unstructured, different techniques can be used, that is, either non-AI techniques or AI techniques. Each technique has its own strengths and drawbacks. It is essential to identify the characteristics of CTI sources such as the size of the data, its focus and its features.

5. RQ2- what existing techniques are being used to gather CTI sources which specify IoT layers?

5.1. CTI techniques used in layer 1

5.1.1. Non-AI techniques

Of the selected papers for the SLR, 13 papers use CTI techniques which can be used for layer 1 in SFI as shown in Table 9. To prevent malicious code injection threats, Lee (2023) addresses the issue of data quality and correlation from heterogeneous devices, aiding in incident prioritisation. Kumar et al. (2019) introduce a multi-Honeypot platform tool employing deep learning for malware classification. Edie et al. (2023) tackle APT threat playbooks dataset analysis, achieving high attribution accuracy. These approaches contribute to an improved understanding and identification of cyber threats, although they may come with computational resource requirements and considerations of false positives. Additionally, false data injection is addressed in (Gao et al., 2020b). The research presents a model based on heterogeneous information networks and graph convolutional networks (GCN) for advanced threat type identification. Zhang et al. (2022a) construct a knowledge graph for automated defence strategy generation, offering structured information on network security. Using the same method, Meier et al. (2018) achieve quicker time processing, for example, work-based linking took 1.789 seconds, and artifact-based linking took 3.724 seconds. Furthermore, they introduce FeedRank, a ranking approach for CTIFs, which assists organizations in feeding selection. Although these models offer substantial benefits, they require infrastructure and expertise. On the other hand, the cyber threats of node capturing are detailed in Rana et al. (2022). It employs honeypots, code analysis, obfuscation, and counterattack strategies to understand and potentially counteract threats. In contrast, Serketzis et al. (2019) enhance digital forensic readiness through IoC analysis and pattern identification, improving incident response capabilities. Czekster et al. (2022) centre on incorporating CTI into active buildings, emphasising encryption for security. Lastly, de Oca et al. (2022) build a global sensor network of honeypots and darknets to capture and analyse network traffic for real-time threat data. These approaches offer various means of threat analysis and defence, each with its unique requirements and advantages.

5.1.2. AI techniques

As shown in Table 10, four papers use supervised learning to address the node capturing threat in layer 1. Wang and Chow (2019) and Irshad and Siddiqui (2023) focus on different aspect—gathering threat intelligence from unstructured data and attribution extraction from reports. Tekin and Yilmaz (2021) employ deep learning for Twitter data, while Khoa et al. (2022) broaden the scope to IIoT networks supported by software-defined networking (SDN), using classification algorithms. In terms of results, Tekin and Yilmaz (2021) utilise deep learning for Twitter data, shows a promising accuracy of 88.61% in classifying cyber threat-related tweets. However, Khoa et al. (2022) focus on SDN-assisted IoT networks with the main contribution in the topic classification of three attack labels, namely *nss* (no shared secret), *zt* (zone transfer) and *qc* (query cache). The research showed a high level of accuracy from 94% to 100% by applying XGBoost. To tackle malicious code injection threats, Gao et al. (2021a) explore cyber threat hunting, Koloveas et al. (2019) combine topic modelling and regex-based filtering for content collection, and Koloveas et al. (2021) follow a three-step process for content ranking. All highlight the versatility of unsupervised approaches. Regarding the results, Gao et al. (2021a), focus on the search for cyber threats using unsupervised techniques, demonstrating its effectiveness in identifying malicious behaviours. Koloveas et al. (2019) propose a combination of topic modelling to exhibit robust content collection and Koloveas et al. (2021) focus on a content ranking approach which suggests a promising way to prioritise threat data. Three papers showcased NLP and text analysis techniques. Kim et al. (2019b) use SyntaxNet for cyberattack analysis, Martins and Medeiros (2022) apply rule-based classification for taxonomy tagging, and Gao et al. (2021b) leverage NLP for threat behaviour extraction. Kim et al. (2019b) employ SyntaxNet with conditional random fields, SVM classifier and LDA. The accuracy of the research reached up to 75% of F1 score with a dataset of 431518 posts in 101711 threats. Martins and Medeiros (2022) indicate rule-based classification is effective for taxonomy tagging classified by 8 different attributes (*URL, network address, network name, file hash, file name, email text, rule* and *agent*) from 1366 cyber-attack events, while Gao et al. (2021b) showcase the potential

Table 8
AI Techniques used in unstructured CTI source.

Paper ID	Domain	Focus	Technique category	Techniques tested	Data name	Datasize	Feature	Accuracy
Kadoguchi et al. (2019)	All	To collect the CTI from Darkweb	Deep learning	Webcrawler (Sixgill), MLP	Critical posts and non-critical posts.	3000 posts	Word2vec range value [-1,1] to classify critical posts and non-critical posts	F-score 79.4%
Kadoguchi et al. (2020)	All	To collect the CTI from Darkweb	NLP and Supervised learning	Doc2vec, K-Means algorithms for clustering and Deep cluster for self-supervised learning	1700 posts	Darkweb	Related and Unrelate malware.	N/A
Wang et al. (2022a)	All	CTI Feed assessment	Deep learning	ML - KNN classifier	21,448 CTI samples	Darkweb	The content assessment contains multi-source verification, content richness, timeliness with two attributes (link, authority)	0.923 accuracy
Adewopo et al. (2020a)	Health-care	Identify texts related to cyber threats	Supervised ML	Logistic regression, Random Forest classifier, Gradient Boosting	Twitter and Dark webs	500,000 tweets and over 128,000 posts from dark web forums.	9 Features: IP Addresses, Domain Names, Malware Signatures, URL Patterns, Hashes, Attack Patterns, Network Traffic Patterns, Patterns of Exploitation, Cryptography Usage	Random Forest Classifier achieved the highest F1-score of 0.81.
Adewopo et al. (2020b)	All	Twitter and Dark web	Supervised + Unsupervised learning	Logistics Regression, Random Forest Classifier, Gradient Boosting + Optimisation	Twitters	500,000 tweets over the period of 90 days and 128,000 posts from different discussion darkweb threads.	The thread titles are related to Carding, Newbie, Scam, Hacking, and Review threads.	Random forest classifier achieved a higher F1-score, is 0.81
Orbinato et al. (2022b)	All	Classify unstructured CTI report	Supervised learning	Naïve Bayes, Logistic Regression, SVM, MLP + Deep Neural network	CTI reports	12945 samples	188 MITRE ATT&CK techniques	SecureBERT produces the best accuracy up to 72%,
Sangher et al. (2023a)	All	Identify Cybercrimes through Dark Web Forum contents	Deep learning	Deep learning (RNN, CNN, LSTM and Transformer)	Agora dataset	109 activities by category.	Classify into three categories Cybercrime, Not Cybercrime and Can't say if cybercrime.	LSTM and BERT significantly outperformed and attained an accuracy of 96%.
Kim et al. (2022)	All	Classifying Tactics, Techniques and Procedures (TTPs) from unstructured CTI data.	Supervised learning	Logistic Regression, Naive Bayes, and MLP	TRAM	578 techniques related to 5,660 sentences.	These elements might include keywords, patterns, syntactic structures, and semantic information to differentiate between different TTPs.	F-score 90% to 95% on an average
Preuve-neers and Joosen (2021)	All	Avoid misclas-sification in CTI with ML application	Supervised learning	Decision Tree Classifier and Random Forest Model, Support Vector Machine models, Neural Network, Deep Learning-Based, Autoencoder	PCAP files from the CSE-CIC-IDS2018.	N/A	A decision is made on two fields: 'resp_bytes','orig_pkts'	Random Forest model produced F1-score 0.99999
Sangher et al. (2023b)	All	classifying various activities on the Dark Web as either cybercrimes, non-cybercrimes	NLP and Deep learning	CNN, RNN, LSTM, and BERT.	Agora DarkNet	N/A	Three labels: cybercrimes, non-cybercrimes, or uncertain cases.	The LSTM and BERT models achieve the highest accuracy of 96%
Hossen et al. (2021b)	All	To classify security-relevant posts from hacker forums	Unsupervised learning	Topic modelling (unsupervised learning) and Knowledge of Information Retrieval	Hacker forum	N/A	Six categories: Credential leaks, keylogger, DDoS attack, Remote access trojans, Crypters and SQL Injection	Logistic regression and decision tree provided highest level of accuracy BOW (93 -94%)
Sun et al. (2021a)	All	Enhance the classification attributes by considering the number of network interfaces involved in the attack.	Supervised learning	XG Boost, MLP, SVM, Decision Tree, Random Forest	OSTIPs	24,835 articles published from 2010 to 2019	CSI-candidate numbers, topic words, dictionary-word ratios, security target-word density, and article length.	an accuracy of 94.29%
Gautam et al. (2020a)	All	Analysing hacker forums	Unsupervised	Learning-based	Hacker fo-rums (CrackingArena), AZSecure-data.org	44927 threads	Classify as relevant or irrelevant.	99% accuracy
Chi et al. (2018)	All	Sentimental analysis from social media to assess the cyber threat	NLP	Sentimental analysis	Twitter	N/A	Tweets sentiments / Political scale category 1,2,3,4,5	Improving 4% of tweet sentiment classification

Table 9
Non-AI Techniques used in layer 1.

Paper ID	Domain	Focus	Techniques tested	Data sources	Data size	Feature	Accuracy
Lee (2023)	All	analysing cyber incidents collected from heterogeneous devices	Data analytics	SIEM, MISP, IntelMQ, CyberTriage and GRR	N/A	18 common attributes from Webserver, IPS/WAF and SIEM attribute.	N/A
Kumar et al. (2019)	All	Integrating multiple classes of Honeypots	Malware classification	Honeypot sensors	N/A	Signature based, and Pattern Knowledge base detection data	N/A
Edie et al. (2023)	All	APT threat playbooks dataset	Rule mining - Calculate code similarity + Activity attack graph	APT threat playbooks dataset	N/A	Similarity metric for attribution (jaccard)	accuracy 95.7%
Gao et al. (2020b)	All	identify threat types.	Meta-Path and Meta-Graph Instances-Based Computing	N/A	N/A	meta-graph based adjacent matrices are aggregated to obtain the weighted adjacent matrix B	N/A
Settanni et al. (2017)	All	Securing cyber physical systems	Artifact based linking, word-based linking and dictionary-based linking.	STIX. IODEF and JSON	N/A	N/A	With a data set of 1023, word-based linking takes 1.789 seconds
Meier et al. (2018)	All	Quality of CTI Feeds	A tamper-resistant ranking metric - correlation graph	Real Feeds	40 million entries	Assigns each feed a score and allows to rank them.	N/A
Yeboah-Ofori et al. (2019)	Supply chain	Cyber supply chain security	Cybersecurity controls and practices	N/A	N/A	N/A	N/A
Rana et al. (2022)	All	enhance counterintelligence and counterattack capabilities	Deception and Honeypots, Obfuscation, Payload Generation, Counterattack	N/A	N/A	Document-Based Tokens, Honeypots, Decoy Files, Data Analysis, Attack Vectors, Persistence, Malicious JavaScript	N/A
Serketzis et al. (2019)	All	improving the operational digital forensic readiness (DFR) of organizations	Data analytics	AlienVault Open threat exchange	1500 malware hash values	Relationships between entities of IoCs	86.85%
Zhang et al. (2022a)	All	generate defense strategies for network security.	CTI Knowledge Graph Construction, CTI Ontology Construction, CTI-KGE (Knowledge Graph Embedding)	Neo4j	224,430 entities, 9 relation types, and 408,885 triples.	Mean Reciprocal Rank (MRR) and Hit@n as evaluation metrics	N/A
Meier et al. (2018)	All	ranking CTI feeds	Graph modelling	N/A	40 million entries	The main feature indicator is the percentage of entries that one CTIF confirms from another CTIF.	N/A
Czekster et al. (2022)	Energy	Incorporating CTI into buildings with sensors and actuators	Encryption	N/A	N/A	N/A	N/A
de Oca et al. (2022)	All	Build a worldwide sensor network of honeypots and darknets.	VPS provider hosted nodes and nodes donated to the project by third-parties acting as endpoints.	N/A	N/A	Remote endpoint sensors, Frontend servers, External partner and third-party systems, Backend servers, External reporting system, Utility server.	N/A

Table 10
AI Techniques used in layer 1.

Paper ID	Domain	Focus	Technique category	Techniques tested	Data sources	Data size	Feature	Accuracy
(Wang and Chow, 2019)	All	Articles, reports, forums	Supervised learning	Java Annotation Pattern Engin	N/A, using web crawling	N/A	N/A	32.6% accuracy
(Tekin and Yilmaz, 2021)	All	Twitter	Deep learning	LSTM	21,000 tweets	N/A	"vulnerability" and "0day" to specific threat types such as "DDoS", "SQL injection", "buffer overflow".	88.61%
(Gao et al., 2021a)	All	Cyberthreat hunting	Unsupervised learning	Unsupervised ML	DAPRA TC dataset	N/A	18 attack cases & IOC types and IOC relations	100% precision, 96.74% recall, and 98.34% F1
(Kumar et al., 2021)	Maritime Transportation Systems	Automated DL-driven CTI modelling	Deep learning	DLTIF	Network sniffing tool (wireshark) gather raw packets at various choke points (e.g., mobile base stations) and can log them into a distributed database (i.e., MySQL cluster database)	N/A	Deep Feature Extractor (DFE) scheme's data is feed into the Bi-GRU based CTI Driven Detection (CTIDD) scheme	Obtained up to 99% accuracy
(Koloveas et al., 2019)	All	Collect zero-day vulnerabilities, exploits, indicators	Unsupervised learning	Topic modelling	Social media and dark web	N/A	Using word2vec with a latent space of 150 dimensions, a training window of 5 words, a minimum occurrence of 1 term instance, and 10 parallel threads. Use user tag for topic vocabulary for a set of N most related terms.	N/A
(Li et al., 2018)	All	Automated discovery and analysis of event-based CTI	ML	NLP, ML and data mining.	N/A	294 articles	N/A	Precision for location events is 76.9%, precision for device events is 92% and precision for organization events is 85.7%
(Khoa et al., 2022)	All	Software-Defined Networking (SDN)-assisted Industrial Internet of Things (IIoT) networks.	Supervised learning	XGBoost, Random Forest, K Neighbors	Dataiku	172.202 records	nss (no shared secret), zt (zone transfer), qc (query cache), and a normal label	94%-100% accuracy
(Kaiser et al., 2022)	All	automating incident responses	Supervised learning	knowledge graph	N/A	N/A	attack techniques, observables, defensive techniques, and relationships between them	N/A
(Pour et al., 2021)	All	an actionable CTI threats	Supervised learning	RF	information of compromised devices with a two-week period..	N/A	IP header, TCP header and TCP options	94.63%
(Irshad and Siddiqui, 2023)	All	CTI attribution extraction from CTI report	Supervised learning	Decision tree, Random Forest, Support Vector machine	CTI reports+CVE+Malware Sample reports	N/A	7 features (Cyber threat actor, TTP, Malware, Tools, target Organisation, Target Country, Target Application)	Accuracy81-96%
(Tundis et al., 2020)	All	evaluating the relevance and quality of various OSINT sources	Supervised learning	Regression analysis	Twitter	1.2 million tweets spanning a three-year period.	registration date, location, followers, connections, retweets, content analysis	0.975
(Kim et al., 2019b)	All	Extract information from cyberattack analysis reports.	NLP	NLP SyntaxNet incorporates the CRF (Conditional Random Fields) algorithm	N/A	190 reports	IP, URL, Hash, Email, CVE, and time objects	F1-score of 76%

Table 10 (*continued*)

Paper ID	Domain	Focus	Technique category	Techniques tested	Data sources	Data size	Feature	Accuracy
(Samtani et al., 2017)	All	analysing malicious hacker assets found in various hacker forums	Unsupervised and Supervised learning	data collection, data preprocessing, SVM classifier, LDA	N/A	431,518 posts in 101,711 threads	source code, attachment, and tutorial topics	98.20%
(Al-Fawa'reh et al., 2022a)	All	Improving intrusion detection system (IDS)	Supervised learning	DNN + PCA	CSE-CICIDS2018 dataset - AWS dataset	CIC IDS 2018 dataset	protocol number, IP address, unique Flow ID, IP source/destination, timestamp, and tag.	98%
(Koloveas et al., 2021)	All	Collecting CTI source	Unsupervised learning	LDA	CVE	N/A	Number of words, Security Action-word density	N/A
(Martins and Medeiros, 2022)	All	Creating Open source Threat intelligence and a unified TI taxonomy	NLP	Rule-based classification	1,366 events	N/A	8 attributes (URL, Network address, Network name, File hash, File name, Email text, Rule, agent)	0.98%
(Riesco and Villagrá, 2019)	Business risk management	Dynamic risk assessment	NLP	Ontology	honeypots and dark/deep Web	N/A	SWRL rules such as asset valuation, threat identification, risk assessment, risk severity classification, security event detection, and risk mitigation strategy selection	N/A
(Gao et al., 2021b)	All	proactive cyber threat hunting within computer systems	NLP	NLP	N/A	N/A	file events, process events, and network events	N/A
(Marques et al., 2022)	All	Identify cybercrime from darkweb	NLP and Deep learning	NLP, CNN, RNN, LSTM, and BERT	Agora DarkNet	N/A	Tcybercrimes, non-cybercrimes, or uncertain cases.	The LSTM and BERT models achieve the highest accuracy of 96%

of NLP for threat behaviour extraction. The effectiveness of techniques varies depend on the specific objectives and datasets.

5.2. CTI techniques used in layer 2

5.2.1. Non-AI techniques

Table 11 summarises the 15 papers which use non-AI techniques in layer 2. In addressing phishing threats in layer 2, Merah and Kenaza (2021b) propose an ontology-based approach with security information event management that integrates CTI with Structured Threat Information eXpression (STIX) for cyber risk monitoring. While the paper does not explicitly provide quantitative results, the integration of CTI with STIX enhances the comprehensiveness of threat intelligence data. Moving on to Darknet Threat Intelligence, Arnold et al. (2019) adopt a unique approach by utilising elastic search with Kibana analytics to identify cyber threats in major darknet data sources to define the motive of the phishing attack. Landauer et al. (2019) focus on the finance sector and found nine large clusters of related phishing threats such as *distribution sites* and *emails*. In contrast, Miles et al. (2014) analysed the interrelationships between malware instances and utilised automated processes. While no accuracy metrics are disclosed, the paper mentions the use of a dataset provided by a major financial institution, enhancing its credibility. On the other hand, to identity authentication issues, Moraliyage et al. (2022) classified CTI based on text and image content. Allegretta et al. (2023a) leveraged the STIX dataset and applied rule-based analysis in 3M cyber incidents. Focusing on attack graphs and CTI integration, Gylling et al. (2021) investigated attack behaviour using a multimodal architecture approach based on tactics, techniques, procedures, indicators of compromises, targeted vulnerabilities, and suspected threat actor groups in high accuracy of 95%.

5.2.2. AI techniques

As shown in Table 12, to tackle one of the cyber threats in layer 2 such as phishing, 22 of the 39 papers in this section used Deep learning techniques such as LSTM RNN (Grisham et al., 2017; Graf and King, 2018; Suryotrisongko et al., 2022b), BERT (Jo et al., 2022; Zhang et al., 2021b; Liu et al., 2022; Kristiansen et al., 2020; Fujii et al., 2022), CNN (Wang et al., 2022a; Ampel et al., 2020; Graf and King, 2018; Song et al., 2022b; Wang et al., 2022b; Sanjeev et al., 2020; Sarhan and Spruit, 2021; Zhao et al., 2020a,b), transfer learning (Ampel et al., 2020), artificial neural networks (Alsaedi et al., 2022), deep neural networks (Al-Fawa'reh et al., 2022b). These deep learning techniques offer a high level of accuracy of more than 98% to minimise misclassification (Zhang et al., 2021b) based on *titles, URLs* and *snippets* in classifying 6000 APT domains from 400 APT attack reports. Sun et al. (2021b) used a graph convolutional network considering 6 types of Indicators of Compromise (IoC) and 9 types of relationships which help to tackle the issue of heterogeneous IoC effectiveness with an accuracy of 98.59% based on 5 variables, namely *attacker, vulnerability, file type, platform* and *device*. Furthermore, the RNN model achieves an accuracy of 99.025% (Wagner et al., 2018b) in detecting Advanced Persistent Threat (APT) attacks with a training model of 7114 threads labelled as relevant out of 44927 threads.

5.3. CTI techniques used in layer 3

5.3.1. Non-AI techniques

As shown in Table 13, Dulaunoy et al. (2022) focus on developing a system for storing historical forensic artifacts collected from SSH connections to address one of cyber threats in this layer, namely insider attacks. The primary technique used is SSH protocol fingerprinting,

Table 11
Non-AI Techniques used in layer 2.

Paper ID	Domain	Focus	Techniques tested	Data name	Data size	Feature	Accuracy
Merah and Kenaza (2021b)	All	Ontology for CTI risk monitoring	Ontology	CTI- XML and JSON reports.	N/A	operational, tactical, and strategic	N/A
Almohannadi et al. (2018)	All	Define adversary's motive	Elastic search with Kibana analytics	Honeypot	log data 500 MB for more than a year through AWS cloud called Kippo and Dionea	Find common attack event that attacker uses	N/A
Landauer et al. (2019)	All	Identify patterns for intrusion detection	Data mining.	N/A	N/A	N/A	Only 16 out of 1000 anomalies were undetected after using the proposed approach
Miles et al. (2014)	Finance	Interrelationships among instances of malware	N/A	N/A	463 malicious	Type of malware artifacts including the binary, code, code semantics, dynamic behaviours, malware metadata, distribution sites and emails.	Found nine large clusters of related malware
Bou-Harb (2016)	All	Filter out misconfiguration traffic from darknet data	Probabilistic distribution, and joint probability computation, normalization	Darknet	One-hour period of CAIDA's darknet dataset for one experiment	Two core metrics ("rareness of access" and "scope of access")	N/A
Atifi and Bou-Harb (2017)	All	Network traffic image	Bloom filters	N/A	10 GB of real darknet data and close to 15 thousand malware traffic samples.	Correlation of network traffic and the generation of actionable CTI.	N/A
Gylling et al. (2021)	All	Attack behaviour integrating with Attack (Defense) Graphs (ADGs)	N/A	N/A	N/A	Tactics, techniques, procedures, indicators of compromise, targeted vulnerabilities, and suspected threat actor groups.	95%
Moraliyage et al. (2022)	All	classify onion services based on the image and text content	Multimodal architecture	Computer Incident Response Center Luxembourg (CIRCL)	N/A	Text and image	N/A
Allegretta et al. (2023a)	CTI	Identify trends in CTI	Rule-based analysis	private STIX dataset	3million cyber incidents	grach of cyber incident components called STIX domain objects (SDO).	N/A
Leite et al. (2022)	All	Improving network intrusion detection and incident response	N/A	N/A	78.5 GB of network traffic data (PCAPs).	file hashes, exploit downloader files, IP addresses	96.22% based on the tested ransomware samples.
Jiang et al. (2023)	All	Sharing threat detection models	Blockchain and Federated Learning	ISCX-IDS-2012 and CIC-DDoS-2019	N/A	FlowID, Source IP, Source Port, Destination IP, Destination Port, and Timestamp	N/A
Zhang et al. (2021a)	All	Accurately extract and automatically identify threat actions in unstructured CTI reports.	Rule-based classification	243 CTI reports	N/A	TF − IDF, frequency, dependence, distance	N/A
Ammi et al. (2022)	All	A cloud-native architecture capable of connecting security-related data	Established semantic technologies for cloud-native security solutions in CTI	N/A	N/A	Event object, EventType, Item, ItemCategory, ItemType, and Relations	N/A
Yoo and Lee (2023)	All	Identify and filter the ordinal scale risk of the source IP in deceptive environment-generated traffic	Naive Bayes discriminant analysis-based ordinary scale classification model	Own dataset (Korea Internet & Security Agency)	N/A	IP	N/A
Shin et al. (2019)	All	Classifying and analysing pivot attacks, a type of cyber attack	Automatic Pivot Classifier Algorithm (APCA)	N/A	N/A	source and destination IPs, ports, and other attributes	N/A

Table 12
AI techniques used in layer 2.

Paper ID	Domain	Focus	Technique category	Techniques tested	Data sources	Data size	Feature	Accuracy
Truvé (2016)	All	Analyse the current state of world affairs or predict future attack events	Supervised learning/ Classification	Predictive models (SVM, risk score calculation)	Open, deep, and dark web	N/A	7,528 samples from 2010-01-01 to 2014-12-31)	accuracy of 0.83 (precision = 0.82, recall = 0.84
Zhang et al. (2022b)	All	Automated breaking of dark web CAPTCHA to facilitate dark web data collection	Vision + Image recognition	Generative Adversarial Network (GAN)	N/A	N/A	N/A	94.4% accuracy
Liao et al. (2016)	All	Automatic extraction/gathering of OpenIOC compatible data from sources like articles, blogs, forums, reports, etc. for cyber threat inteligence	Supervised - logistic regression	Graph mining technique - KLR classifier	N/A	N/A	N/A	95% accuracy and 90% coverage
Grisham et al. (2017)	All	Mobile malware from zipped Androi apps attachment in hacker forums	Deep learning + AI neural network	LSTM RNN+Social network analysis	Forums - Ashiyane, Hackhound, VBSpiders, Zloy	N/A	Key threat actors	Precision 95%, Recall 81% and Fmeasure 87%
Wheelus et al. (2016)	All	Designed a multi-layered Big Data architecture to automate the generation of cyber threat artifacts for adaptive CTI	Supervised + Classification + AI neural network + NLP	Designed a Multi-Layered Big Data Architecrure to automate the generation of cyber threat artifacts to effectively feed to ML techniques for adaptive CTI.	SANTA Dataset	N/A	SANTA Dataset	N/A
Sury-otrisongko et al. (2022a)	All	Topic modelling in CTI for OSINT - https://pypi.org/project/maryam/	Deep learning	BERTopic and Top2Vec	Nulled.io hacker forum database	N/A	Using a not-cybersecurity -word list to be able to filter unrelated words.	N/A
Deliu et al. (2017)	All	Extract CTI from Hacker forums	Deep learning	SVM and CNN	Nulled.io	16000 posts (relevant and irrelevant)	w2vInternal-CNN D = 300	SVM (trigrams) produces a highest result accuracy of 0.83
Deliu et al. (2018)	All	Extract CTI from Hacker forums	Supervised learning	SVM and LDA	Nulled.io	16,000 posts	Classify with 5 topics with relevant, timely and actionable CTI.	N/A
Williams et al. (2018)	All	An incremental crawling approach designed to gather hacker forum attachments on an ongoing basis	Deep learning	LSTM RNN	Hacker forums: OpenSC, Garage4hackers, Hacksden, AntiOnline, Crackingzilla, WebCracking, SafeSkyHacks, Ashiyane, Hack, and Haker	N/A	exploit name, author activity, forum, sub-forum, thread, and URL	N/A
Sury-otrisongko et al. (2022b)	All	OSINT and XAI to detect based DGA-based traffic (malicious DNS traffic)	Supervised learning	XAI - Logistic regression, random forest, Naïve bayes, extra tree and ensemble	Alexa and Bonet	Alexa Top 1M (1,000,000 domain names) and 803,333 domain names of ten botnet DGA families	7 features: Charlength, TreeNewFeature, ReputationAlexa, RE-Alexa, Min-RE-Botnets, Entropy and IRad	The highest accuracy (96.2%)
Ampel et al. (2020)	All	Multi-class Text classification for hacker exploits	Deep learning	transfer learning - Feature-representation - transfer - BiLSTM layer - Convolutional layer	Traditional hacker forums	18 sources, English and Russian includes 8592134 posts and 264574 source codes.	8 exploit labels such as Web applications, DoS, Remote, Local, SQL injection, XSS, File inclusion, Overflow.	DTL-EL leads to statistically significant performance increases in accuracy at a 3.22% increase

(*continued on next page*)

Table 12 (*continued*)

Paper ID	Do-main	Focus	Technique category	Techniques tested	Data sources	Data size	Feature	Accuracy
Graf and King (2018)	All	Neural network and Blockchain and situational awareness - Proactive CTI	Deep learning	Neural network - Deep autoencoder - Smart contract	OSINT sources	The dataset contained 5,850 training documents and 584 test documents.	'number of related incidents', 'number of related words', 'number of original words', 'detected significant terms' and 'vulnerability score'.	0.942
Mavroeidis et al. (2021)	All	Non-uniform, unstructured and ambiguous high-level information.	Deep learning	Ontology	STIX, MITRE	N/A	N/A	N/A
Zhang et al. (2021b)	All	Mining open-source CTI to minimise misclassification	Deep learning	article proposes two variants of the networks: CNN+mi-NET, CNN+MI-NET, BiLSTM+mi-NET, and BiLSTM+MI-NET.	N/A	6000 APT domains from 400 APT attack reports	titles, URLs, and snippets	95.39% to 98.14%, with the CNN+mi-NET model achieving the highest accuracy.
Wang et al. (2022c)	All	CTI entity recognition model	Deep learning	KE-BERT-BiLSTM-CRF based on knowledge engineering	CyTiner Dataset	N/A	group, time, user, methods	N/A
Song et al. (2022b)	All	diachronic graph embedding	Deep learning	hyperbolic graph neural networks and hyperbolic gated recurrent neural networks.	N/A	32,766 posts made by 8,429 hackers between January 1, 1996 and July 10, 2019 (23-year period)	future threat type (local or remote attack) and platform (attack from Linux or windows).	In terms of F1 score of 82.6%.
Bose et al. (2021)	All	Twitter user account	NLP	Web crawler Tweepy + IBM's Watson Natural + rule based classification Language Understanding (NLU) service	Twitter	50,000 Twitter user accounts	three categories "antivirus and malware", "Technology and Computing", and "computer science"	55% - 67%
Zuo et al. (2022)	All	extracting entities from cyber-security texts.	Deep learning	BERT, Bidirectional LSTM (BiLSTM), Conditional Random Fields (CRF),	Symantec, Fireeye, and Threatpost	1087 cyber-security texts	BERT, LSTM-based sequence	F1 score of 0.758
Wang et al. (2022b)	All	NER tasks related to APTs	Deep learning	BiLSTM - CRF, CNN - BiLSTM - CRF, LM - LSTM - CRF	APTNER	10,984 sentences, 260,134 tokens and 39,565 entities	21 predefined entity categories such as IP, URL, malware, and location.	N/A
Panagiotou et al. (2021)	All	Extract information from blog and websites	Supervised learning	SVM, RF	N/A	920 web pages	CVE applied TF-IDF	SVM performs better than RF.
Kristiansen et al. (2020)	All	Collect, process, analyse, and generate threat-specific knowledge from tweets shared by multiple users on Twitter.	Both ML and supervised learning	BERT, CNN, K-means clustering LDA	Twitter	76,047 tweets	"Covidlock" ransomware	BERT with 88% accuracy
Dhake et al. (2023)	All	Using internet hacker forum as a source of gaining CTI and developing proactive type of CTI	Supervised learning	ML	Privately created upon CrackingArena	Number of threads of 44,927 is used, while there are 5,047 relevant threads	Common cybersecurity keywords	N/A
Sanjeev et al. (2020)	All	Automated for CTI generation is presented that can act as attack indicator for the security defence mechanism such as SIEM	Deep learning	deep learning neural network-based CTI generation for cyber threat prediction	open sources intelligence (OSINT) database	N/A	Syslog, firewall, IDS/IPS,	N/A

Table 12 (*continued*)

Paper ID	Do-main	Focus	Technique category	Techniques tested	Data sources	Data size	Feature	Accuracy
Yu et al. (2022)	All	Tactics And Techniques Classification	NLP	HM-ACNN	N/A	N/A	natural text	N/A
Guarascio et al. (2022)	All	Sharing threat events and Indicators of Compromise (IoCs) to improve decision-making and countermeasures against cyberattacks, especially in the context of Intrusion Detection Systems (IDS)	Supervised learning	Active learning	CICIDS2017 dataset	N/A	Honeypot-based data enrichment.	91-97%
Sarhan and Spruit (2021)	All	unstructured Advanced Persistent Threat (APT) reports	Deep learning	CNN	MalwareDB dataset	39 APT reports contained 6,819 sentences,	, we were only able to classify 1,910 sentences	achieving a higher F-measure by 4.2%.
Li et al. (2022)	All	Attack behaviour graphs	NLP	AttacKG	N/A	1515 reports	identifying attack techniques, extracting dependencies among entities, recognizing domain-specific terms (IoCs)	F1-scores of 0.887, 0.896
Alsaedi et al. (2022)	All	detecting malicious URLs	Supervised learning and Deep learning, ensemble learning	Random Forest algorithm and an Artificial Neural Network (ANN) classifier.	Kaggle and Phishtank	20,000 URLs was drawn and used in this study	feature extraction (N-gram technique), feature representation (TF-IDF), feature selection (using information gain), ensemble learning-based prediction (Random Forest), and decision making (Artificial Neural Network)	96.8%
Jo et al. (2022)	All	Fast updating ransomware attacks from CTI reports	Unsupervised learning	BERT, NER and RE	ThreatPost, and Malwarebytes	540 K articles (roughly 11 M sentences)	ransomware names, attack vectors, vulnerabilities, platforms, algorithms, and tools	F -score of 0.972
Zhao et al. (2020a)	All	Analys and classify cyber-attacks	Deep learning	CNN	TI-Spider	118,000 threat-related descriptions over the past 16 years from January 2002 to November 2018.	IP addresses, domain names, URLs, hashes.	84% and 94%
Liu et al. (2022)	All	Trigger-Enhanced Actionable CTI Discovery System	NLP + Deep learning	NLP + BERT	N/A	29,000 cybersecurity reports	“campaign triggers,”	86.99% and an F1 score of 87.02%.
Li et al. (2023)	All	CTI automatic analysis tasks	Deep learning	BERT	APT notes	634 APT reports	Converting tree-shaped input structures into linear structures	0.941
Sun et al. (2021b)	All	Automatically gathering CTI records	Supervised learning	ML and rule-based techniques	Neo4j	24,835 articles published between 2010 and 2019	CSI-candidate number, topic word, article length, dictionary-words density, security action-word density, securityTarget-word density.	N/A

(*continued on next page*)

Table 12 (*continued*)

Paper ID	Domain	Focus	Technique category	Techniques tested	Data sources	Data size	Feature	Accuracy
Fujii et al. (2022)	All	automates the conversion of unstructured CTI into structured STIX format	Deep learning	NER, BERT	N/A	34 sites	N/A	RoBERTa-large achieved the highest accuracy among the models with an F-measure of 0.8012
Zhao et al. (2020b)	All	Heterogeneous IOCs to quantify the interdenpendent relationship among IOCs	Deep learning	Graph Convolutional Network	73 international security blogs	30,000 training samples from 5,000 threat description texts.	Attacker, vulnerability, file, type, platform and device in the network as nodes. It considers 6 types of IOCs and 9 types of relationship.	Multi-granular results a highest accuracy 98.59%
Al-Fawa'reh et al. (2022b)	All	Intrusion detection system (IDS)	Deep learning	Deep Neural Network (DNN) and PCA	CSE-CICIDS2018 Dataset	19141630 normal flow, 714290 attacks	81 features in which flow-level features (F2, F4, F9, F10, F11, F12, F13, F14, F15, F17, F18 and F19,)	Classify malicious and benign flowes _ 98% accuracy
Sakthivelu and Vinoth Kumar (2022)	All	detection of Advanced Persistent Threat (APT) attacks	Supervised learning and Deep learning	Bayesian algorithm, the C5.0 decision tree algorithm	NSL-KDD dataset	N/A	network traffic, behaviour patterns, and communication to detect anomalies	Deep learning outperforms 99.55% respectively.
Gautam et al. (2020b)	All	cybercrime in hacker forums	Deep learning	RNNs	CrackingArena	44,927 threads, with 7,114 threads labelled as relevant.	thread date, author name, post data, and thread title include terms related to cybersecurity, such as antivirus, backdoor, botnet, malware	RNN GRU model achieved an accuracy of 99.025%
Al-Fawa'reh et al. (2022c)	All	Abnormal Network Behaviour Detection	Supervised learning	Learning-based	CSE-CICIDS 2018	125,973 network traffic samples in the KDDTrain	40 features in CSE-CICIDS2018dataset mapped into Normal, DoS, Remote to Local (R2L), Probe, and User to Root (U2R) categories	95.6% - 99.67%

leveraging a passive SSH scanner and a Redis database for storing key materials. Husari et al. (2018) investigate text mining from CTI reports. The technique employed is known as ActionMiner, which involves calculating entropy and mutual information to ensure meaningful threat action extraction from Verb-Object combinations. Molloy et al. (2022) present the JARV1S system. JARV1S utilises a phenotype-based approach, combining static analysis, binary clone search, and information retrieval to handle the proliferation of malware variants. Results indicate high matching ratios for identifying malware families, with only a few false negatives. The system is shown to be scalable, efficient, and effective against zero-day malware. Gong and Lee (2021b) address the prevention of cyberattacks in energy infrastructure cloud environments. The technique used is a rule-based threat detection mechanism. Nagasawa et al. (2021) assign appropriate labels to security blog posts using the Topics Over Time (TOT) model.

5.3.2. AI techniques

To prevent threats such as insider attacks, Chen et al. (2023b) focus on CTI extraction using deep learning techniques, particularly BERT as shown in Table 14. They introduce the CARE system for extracting critical threat entities and their relationships from cybersecurity articles. However, specific results or features are not mentioned. Samtani et al. (2016) collect and analyse malicious assets from hacker communities using supervised classification and NLP to process 2777 source code and 1709 attachments in 10 topics. The paper shows that SVM achieves the highest accuracy of 96.67%. Ge and Wang (2022) emphasise its contribution to image processing in combination with deep learning (RNN, CNN) and supervised learning (SVM), however, they do not specify the level of accuracy of the method. As layer 4 is the closest layer to the user interface, (Merah and Kenaza, 2021b) emphasises the use of Explainable AI in interpreting threat intelligence from event logs. Kattamuri et al. (2023) improve static malware detection using ML and optimization algorithms like Ant Colony Optimization (ACO), Cuckoo Search Optimization (CSO), and Grey Wolf Optimization (GWO) from 51049 samples including a total of 108 pure PE file header attributes. This research indicates the 12 most relevant features from the PE file header attributes with an accuracy of 99.37% with the assistance of NLP. Robertson et al. (2017) develop a threat model for extracting the structure and behaviour of online hacker communities. It employs various approaches, including supervised, semi-supervised, and unsupervised learning to propose 9 attributes *product fields, item reviews, topic content, post content, topic author, post author, author status, reputation* and *topic interest*. Of the techniques applied, the validation reveals SVM achieves the highest accuracy of 87%. The dataset includes hacker forum data, and the paper discusses the collection of cyber threat warnings. Gong and Lee (2021c) propose a CTI framework to enhance the security of energy cloud environments. They adopt deep learning (GNN) techniques with an F1 score of 82.2% from 20480 IoC data in *energy consumption and production, communication protocols, user iden-*

Table 13
Non-AI technique in layer 3.

Paper ID	Domain	Focus	Techniques tested	Data sources	Data size	Feature	Accuracy
Dulaunoy et al. (2022)	All	Develop a system to store historical forensic artifacts	Finger printing SSH protocol	Passive SSH	N/A	A simple SSH scanner including hash or host.	N/A
Husari et al. (2018)	All	Text mining from CTI reports	Text mining (ActionMiner)	Wikipedia	2200 malware reports.	Verb-Object combination is meaningful threat action.	N/A
Molloy et al. (2022)	All	malware analysis	Static analysis, binary clone search, information retrieval	Malware Repository, Begin Repository and Zero-day set	200,000 malware samples along with 100,000 benign samples	Extract phenotypes, that is, observable characteristics from the assembly code, to match functional level clones	Matched 100% ratio
Gong and Lee (2021b)	Energy grid	Preventing energy infrastructure cloud environments from cyberattacks.	Rule-based threat detection mechanism	N/A	20,480 attacks	N/A	F1 score of 0.822
Nagasawa et al. (2021)	All	Assign appropriate labels to security blog posts.	Topics Over Time (TOT) model.	N/A	N/A	N/A	N/A

Table 14
AI Techniques used in layer 3.

Paper ID	Domain	Focus	Technique category	Techniques tested	Data sources	Data size	Feature	Accuracy
Chen et al. (2023b)	All	Situational awareness	Deep learning	N/A	N/A	N/A	Extracts critical threat entities and presents their relationship in both graphical and textual forms	N/A
Samtani et al. (2016)	All	Collect and analyze malicious assets	Supervised learning, Classification, NLP	Webcrawler, SVM, Topic modelling (LDA)	AZSecure Hacker	2777 source code and 1709 attachments.	10 topics with provided keywords, exploit type.	SVM provides a high level of accuracy 96.67%
Afzaliseresht et al. (2020)	All	Mining threat ingelligence data from event logs.	Explainable AI.	NLP	TCP/IP data related to security threats.	N/A	N/A	Story telling report included in the story telling, log files
Ge and Wang (2022)	All	Assessing system risks.	Image processing, Deep learning and Supervised learning	RNN, CNN, SVM	TT&CK V8 tactics and techniques, TTPDrill, rcATT and Drebin	6500 examples of tactics and techniques of ATT&CK	Enable computers to read, understand, and generalise the meaning of texts	SeqMask achieved F1 scores of 86.07% and 73.99% for TTPs classifications.
Kattamuri et al. (2023)	All	Improving static malware detection of Windows executable files	supervised learning	ML, ACO, CSO, GWO	SOMLAP.	51,409 samples including a total of 108 pure PE file header attributes	Select the 12 most relevant features from the PE file header attributes	an accuracy of 99.37% in malware detection
Robertson et al. (2017)	All	Detect hackers' behaviour for a pro-CTI	Supervised, Semi-supervised learning and Unsupervised learning	Web-scrawler, ML + Data ana-lytics, Game theory	Dark and Deep webs	N/A	Product fields, item reviews, topic content, post content, topic author, post author, author status, reputation, topic interest	SVM achieved 87%
Gong and Lee (2021c)	Energy	protect energy cloud systems from cyberattacks	Deep learning	GCN	N/A	20,480IoC data.	energy consumption and production, communication protocols, user identifiers, port information, process information, process resource usage, and energy object statistics.	A macro-F1 score of 0.822

Table 15
Non-AI Techniques used in layer 4.

Paper ID	Domain	Focus	Techniques tested	Data sources	Data size	Feature	Accuracy
Arıkan and Acar (2021)	All	Network security data	Data mining	KDD CUP 99, NSL-KDD, CART	N/A	N/A	N/A
Dietz et al. (2022)	All	Simulating attack scenarios on industrial control systems (ICS).	Digital twin security simulation	Honeypot	N/A	463 relationships	N/A.
Chakir et al. (2023)	All	Evaluation of open-source Web Application Firewalls (WAFs)	Signature-based and hybrid-based detection approaches	Payload All The Things	N/A	Three main types of attacks: SQLI, XSS, and XXE	AQTRONIX Webknight v4.4 with a high recall value of 98.5%

tifiers, port information, process information, process resource usage and *energy object statistics*.

5.4. CTI techniques used in layer 4

5.4.1. Non-AI techniques

As shown in Table 15, the CTI techniques used in this paper can address data breaches and service interruptions in layer 4 (the application layer) as explained in section 2.2. Focusing on various aspects of application data, Arıkan and Acar (2021) employ data mining techniques and utilise datasets like KDD CUP 99, NSL-KDD, and CART for its analysis. Centred on digital twin security simulations for generating structured CTI, (Dietz et al., 2022) utilises the MiniCPS tool for ICS simulation and applies digital twin technology. The paper emphasises the integration of digital twin security simulations into CTI generation, highlighting the potential for improving threat information sharing. However, the paper does not specify key results or accuracy metrics but outlines a framework for this integration. To prevent application vulnerabilities and software bugs, it is essential to evaluate open-source Web Application Firewalls (WAFs) (Chakir et al., 2023). Chakir et al. (2023) assess the effectiveness of two open-source WAFs, AQTRONIX Webknight v4.4 and ModSecurity v3.0.4, in detecting various web application attacks. The paper employs signature-based and hybrid-based detection approaches and evaluates these WAFs based on their performance metrics, including recall, precision, F-value, and false positive rate. AQTRONIX Webknight v4.4 demonstrates strong performance in detecting SQL injection, XSS, and XXE attacks, achieving a high recall value of 98.5%. However, it also results in a high false positive rate of 99.6%. On the other hand, ModSecurity v3.0.4's effectiveness varies with the level of paranoia configured, with PL4 showing a high false positive rate of 60.3%. The paper concludes that while these WAFs have strengths, they are not completely suitable for web application security due to their reliance on signature-based approaches. It suggests exploring new WAF approaches based on ML and deep learning for improved detection and lower false positive rates. Chakir et al. (2023) stands out for its comprehensive evaluation of open-source WAFs, providing specific performance metrics and highlighting the need for future advancements in web application security.

5.4.2. AI techniques

In the application layer, Samtani et al. (2021) highlight that the existing CTI with the analysis of event log files leads to a lack of a proactive CTI system of potential threats before an attack occurs, as shown in Table 16. The paper recommends using the Dark Web as an unstructured CTI data source to collect and analyse the hackers' motivation and their criminal assets. This approach improves situational awareness in a proactive CTI system. The authors used SVM, NLP and deep learning (LDA). The paper contributed a huge database consisting of 10,975,390 records from 90 platforms to identify and classify two features that are potential threats along with key hackers and their criminal assets from multi-lingual sources. However, the level of accuracy was not presented clearly in the paper.

The analysis in Sections 5.1 - 5.4 shows that different CTI techniques can be used to address a particular SFI layer. Nevertheless, non-AI techniques' performance is not specified clearly in most of the selected papers but its processing time can be indicated in terms of whether it is quicker or not. On the other hand, AI techniques can be measured by the level of accuracy, for example. Deep learning has been applied widely across different layers and it offers a high level of accuracy which can reach up to 100% depending on its selected features.

6. RQ3 - how does vCISO utilise the intelligence from current CTI techniques to estimate the potential monetary impact of an exploited vulnerability on a farm?

As shown in the analysis in Sections 4 and 5, none of the articles in the SLR discuss the application of CTI to proactively prevent cyber attacks in the agricultural context, including crop production, livestock, fishery, forestry and their supply chains. It is important to have a specifically designed model for agriculture to meet the increasing needs. However, based on the literature review, we can use the existing techniques and CTI database to develop a comprehensive platform that helps vCISO and farmers monitor and be aware of potential threats to their smart farming system. We provide the results in Sections 6.1 and 6.2. Section 6.1 proposes a taxonomy of CTI sources that can be used in an agricultural context. Section 6.2 presents a taxonomy of the most commonly used techniques in the literature which can be used in our future work. In this section, we demonstrate a scenario of an agricultural ground vehicle attack surface. As discussed in the introduction section, in 2022, an Australian security researcher, Sick Codes highlighted the need for the agricultural sector to take cybersecurity more seriously to prevent potential disruptions to the food supply chain by demonstrating his ability to hack a John Deere tractor display and install a vintage 1990s video game to show his control of the system proofread (ABC Rural et al., 2022). This indicates the need for a CTI platform in existing smart farming systems.

6.1. Taxonomy of CTI sources aligned with the agriculture context

Based on our SLR and under the context of agriculture, we propose an updated appropriate accessible and available CTI source to establish and improve proactive CTI in agriculture. As shown in Fig. 2 which illustrates a cyber DDoS attack, it can gain access control to a vehicle such as a tractor, stealing data through network communication related to third-parties such as data loggers, dealerships, and supporting services parties (agro firms, precision farming third-parties). As a result, it will cause disruptions to the whole farming supply chain. As an application of our systematic literature review, we can list an available CTI source that will help cyber expertise update with the fast-evolving cyber threats to layer 2 (network communication). As shown in Table 17, the accessible CTI source is assessed as to whether it can help in cyber threat detection and with which features. There are 8 CTI sources with available links, as shown in Tables 18 and 19, however, none of the

Table 16
AI Techniques used in layer 4.

Paper ID	Domain	Focus	Technique category	Techniques tested	Data sources	Data size	Feature	Accuracy
(Samtani et al., 2021)	All	Dark web situational awareness by collecting, analysing HAP and reports on major Dark web data sources + Strategic intelligence	Supervised learning, Classification, NLP	Webcrawler + Text mining (Latent Dirichlet allocation) + SVM	HAP data collected from forums, IRC, DNMS, network logs, Source code in forums, Shops. The authors propose techniques but do not give the accuracy.	10975390 records from 90 platforms	Threat detection, key hacker identification,multi-lingual analysis, global hacker surveillance and cybersecurity visualisation	N/A

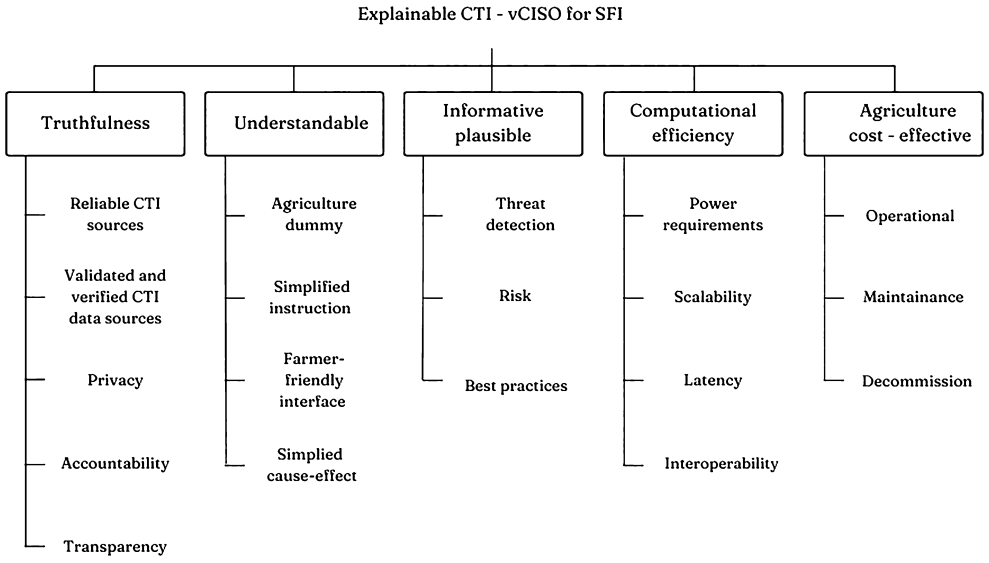

Fig. 4. A taxonomy of a farmer-friendly CTI for vCSIO and non-technical stakeholders in Agriculture.

sources offers a user-friendly interface for non-technical stakeholders to understand. This highlights a need to have an explainable platform as a virtual CSIO to assist non-technical stakeholders in coping with cybersecurity and enhance CTI to be accessible to anyone in the agriculture chain.

6.2. Taxonomy of CTI techniques aligned with the IoT layer in the agriculture context

Based on our SLR and in the context of agriculture, we proposed updated, appropriate, accessible and available CTI techniques to establish and improve proactive CTI in IoT agriculture systems. Fig. 4 shows the taxonomy of an explainable CTI under five main terms which are truthfulness, understandable, informative and plausible, computational efficiency and agriculture cost-effective. Under each term, there are different features and a detailed definition of each feature is given below its structure shown in Table 20. The definition includes Description, Techniques and Drawbacks. This enables CTI to be more applicable and explainable in the agricultural context where farmers and their stakeholders are not yet in the mature stage of technology adoption and are unfamiliar with specific advanced technological terms in general and also cybersecurity specifically. Regarding the survey from farmers and stakeholders and the standard of an explainable application in digital agriculture (Dara et al., 2022), we propose the taxonomy of a farmer-friendly CTI for vCSIO and non-technical stakeholders in the era of Agriculture 4.0. There is an increasing use of supervised learning, deep learning, and NLP in explainable AI. As can be seen in Equation (3) (Wang et al., 2014), the algorithm of logistic regression as an example, the strength of regression is simplicity, high interpretability, low data requirements and is inclined to be less overfitting. Thus, it enhances transparency and simplifies instruction and operation.

$$P(Y=1)=\frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n)}} \tag{3}$$

where: $P(Y=1)$ represents the probability of the binary outcome being 1. $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are the coefficients associated with the predictor variables $X_1, X_2, \ldots, X_n$ and e is the base of the natural logarithm, approximately equal to 2.71828.

However, the weakness of regression is that it does not consider interrelationships and cross-information, in this case, XGboost which can be seen in Equation (4) (Chen and Guestrin, 2016) or neural network techniques which would help to produce a better result to achieve the best practices, risk assessment and interoperability as shown in the taxonomy.

$$F(y_t^s, \hat{y}_t^s) = L(y_t^s, \hat{y}_t^s) + \beta\Omega(f) \tag{4}$$

where the objective function $F(y_t^s, \hat{y}_t^s)$ has two components: a loss function $L(y_t^s, \hat{y}_t^s)$ that measures the distance between the actual value y_t^s and the model predicted value $\hat{y}_t^s$; and a regularization term $\Omega(f)$, weighted by hyperparameter β which penalises the number of parameters used in the model to avoid overfitting.

However, some would argue that the use of non-AI techniques would fit in relation to several aspects, such as accountability, user-friendly interface, operational considerations and maintenance. Therefore, the proposed taxonomy enables us to optimise the use of each learning type or non-AI technique to achieve an appropriate CTI-vCISO which is cost-effective for farmers and their stakeholders.

Table 17
Dataset aligned with networking communication in a farm system.

ID paper	CTI database	Features	User-friendly	Zigbee	LoRan-Wan	WiFi	LTE and 5G	NFC (Near Commu-nication)	Satel-lite	Bluetooth	RFID
Gao et al. (2021a)	CVE	CVE IDs and Event description	●	✓	✓	✓	✓	✓	✓	✓	✓
Merah and Kenaza (2021b)	CVSS	Score	●	✓	✓	✓	✓	✓	✓	✓	✓
Merah and Kenaza (2021b)	CCE	System con-figuration issues	●	●	Partly	Partly	Partly	●	●	Partly	●
Gao et al. (2021a)	DAPRA TC	IoC types, IoC relations and Threator's behaviour	●	✓	✓	✓	●	●	●	●	●
Xu et al. (2020)	NVD	CVE, CPE API, Feeds, Vendor comments	Partly	✓	✓	✓	✓	✓	✓	✓	✓
Dulaunoy et al. (2022)	SSH scanner, IP, fingerprint by types	SSH indicators	●	Partly	Partly	Partly	●	●	●	●	●
Edie et al. (2023)	APT Threat Playbooks	Threator behaviours	●	●	●	●	●	●	●	●	●
Wang et al. (2022b)	APTNER	Threator behaviours	●	●	●	●	●	●	●	●	●
Samtani et al. (2021), (Samtani et al., 2016), (Orbinato et al., 2022c)	Hacker Asset Portal	Threator behaviours	Partly	●	●	●	●	●	●	●	●
Sarhan and Spruit (2021)	APT	Threator behaviours	●	●	●	●	●	●	●	●	●
Orbinato et al. (2022a), (Orbinato et al., 2022b)	STIX format	Threator behaviours	●	Partly	Partly	Partly	Partly	Partly	Partly	Partly	Partly
Alsaedi et al. (2022)	Malicious URLs dataset	URLs	●	●	●	●	●	●	●	●	●
Alsaedi et al. (2022), (Meier et al., 2018)	Phishtank	Threator behaviours	●	●	●	●	●	●	●	●	●
Alsaedi et al. (2022)	Malicious URLs dataset	URLs	●	●	●	●	●	●	●	●	●
Suryotrisongko et al. (2022b)	Google Safe browser	TTPs	Partly	●	●	●	●	●	●	●	●
Suryotrisongko et al. (2022b), (Serketzis et al., 2019)	AlienValut	TTPs	●	Partly	Partly	Partly	Partly	Partly	Partly	Partly	Partly
Meier et al. (2018)	The CINS Army List	Threator behaviours	●	●	●	●	●	●	●	●	●
Meier et al. (2018)	Nothink	Threator behaviours	●	●	●	●	●	●	●	●	●
Meier et al. (2018)	Feodo Tracker	traffic network	●	✓	✓	✓	✓	✓	✓	✓	✓
Meier et al. (2018)	SSLIPBlack-list	traffic network	●	✓	✓	✓	✓	✓	✓	✓	✓
Evangelatos et al. (2021)	DNRTI	Threator behaviours	●	●	●	●	●	●	●	●	●

Table 17 (*continued*)

ID paper	CTI database	Features	User-friendly	Zigbee	LoRan-Wan	WiFi	LTE and 5G	NFC (Near Commu-nication)	Satel-lite	Bluetooth	RFID
Dietz et al. (2022)	STIX format	techniques intelligence	●	Partly	Partly	Partly	Partly	Partly	Partly	Partly	Partly
Jo et al. (2022)	TTP	TTPs	●	●	●	●	●	●	●	●	●
Kattamuri et al. (2023)	SOMLAP	Threator behaviours	●	●	●	●	●	●	●	●	●
Robertson et al. (2017), (Al-Fawa'reh et al., 2022b), (Al-Fawa'reh et al., 2022c)	CSE-CIC-IDS2018 on AWS	CVE, CPE API, Feeds, Vendor comments	●	✓	✓	✓	✓	✓	✓	✓	✓
Jiang et al. (2023)	ISCX-IDS-2012 and ISCX-IDS-2019	CVE, CPE API, Feeds, Vendor comments	Partly	✓	✓	✓	✓	✓	✓	✓	✓

● means the dataset does not include feature such as user-friendly or networking communication such as Zigbee, LTE & 5G, etc.
✓means the dataset includes the indicator. **Partly** means the dataset partly includes the indicator.

Table 18
Structured CTI source.

ID Paper	Name	Subname	URLs
Gao et al. (2021a)	CVE		https://cve.mitre.org/
Merah and Kenaza (2021b)	CVSS		https://www.first.org/cvss/
Merah and Kenaza (2021b)	CCE		https://ncp.nist.gov/cce
Gao et al. (2021a)	TTP	DAPRA TC	https://github.com/darpa-i2o/Transparent-Computing/blob/master/README-E3.md
Xu et al. (2020)	NVD		https://nvd.nist.gov/
Dulaunoy et al. (2022)	IP	SSH scanner	https://github.com/D4-project/passive-ssh
Edie et al. (2023)	APT	APT Threat Playbooks	https://github.com/KelsieEdie/Extending-Threat-Playbooks-for-APT-Attribution
Wang et al. (2022b)	APT	APTNER	https://github.com/wangxuren/APTNER
Samtani et al. (2021), (Samtani et al., 2016), (Orbinato et al., 2022c)	APT	Hacker Assest Portal	https://www.azsecure-data.org/hacker-assets-portal.html
Sarhan and Spruit (2021)	APT	No specifc name	https://github.com/IS5882/Open-CyKG/tree/main
Orbinato et al. (2022a), (Orbinato et al., 2022b)	STIX format	No specifc name	https://github.com/dessertlab/cti-to-mitre-with-nlp
Alsaedi et al. (2022)	Malicious URLs dataset	No specific name	https://www.kaggle.com/datasets/saxn/malicious-urls-dataset
Alsaedi et al. (2022), (Meier et al., 2018)	Malicious URLs dataset	Phishtank	https://phishtank.org/
Alsaedi et al. (2022)	Malicious URLs dataset	No specific name	https://www.unb.ca/cic/datasets/url-2016.html
Suryotrisongko et al. (2022b)	OSINT	Google Safe browser	https://safebrowsing.google.com/
Suryotrisongko et al. (2022b), (Serketzis et al., 2019)	OSINT	AlienValut	https://otx.alienvault.com/
Meier et al. (2018)	OSINT	The CINS Army List	https://cinsscore.com/#list
Meier et al. (2018)	OSINT	Nothink	http://www.nothink.org/
Meier et al. (2018)	OSINT	Feodo Tracker	https://feodotracker.abuse.ch/
Meier et al. (2018)	OSINT	SSLIPBlacklist	https://sslbl.abuse.ch/
Evangelatos et al. (2021)	OSINT	DNRTI	https://github.com/SCreaMxp/DNRTI-A-Large-scale-Dataset-for-Named-Entity-Recognition-in-Threat-Intelligence
Dietz et al. (2022)	STIX format	No specific name	https://github.com/digitaltwinCTI/CTI-DT-utilities/tree/master/data
Jo et al. (2022)	TTP	No specific name	https://github.com/MuscleFish/SeqMask/tree/main/datas
Kattamuri et al. (2023)	STIX format	SOMLAP	https://www.kaggle.com/datasets/ravikiranvarmap/somlap-data-set
Robertson et al. (2017), (Al-Fawa'reh et al., 2022b), (Al-Fawa'reh et al., 2022c)	STIX format	CSE-CIC-IDS2018 on AWS	https://www.unb.ca/cic/datasets/ids-2018.html
Jiang et al. (2023)	CTI Feeds	ISCX-IDS-2012 and ISCX-IDS-2019	https://www.unb.ca/cic/datasets/index.html

Table 19
Unstructured CTI source.

ID Paper	Name	Subname	Link
Orbinato et al. (2022c)	Dark web	AZSECURE	https://www.azsecure-data.org/
Orbinato et al. (2022c), Gautam et al. (2020a)	Hacker forums	AZSECURE	https://www.azsecure-data.org/
Hossen et al. (2021b)	Hacker forums	Hacker5	https://forums.hak5.org/
Grisham et al. (2017)	Hacker forums	Ashiyane, Hackhound, VBSpiders, Zloy	
Suryotrisongko et al. (2022a), Deliu et al. (2017), Deliu et al. (2018), Hossen et al. (2021b)	Hacker forums	Nulled.io	https://archive.org/download/nulled.io_database_dump_06052016
Williams et al. (2018)	Hacker forums		OpenSC, Garage4hackers, Hacksden,AntiOnline,Crackingzilla, WebCracking, SafeSkyHacks, Ashiyane, Hack,and Haker
Graf and King (2018)	CTI Feeds	Security Mailing List	https://seclists.org/
Meier et al. (2018)	CTI Feeds	Nothink	http://www.nothink.org/
Meier et al. (2018), Allegretta et al. (2023b)	CTI Feeds	AlienvaultReputation IP	https://otx.alienvault.com/
Meier et al. (2018)	CTI Feeds	Binary Defence	https://www.binarydefense.com/
Meier et al. (2018)	CTI Feeds	Emerging Threats	https://rules.emergingthreats.net/
Meier et al. (2018)	CTI Feeds	Feodo Tracker	https://feodotracker.abuse.ch/
Zhang et al. (2021b)	CTI Feeds	Threat Miner	https://www.threatminer.org/index.php
Zhang et al. (2021b)	CTI Reports	Kaspersky	https://www.kaspersky.com.au/
Khoa et al. (2022)	CTI Feeds	Dataiku	https://www.kaggle.com/datasets/charleswheelus/dataiku-cti
Song et al. (2022b)	Hacker forums		https://github.com/HongyiZhu/D-GEF/tree/master/case_study
Panagiotou et al. (2021)	CTI Feeds	Clear web	https://www.govcert.ch/
Panagiotou et al. (2021)	CTI Feeds	Clear web	https://thehackernews.com/
Panagiotou et al. (2021)	CTI Feeds	Clear web	https://securitynews.sonicwall.com/
Panagiotou et al. (2021)	CTI Feeds	Clear web	https://securelist.com/
Panagiotou et al. (2021)	CTI Feeds	Clear web	https://www.auscert.org.au/
Panagiotou et al. (2021)	CTI Feeds	Clear web	https://www.cbronline.com/
Panagiotou et al. (2021)	CTI Feeds	Clear web	https://us-cert.cisa.gov/
Panagiotou et al. (2021)	CTI Feeds	Clear web	https://www.zdnet.com/
Panagiotou et al. (2021)	CTI Feeds	Clear web	https://edition.cnn.com
Sangher et al. (2023a)	CTI Feeds	Darknet Market place	https://www.kaggle.com/datasets/philipjames11/dark-net-marketplace-drug-data-agora-20142015
Fujii et al. (2022)	CTI Feeds	34 sites	https://link.springer.com/chapter/10.1007/978-3-031-15255-9_5/tables/6

6.3. Limitations and assumptions of vCISO aligning with the SFI context

The proposed vCISO has the following limitations in the agriculture context:

Dissemination of Information (Privacy): vCISO in SFIs assumes that the privacy of disseminating information is secure when it is delivered to third parties or suppliers of IoT equipment in SFIs.

Adaptability (Different Levels of Explainability): Operating within the SFI context, vCISO encounters adaptability challenges due to varying stakeholder understanding levels. It requires effective communication and making cybersecurity decisions that cater to different users of SFIs, as explained in Section 2.1 from agricultural scientists, CISO experts to farmer managers and farmers who have different needs and management perspectives.

Quantum-Safe Explainability: It is assumed that the proposed vCISO is integrable with quantum-safe explainability techniques (Suhasini et al., 2023). This ensures that security decisions remain explainable in a post-quantum computing era, demanding strategic foresight to maintain relevance amid rapid technological evolution in the SFI landscape.

7. Conclusion and future research direction

Our study has presented a comprehensive literature review of CTI techniques and sources that can be tailored to the unique challenges of smart agriculture. By conducting an SLR, the gap in the literature is highlighted and it is clear that none of the selected papers applied CTI techniques using CTI data sources from the agriculture domain or addressed a smart farming infrastructure. Regarding the objectives and the focus of each selected paper, we classified the existing research into different groups that align with the main CTI sources (unstructured and structured sources), as presented in Section 4 of this paper, and detailed the four layers of SFI, as presented in Section 5 of this paper. Depending on the main focus, size and features of the dataset, suitable CTI techniques, either non-AI or AI techniques, can be used. Our SLR provides a benchmark for applying CTI techniques, ensuring the integrity of the agricultural ecosystem and the security of the global food supply chain. Our future research direction will utilise the proposed taxonomy of CTI sources and techniques to design a user-friendly prototype using Explainable AI as a reliable vCISO platform. This design helps other non-technical stakeholders in smart farming architecture to understand and identify the threat intelligence in a qualitative and quantitative approach that may impact their farming system and their supply chain. Emerging AI techniques, such as federated learning and online learning will be applied in the model as a comparison with existing AI techniques as a robustness check. We have aligned and verified CTI in the context of Agriculture 4.0. It will be a solid foundation for us to make farming and agriculture digital transformation more secure in Agriculture 5.0 with twin technology explained in Section 2.2.3, offering solutions to enhance cyber safety and reliability and, more importantly, explainability.

CRediT authorship contribution statement

Hang Thanh Bui: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hamed Aboutorab:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation,

Table 20
Detailed explainable of a taxonomy for vCISO's farmer- friendly CTI in Agriculture.

	Description	Techniques	Drawbacks
Responsibilities	This principle ensures CTI is trustworthy for farmers and its stakeholders to use containing 4 terms as follows.		
Reliable CTI sources	CTI sources played an important role in the accuracy and quality of a CTI as presented in Section 4. Therefore, a reliable CTI source must be verified and validated by the existing literature or the specific authority such as a governmental institute.	As the result of section 4, Tables 5-8 provided the verified and validated CTI source	Each CTI source would fit for different purposes, for example, in Table 17, not every CTI database can provide all essential information for SFI - farmers network communications
Privacy	As mentioned in Section 2.2, one of the major concerns for farmers is privacy when using CTI. CTI should ensure data privacy and confidentiality of SFI, farmers and their stakeholders, particularly in CTI sharing.	Encryption (Czekster et al., 2022), Federated learning (Jiang et al., 2023)	High computational cost
Accountability	Accountability enables CTI to address any mismanagement, misclassification and error during its process. It should be held accountable and audited frequently by the third parties and their stakeholders.	Xgboost (Serketzis et al., 2019)	It is possible to cause misclassification
Transparency	It is a principle of ethical AI to reinforce the trust of farmers and their stakeholders when using CTI. CTI output should be open regarding which data from the SFI, farmers and their stakeholders used.	Federated learning, Blockchain (Jiang et al., 2023), decision trees, linear models, or rule-based systems	High-computational cost
Understandable	It contains 4 terms to ensure the CTI-vCISO platform fits with the Agricultural context and their essential needs.		
Agriculture dummy	SFI has been accelerated applied in agriculture in recent years leaving a gap in farmers and their stakeholders as earlier technology adopters. Therefore, CTI should make sure farmers understand cybersecurity, incidents, and cyber threats and how to prevent them in the simplest and most relatable ways.		
Simplified instruction	The CTI platform must interpret the cyber threat, impact and action in a clear and simple approach to ensure the farmers and their stakeholders can take action to prevent potential threats	Decision tree, XgBoost	A low level of similarity and semantic search
Farmer-friendly interface	It visualises CTI- vCISO including the functionality, display to deliver the explanation of CTI in a simplest and simplified instruction	Elastic search (Almohannadi et al., 2018)	High cost to update as customise for farmer only
Simplified cause-effect	It means CTI-vCISO prioritise the most relevant cause and its effect on SFI	Graph Convolutional Network (Zhao et al., 2020b)	High computation cost and may be overfitting.
Informative plausible	It means CTI-vCISO must provide the essential information to detect, assess risks, best practices described in three following terms.		
Threat detection	The identification of potential cybersecurity threats and be able to provide clear and understandable explanations of the seriousness of these threats.	Naive Bayes, Logistic Regression, SVM, MLP, RNN with LSTM, CNN, and SecureBERT (Evangelatos et al., 2021; Orbinato et al., 2022a)	It requires a high computation cost and may cause misclassification. Preuveneers and Joosen (2021) proposed a model to mitigate the misclassification with F1-score 0.99999.
Risk	It means the assessment of the potential damage, vulnerabilities, and consequences associated with identified threats allows for informed decision-making and remediation strategies.	Data analytics (Spatiotemporal analysis) (Allegretta et al., 2023b)	It is hard to measure the effect of cross-information and interrelationship if two risk categories are similar or under a broad category. On the other hand, it cannot measure which factor should be prioritised.
Best practices	The development of recommendations and actions, policies that represent the most effective and safest ways to conduct explainable CTI-vCISO activities.	Strategic intelligence (Kadoguchi et al., 2019; Samtani et al., 2021)	It does not provide the rank and prioritisation for specific circumstances
Computational efficiency	It means the ability of a CTI-vCISO system to perform tasks and processes efficiently and quickly, thereby optimizing resource utilisation. It includes 4 terms.	Elastic search, Data analytics (Lee, 2023; Zhang et al., 2022a)	It relies on third-party and slow update
Power requirements	It means the amount of electrical energy required to operate a CTI-vCISO system, affecting its energy efficiency and cost		
Scalability	A CTI system's ability to handle increased workloads and data volumes while maintaining performance and functionality		
Latency	It means the time delay between starting a task or query and receiving a response, which affects the real-time nature of CTI-vCISO operations		
Interoperability	The ability of a CTI-vCISO system to work seamlessly with other related systems, tools or protocols to achieve data exchange and collaboration.	Decision tree (Sakthivelu and Vinoth Kumar, 2022)	
Agricultural cost-effectiveness	This criterion ensures CTI-vCISO to be effective and efficient for farmers and their stakeholders to use. It contains the three following terms.		
Operational	The day-to-day operations and activities of CTI-vCISO systems within an organization's cybersecurity infrastructure.	Statis analysis, information retrieval (Molloy et al., 2022)	Slow update
Maintenance	The ongoing efforts and tasks required to ensure the continued functionality, safety, and performance of CTI-vCISO systems	Automated DL-driven CTI modelling (Kumar et al., 2021)	High computational cost
Decommission	The process of decommissioning or shutting down a CTI-vCISO system or component at the end of its life cycle or when it is no longer needed	Kibanna analytics (Almohannadi et al., 2018)	

Conceptualization. **Arash Mahboubi:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **Yansong Gao:** Writing – review & editing, Formal analysis, Data curation. **Nazatul Haque Sultan:** Writing – review & editing, Formal analysis, Data curation. **Aufeef Chauhan:** Writing – review & editing, Formal analysis, Data curation. **Mohammad Zavid Parvez:** Formal analysis, Data curation. **Michael Bewong:** Writing – review & editing. **Rafiqul Islam:** Writing – review & editing. **Zahid Islam:** Writing – review & editing. **Seyit A. Camtepe:** Writing – review & editing. **Praveen Gauravaram:** Writing – review & editing. **Dineshkumar Singh:** Writing – review & editing. **M. Ali Babar:** Writing – review & editing. **Shihao Yan:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

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

Acknowledgements

The work has been supported by the Cyber Security Research Centre Limited (Grant number C11-00239) whose activities are partially funded by the Australian Government's Cooperative Research Centre Program.

References

ABC Rural, Buchanan, Kallee, Murphy, Tanya, 2022. John Deere tractor hack reveals food supply vulnerable to cyber attacks. [Online]. Available https://www.abc.net.au/news/rural/2022-08-24/tractor-hack-reveals-food-supply-vulnerable/101360062.

Adewopo, V., Gonen, B., Adewopo, F., 2020a. Exploring open source information for cyber threat intelligence. In: 2020 IEEE International Conference on Big Data (Big Data), pp. 2232–2241.

Adewopo, V., Gonen, B., Adewopo, F., 2020b. Exploring open source information for cyber threat intelligence. In: 2020 IEEE International Conference on Big Data (Big Data). IEEE, pp. 2232–2241.

Afzaliseresht, N., Miao, Y., Michalska, S., Liu, Q., Wang, H., 2020. From logs to stories: human-centred data mining for cyber threat intelligence. IEEE Access 8, 089.

Ahmed, M.A., Gallardo, J.L., Zuniga, M.D., Pedraza, M.A., Carvajal, G., Jara, N., Carvajal, R., 2022. Lora based IoT platform for remote monitoring of large-scale agriculture farms in Chile. Sensors 22 (8), 2824.

Al-Fawa'reh, M., Al-Fayoumi, M., Nashwan, S., Fraihat, S., 2022a. Cyber threat intelligence using PCA-DNN model to detect abnormal network behavior. Egypt. Inform. J. 23 (2), 173–185.

Al-Fawa'reh, M., Al-Fayoumi, M., Nashwan, S., Fraihat, S., 2022b. Cyber threat intelligence using PCA-DNN model to detect abnormal network behavior. Egypt. Inform. J. 23 (2), 173–185.

Al-Fawa'reh, M., Al-Fayoumi, M., Nashwan, S., Fraihat, S., 2022c. Cyber threat intelligence using PCA-DNN model to detect abnormal network behavior. Egypt. Inform. J. 23 (2), 173–185.

Al-Ofeishat, H.A., Al Rababah, M.A., 2012. Near field communication (NFC). Int. J. Comput. Sci. Netw. Secur. 12 (2), 93.

Alahmadi, A.N., Rehman, S.U., Alhazmi, H.S., Glynn, D.G., Shoaib, H., Solé, P., 2022. Cyber-security threats and side-channel attacks for digital agriculture. Sensors 22 (9), 3520.

Allegretta, M., Siracusano, G., Gonzalez, R., Gramaglia, M., 2023b. Are crowd-sourced CTI datasets ready for supporting anti-cybercrime intelligence? Comput. Netw. 234, 109920.

Allegretta, M., Siracusano, G., Gonzalez, R., Vallina, P., Gramaglia, M., 2023a. Using CTI data to understand real world cyberattacks. In: 2023 18th Wireless On-Demand Network Systems and Services Conference (WONS). IEEE, pp. 100–103.

Almohannadi, H., Awan, I., Al Hamar, J., Cullen, A., Disso, J.P., Armitage, L., 2018. Cyber threat intelligence from honeypot data using elasticsearch. In: 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA). IEEE, pp. 900–906.

Alnowaiser, K.K., Ahmed, M.A., 2023. Digital twin: current research trends and future directions. Arab. J. Sci. Eng. 48 (2), 1075–1095.

Alsaedi, M., Ghaleb, F.A., Saeed, F., Ahmad, J., Alasli, M., 2022. Cyber threat intelligence-based malicious URL detection model using ensemble learning. Sensors 22 (9), 3373.

Ammi, M., Adedugbe, O., Alharby, F.M., Benkhelifa, E., 2022. Leveraging a cloud-native architecture to enable semantic interconnectedness of data for cyber threat intelligence. Clust. Comput. 25 (5), 3629–3640.

Ampel, B., Samtani, S., Zhu, H., Ullman, S., Chen, H., 2020. Labeling hacker exploits for proactive cyber threat intelligence: a deep transfer learning approach. In: 2020 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, pp. 1–6.

Arıkan, S.M., Acar, S., 2021. A data mining based system for automating creation of cyber threat intelligence. In: 2021 9th International Symposium on Digital Forensics and Security (ISDFS). IEEE, pp. 1–7.

Arnold, N., Ebrahimi, M., Zhang, N., Lazarine, B., Patton, M., Chen, H., Samtani, S., 2019. Dark-net ecosystem cyber-threat intelligence (CTI) tool. In: 2019 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, pp. 92–97.

Atifi, A., Bou-Harb, E., 2017. On correlating network traffic for cyber threat intelligence: a bloom filter approach. In: 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC). IEEE, pp. 384–389.

Attri, I., Awasthi, L.K., Sharma, T.P., Rathee, P., 2023. A review of deep learning techniques used in agriculture. Ecol. Inform., 102217.

Baker, L., Green, R., 2019. Cyber Security in UK Agriculture. NCC Group.

Bartnes, M., Zand, A., Stringhini, G., Kemmerer, R., 2014. Targeted attacks against industrial control systems: is the power industry prepared? In: Proceedings of the ACM Conference on Computer and Communications Security, pp. 13–22.

Baryshnikova, N., Altukhov, P., Naidenova, N., Shkryabina, A., 2022. Ensuring global food security: transforming approaches in the context of agriculture 5.0. IOP Conf. Ser. Earth Environ. Sci. 988 (3), 032024. IOP Publishing.

Becker, J., 2020. Cyber attack forces cancellation of wool sales across Australia. [Online]. Available https://www.abc.net.au/news/rural/2020-02-27/ransomware-cyber-attack-cripples-australian-wool-sales/12007912. (Accessed 22 January 2024).

Borchi, M.R. John, Woodcock, Melanie, Raniga, B., 2021. A threat-based assessment of the cyber resilience of the Australian agricultural sector. AgriFutures Aust.. Cyber security threats – are we prepared? Publication No. 21-070 Project No. PRJ-012909. [Online]. Available https://agrifutures.com.au/wp-content/uploads/2021/07/21-070.pdf.

Bose, A., Sundari, S.G., Behzadan, V., Hsu, W.H., 2021. Tracing relevant Twitter accounts active in cyber threat intelligence domain by exploiting content and structure of Twitter network. In: 2021 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, pp. 1–6.

Bou-Harb, E., 2016. A probabilistic model to preprocess darknet data for cyber threat intelligence generation. In: 2016 IEEE International Conference on Communications (ICC). IEEE, pp. 1–6.

Chakir, O., Sadqi, Y., Maleh, Y., 2023. Evaluation of open-source web application firewalls for cyber threat intelligence. In: Big Data Analytics and Intelligent Systems for Cyber Threat Intelligence, pp. 35–48.

Chen, C.-M., Hsu, F.-H., Hwang, J.-N., 2023b. Useful cyber threat intelligence relation retrieval using transfer learning. In: Proceedings of the 2023 European Interdisciplinary Cybersecurity Conference, pp. 42–46.

Chen, M., Liu, W., Zhang, N., Li, J., Ren, Y., Yi, M., Liu, A., 2022. GPDS: a multi-agent deep reinforcement learning game for anti-jamming secure computing in MEC network. Expert Syst. Appl. 210, 118394. [Online]. Available https://www.sciencedirect.com/science/article/pii/S0957417422015044.

Chen, M., Liu, A., Xiong, N.N., Song, H., Leung, V.C.M., 2023a. SGPL: an intelligent game-based secure collaborative communication scheme for metaverse over 5g and beyond networks. IEEE J. Sel. Areas Commun., 1.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Chi, H., Martin, A.R., Scarlett, C.Y., 2018. Data analytics for cyber threat intelligence. Anal. Knowl. Manag., 407–431.

Czekster, R.M., Metere, R., Morisset, C., 2022. Incorporating cyber threat intelligence into complex cyber-physical systems: a stix model for active buildings. Appl. Sci. 12 (10), 5005.

Dalziel, H., 2014. How to Define and Build an Effective Cyber Threat Intelligence Capability. Syngress.

Dara, R., Hazrati Fard, S.M., Kaur, J., 2022. Recommendations for ethical and responsible use of artificial intelligence in digital agriculture. Front. Artif. Intell. 5, 884192.

de Oca, E.M., Armin, J., Consoli, A., 2022. Cyber-threat intelligence from European-wide sensor network in SISSDEN. In: Challenges in Cybersecurity and Privacy-the European Research Landscape. River Publishers, pp. 117–128.

Deliu, I., Leichter, C., Franke, K., 2017. Extracting cyber threat intelligence from hacker forums: support vector machines versus convolutional neural networks. In: 2017 IEEE International Conference on Big Data (Big Data). IEEE, pp. 3648–3656.

Deliu, I., Leichter, C., Franke, K., 2018. Collecting cyber threat intelligence from hacker forums via a two-stage, hybrid process using support vector machines and latent Dirichlet allocation. In: 2018 IEEE International Conference on Big Data (Big Data). IEEE, pp. 5008–5013.

Demestichas, K., Peppes, N., Alexakis, T., 2020. Survey on security threats in agricultural IoT and smart farming. Sensors 20 (22), 6458.

Department of Agriculture, Water and the Environment (Australia). Snapshot of Australian agriculture [Online]. Available https://www.agriculture.gov.au/abares/products/insights/snapshot-of-australian-agriculture#around-72-of-agricultural-production-is-exported. (Accessed 9 October 2023).

Dhake, B., Shetye, C., Borhade, P., Gawas, D., Nerurkar, A., 2023. Stratification of hacker forums and predicting cyber assaults for proactive cyber threat intelligence. In: 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS). IEEE, pp. 1–6.

Dietz, M., Schlette, D., Pernul, G., 2022. Harnessing digital twin security simulations for systematic cyber threat intelligence. In: 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC). IEEE, pp. 789–797.

Dulaunoy, A., Huynen, J.-L., Thirion, A., 2022. Active and passive collection of SSH key material for cyber threat intelligence. Digit. Treats Res. Pract. 3 (3), 1–5.

Edie, K., Mckee, C., Duby, A., 2023. Extending threat playbooks for cyber threat intelligence: a novel approach for APT attribution. In: 2023 11th International Symposium on Digital Forensics and Security (ISDFS). IEEE, pp. 1–6.

Elsaeidy, A.A., Jagannath, N., Sanchis, A.G., Jamalipour, A., Munasinghe, K.S., 2020. Replay attack detection in smart cities using deep learning. IEEE Access 8, 825–137 837.

Evangelatos, P., Iliou, C., Mavropoulos, T., Apostolou, K., Tsikrika, T., Vrochidis, S., Kompatsiaris, I., 2021. Named entity recognition in cyber threat intelligence using transformer-based models. In: 2021 IEEE International Conference on Cyber Security and Resilience (CSR). IEEE, pp. 348–353.

Ferrag, M.A., Shu, L., Djallel, H., Choo, K.-K.R., 2021. Deep learning-based intrusion detection for distributed denial of service attack in agriculture 4.0. Electronics 10 (11) [Online]. Available https://www.mdpi.com/2079-9292/10/11/1257.

Fuentealba, D., Flores, C., Soto, I., Zamorano, R., Reid, S., 2022. Guidelines for digital twins in 5g agriculture. In: 2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), pp. 613–618.

Fujii, S., Kawaguchi, N., Shigemoto, T., Yamauchi, T., 2022. CyNER: information extraction from unstructured text of CTI sources with noncontextual IOCs. In: International Workshop on Security. Springer, pp. 85–104.

Gajek, S., Jensen, M., Liao, L., Schwenk, J., 2009. Analysis of signature wrapping attacks and countermeasures. In: 2009 IEEE International Conference on Web Services, pp. 575–582.

Gao, P., Shao, F., Liu, X., Xiao, X., Liu, H., Qin, Z., Xu, F., Mittal, P., Kulkarni, S.R., Song, D., 2021b. A system for efficiently hunting for cyber threats in computer systems using threat intelligence. In: 2021 IEEE 37th International Conference on Data Engineering (ICDE), pp. 2705–2708.

Gao, P., Shao, F., Liu, X., Xiao, X., Qin, Z., Xu, F., Mittal, P., Kulkarni, S.R., Song, D., 2021a. Enabling efficient cyber threat hunting with cyber threat intelligence. In: 2021 IEEE 37th International Conference on Data Engineering (ICDE). IEEE, pp. 193–204.

Gao, Y., Doan, B.G., Zhang, Z., Ma, S., Zhang, J., Fu, A., Nepal, S., Kim, H., 2020a. Backdoor attacks and countermeasures on deep learning: a comprehensive review. arXiv preprint. arXiv:2007.10760.

Gao, Y., Li, X., Peng, H., Fang, B., Philip, S.Y., 2020b. HinCTI: a cyber threat intelligence modeling and identification system based on heterogeneous information network. IEEE Trans. Knowl. Data Eng. 34 (2), 708–722.

Gautam, A.S., Gahlot, Y., Kamat, P., 2020a. Hacker forum exploit and classification for proactive cyber threat intelligence. In: Inventive Computation Technologies 4. Springer, pp. 279–285.

Gautam, A.S., Gahlot, Y., Kamat, P., 2020b. Hacker forum exploit and classification for proactive cyber threat intelligence. In: Inventive Computation Technologies 4. Springer, pp. 279–285.

Ge, W., Wang, J., 2022. SeqMask: behavior extraction over cyber threat intelligence via multi-instance learning. Comput. J., bxac172.

Gong, S., Lee, C., 2021a. Efficient data noise-reduction for cyber threat intelligence system. In: Advances in Computer Science and Ubiquitous Computing: CSA-CUTE 2019. Springer, pp. 591–597.

Gong, S., Lee, C., 2021b. Cyber threat intelligence framework for incident response in an energy cloud platform. Electronics 10 (3), 239.

Gong, S., Lee, C., 2021c. Cyber threat intelligence framework for incident response in an energy cloud platform. Electronics 10 (3), 239.

Graf, R., King, R., 2018. Neural network and blockchain based technique for cyber threat intelligence and situational awareness. In: 2018 10th International Conference on Cyber Conflict (CyCon), pp. 409–426.

Grisham, J., Samtani, S., Patton, M., Chen, H., 2017. Identifying mobile malware and key threat actors in online hacker forums for proactive cyber threat intelligence. In: 2017 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, pp. 13–18.

Guarascio, M., Cassavia, N., Pisani, F.S., Manco, G., 2022. Boosting cyber-threat intelligence via collaborative intrusion detection. Future Gener. Comput. Syst. 135, 30–43.

Gylling, A., Ekstedt, M., Afzal, Z., Eliasson, P., 2021. Mapping cyber threat intelligence to probabilistic attack graphs. In: 2021 IEEE International Conference on Cyber Security and Resilience (CSR). IEEE, pp. 304–311.

Haque, M.A., Shetty, S., Kamhoua, C.A., Gold, K., 2023. Adversarial technique validation & defense selection using attack graph & ATT&CK matrix. In: 2023 International Conference on Computing, Networking and Communications (ICNC), pp. 181–187.

Haque, M.F., Krishnan, R., 2021. Toward automated cyber defense with secure sharing of structured cyber threat intelligence. Inf. Syst. Front., 1–14.

Haxhibeqiri, J., De Poorter, E., Moerman, I., Hoebeke, J., 2018. A survey of lorawan for IoT: from technology to application. Sensors 18 (11), 3995.

Hossen, M.I., Islam, A., Anowar, F., Ahmed, E., Rahman, M.M., 2021a. Generating cyber threat intelligence to discover potential security threats using classification and topic modeling. In: Cyber Security Using Modern Technologies. CRC Press, pp. 141–153.

Hossen, M.I., Islam, A., Anowar, F., Ahmed, E., Rahman, M.M., 2021b. Generating cyber threat intelligence to discover potential security threats using classification and topic modeling. In: Cyber Security Using Modern Technologies. CRC Press, pp. 141–153.

Hu, Y., Kuang, W., Qin, Z., Li, K., Zhang, J., Gao, Y., Li, W., Li, K., 2021. Artificial intelligence security: threats and countermeasures. ACM Comput. Surv. 55 (1), 1–36.

Husari, G., Niu, X., Chu, B., Al-Shaer, E., 2018. Using entropy and mutual information to extract threat actions from cyber threat intelligence. In: 2018 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, pp. 1–6.

Irfan, A.N., Chuprat, S., Mahrin, M.N., Ariffin, A., 2022. Taxonomy of cyber threat intelligence framework. In: 2022 13th International Conference on Information and Communication Technology Convergence (ICTC). IEEE, pp. 1295–1300.

Irshad, E., Siddiqui, A.B., 2023. Cyber threat attribution using unstructured reports in cyber threat intelligence. Egypt. Inform. J. 24 (1), 43–59.

Jiang, T., Shen, G., Guo, C., Cui, Y., Xie, B., 2023. BFLS: blockchain and federated learning for sharing threat detection models as cyber threat intelligence. Comput. Netw. 224, 109604.

Jo, H., Lee, Y., Shin, S., 2022. Vulcan: automatic extraction and analysis of cyber threat intelligence from unstructured text. Comput. Secur. 120, 102763.

Juels, A., 2006. RFID security and privacy: a research survey. IEEE J. Sel. Areas Commun. 24 (2), 381–394.

Kadoguchi, M., Hayashi, S., Hashimoto, M., Otsuka, A., 2019. Exploring the dark web for cyber threat intelligence using machine leaning. In: 2019 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, pp. 200–202.

Kadoguchi, M., Kobayashi, H., Hayashi, S., Otsuka, A., Hashimoto, M., 2020. Deep self-supervised clustering of the dark web for cyber threat intelligence. In: 2020 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, pp. 1–6.

Kaiser, F.K., Andris, L.J., Tennig, T.F., Iser, J.M., Wiens, M., Schultmann, F., 2022. Cyber threat intelligence enabled automated attack incident response. In: 2022 3rd International Conference on Next Generation Computing Applications (NextComp). IEEE, pp. 1–6.

Karunathilake, E.M.B.M., Le, A.T., Heo, S., Chung, Y.S., Mansoor, S., 2023. The path to smart farming: innovations and opportunities in precision agriculture. Agriculture 13 (8). [Online]. Available https://www.mdpi.com/2077-0472/13/8/1593.

Kattamuri, S.J., Penmatsa, R.K.V., Chakravarty, S., Madabathula, V.S.P., 2023. Swarm optimization and machine learning applied to PE malware detection towards cyber threat intelligence. Electronics 12 (2), 342.

Khoa, N.H., Trung, D.M., Khoa, B.C.N., Au, T.T.M., Ung, V.-G., Duy, P.T., Pham, V.-H., 2022. Cyber threat intelligence for proactive defense against adversary in SDN-assisted IIoTs context. In: 2022 RIVF International Conference on Computing and Communication Technologies (RIVF). IEEE, pp. 1–6.

Kim, H., Kim, H., et al., 2022. Comparative experiment on TTP classification with class imbalance using oversampling from CTI dataset. Secur. Commun. Netw. 2022.

Kim, H.-S., Kumar, S., Culler, D.E., 2019a. Thread/openthread: a compromise in low-power wireless multihop network architecture for the Internet of things. IEEE Commun. Mag. 57 (7), 55–61.

Kim, N., Kim, M., Lee, S., Cho, H., Kim, B.-i., Park, J.-h., Jun, M., 2019b. Study of natural language processing for collecting cyber threat intelligence using syntaxnet. In: Proceedings of the 3rd International Symposium of Information and Internet Technology (SYMINTECH 2018). Springer, pp. 10–18.

Kitchenham, B., Pretorius, R., Budgen, D., Brereton, O.P., Turner, M., Niazi, M., Linkman, S., 2010. Systematic literature reviews in software engineering–a tertiary study. Inf. Softw. Technol. 52 (8), 792–805.

Kitchenham, B.A., 2012. Systematic review in software engineering: where we are and where we should be going. In: Proceedings of the 2nd International Workshop on Evidential Assessment of Software Technologies, pp. 1–2.

Koloveas, P., Chantzios, T., Tryfonopoulos, C., Skiadopoulos, S., 2019. A crawler architecture for harvesting the clear, social, and dark web for IoT-related cyber-threat intelligence. In: 2019 IEEE World Congress on Services (SERVICES), vol. 2642. IEEE, pp. 3–8.

Koloveas, P., Chantzios, T., Alevizopoulou, S., Skiadopoulos, S., Tryfonopoulos, C., 2021. Intime: a machine learning-based framework for gathering and leveraging web data to cyber-threat intelligence. Electronics 10 (7), 818.

Konečnỳ, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T., Bacon, D., 2016. Federated learning: strategies for improving communication efficiency. arXiv preprint. arXiv:1610.05492.

Kristiansen, L.-M., Agarwal, V., Franke, K., Shah, R.S., 2020. CTI-Twitter: gathering cyber threat intelligence from Twitter using integrated supervised and unsupervised learning. In: 2020 IEEE International Conference on Big Data (Big Data). IEEE, pp. 2299–2308.

Kumar, P., Gupta, G.P., Tripathi, R., Garg, S., Hassan, M.M., 2021. DLTIF: deep learning-driven cyber threat intelligence modeling and identification framework in IoT-enabled maritime transportation systems. IEEE Trans. Intell. Transp. Syst.

Kumar, S., Janet, B., Eswari, R., 2019. Multi platform honeypot for generation of cyber threat intelligence. In: 2019 IEEE 9th International Conference on Advanced Computing (IACC). IEEE, pp. 25–29.

Landauer, M., Skopik, F., Wurzenberger, M., Hotwagner, W., Rauber, A., 2019. A framework for cyber threat intelligence extraction from raw log data. In: 2019 IEEE International Conference on Big Data (Big Data). IEEE, pp. 3200–3209.

Lavric, A., Petrariu, A.I., Popa, V., 2019. Sigfox communication protocol: the new era of IoT? In: 2019 International Conference on Sensing and Instrumentation in IoT Era (ISSI). IEEE, pp. 1–4.

Lee, H.-W., 2023. Analysis of digital forensic artifacts data enrichment mechanism for cyber threat intelligence. In: Proceedings of the 2023 12th International Conference on Software and Computer Applications, pp. 192–199.
Leite, C., den Hartog, J., Ricardo dos Santos, D., Costante, E., 2022. Actionable cyber threat intelligence for automated incident response. In: Nordic Conference on Secure IT Systems. Springer, pp. 368–385.
Li, K., Wen, H., Li, H., Zhu, H., Sun, L., 2018. Security OSIF: toward automatic discovery and analysis of event based cyber threat intelligence. In: 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, pp. 741–747.
Li, Z., Zeng, J., Chen, Y., Liang, Z., 2022. AttacKG: constructing technique knowledge graph from cyber threat intelligence reports. In: European Symposium on Research in Computer Security. Springer, pp. 589–609.
Li, Z.-X., Li, Y.-J., Liu, Y.-W., Liu, C., Zhou, N.-X., 2023. K-CTIAA: automatic analysis of cyber threat intelligence based on a knowledge graph. Symmetry 15 (2), 337.
Liao, X., Yuan, K., Wang, X., Li, Z., Xing, L., Beyah, R., 2016. Acing the IOC game: toward automatic discovery and analysis of open-source cyber threat intelligence. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 755–766.
Liu, J., Yan, J., Jiang, J., He, Y., Wang, X., Jiang, Z., Yang, P., Li, N., 2022. TriCTI: an actionable cyber threat intelligence discovery system via trigger-enhanced neural network. Cybersecurity 5 (1), 8.
Liu, X., Jiang, D., Tao, B., Xiang, F., Jiang, G., Sun, Y., Kong, J., Li, G., 2023. A systematic review of digital twin about physical entities, virtual models, twin data, and applications. Adv. Eng. Inform. 55, 101876. [Online]. Available https://www.sciencedirect.com/science/article/pii/S1474034623000046.
Madsen, A., Reddy, S., Chandar, S. Post-hoc interpretability for neural NLP: a survey. https://arxiv.org/abs/2108.04840.
Marques, R.S., Al-Khateeb, H., Epiphaniou, G., Maple, C., 2022. Pivot attack classification for cyber threat intelligence. J. Inf. Secur. Cybercrimes Res. 5 (2), 91–103.
Martins, C., Medeiros, I., 2022. Generating quality threat intelligence leveraging OSINT and a cyber threat unified taxonomy. ACM Trans. Priv. Secur. 25 (3), 1–39.
Mavroeidis, V., Hohimer, R., Casey, T., Jesang, A., 2021. Threat actor type inference and characterization within cyber threat intelligence. In: 2021 13th International Conference on Cyber Conflict (CyCon), pp. 327–352.
Meier, R., Scherrer, C., Gugelmann, D., Lenders, V., Vanbever, L., 2018. Feedrank: a tamper- resistant method for the ranking of cyber threat intelligence feeds. In: 2018 10th International Conference on Cyber Conflict (CyCon), pp. 321–344.
Merah, Y., Kenaza, T., 2021a. Proactive ontology-based cyber threat intelligence analytic. In: 2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI). IEEE, pp. 1–7.
Merah, Y., Kenaza, T., 2021b. Ontology-based cyber risk monitoring using cyber threat intelligence. In: Proceedings of the 16th International Conference on Availability, Reliability and Security, pp. 1–8.
Miles, C., Lakhotia, A., LeDoux, C., Newsom, A., Notani, V., 2014. Virusbattle: state-of-the-art malware analysis for better cyber threat intelligence. In: 2014 7th International Symposium on Resilient Control Systems (ISRCS). IEEE, pp. 1–6.
Molloy, C., Charland, P., Ding, S.H., Fung, B.C., 2022. JARV1S: phenotype clone search for rapid zero-day malware triage and functional decomposition for cyber threat intelligence. In: 2022 14th International Conference on Cyber Conflict: Keep Moving!(CyCon), vol. 700. IEEE, pp. 385–403.
Montasari, R., Carroll, F., Macdonald, S., Jahankhani, H., Hosseinian-Far, A., Daneshkhah, A., 2021a. Application of artificial intelligence and machine learning in producing actionable cyber threat intelligence. In: Digital Forensic Investigation of Internet of Things (IoT) Devices, pp. 47–64.
Montasari, R., Carroll, F., Macdonald, S., Jahankhani, H., Hosseinian-Far, A., Daneshkhah, A., 2021b. Application of artificial intelligence and machine learning in producing actionable cyber threat intelligence. In: Digital Forensic Investigation of Internet of Things (IoT) Devices, pp. 47–64.
Moraliyage, H., Sumanasena, V., De Silva, D., Nawaratne, R., Sun, L., Alahakoon, D., 2022. Multimodal classification of onion services for proactive cyber threat intelligence using explainable deep learning. IEEE Access 10, 044.
Morris, J.X., Lifland, E., Yoo, J.Y., Qi, Y., 2020. Textattack: a framework for adversarial attacks in natural language processing. arXiv. In: Proceedings of the 2020 EMNLP.
Musa, S., Parish, D.J., 2007. Visualising communication network security attacks. In: 2007 11th International Conference Information Visualization (IV'07). IEEE, pp. 726–733.
Nagasawa, R., Furumoto, K., Takita, M., Shiraishi, Y., Takahashi, T., Mohri, M., Takano, Y., Morii, M., 2021. Partition-then-overlap method for labeling cyber threat intelligence reports by topics over time. IEICE Trans. Inf. Syst. 104 (5), 556–561.
Orbinato, V., Barbaraci, M., Natella, R., Cotroneo, D., 2022a. Automatic mapping of unstructured cyber threat intelligence: an experimental study:(practical experience report). In: 2022 IEEE 33rd International Symposium on Software Reliability Engineering (ISSRE). IEEE, pp. 181–192.
Orbinato, V., Barbaraci, M., Natella, R., Cotroneo, D., 2022b. Automatic mapping of unstructured cyber threat intelligence: an experimental study:(practical experience report). In: 2022 IEEE 33rd International Symposium on Software Reliability Engineering (ISSRE). IEEE, pp. 181–192.
Orbinato, V., Barbaraci, M., Natella, R., Cotroneo, D., 2022c. Automatic mapping of unstructured cyber threat intelligence: an experimental study: (practical experience report). In: 2022 IEEE 33rd International Symposium on Software Reliability Engineering (ISSRE). IEEE, pp. 181–192.
Panagiotou, P., Iliou, C., Apostolou, K., Tsikrika, T., Vrochidis, S., Chatzimisios, P., Kompatsiaris, I., 2021. Towards selecting informative content for cyber threat intelligence. In: 2021 IEEE International Conference on Cyber Security and Resilience (CSR). IEEE, pp. 354–359.
Post, T.S.C.M., 2019. Chinese pig farm's jammer disrupts GPS signals for aircraft. The South China Morning Post [Online]. Available https://www.scmp.com/news/china/society/article/3042991/china-flight-systems-jammed-pig-farms-african-swine-fever.
Pour, M.S., Watson, D., Bou-Harb, E., 2021. Sanitizing the IoT cyber security posture: an operational CTI feed backed up by Internet measurements. In: 2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). IEEE, pp. 497–506.
Preuveneers, D., Joosen, W., 2021. Sharing machine learning models as indicators of compromise for cyber threat intelligence. J. Cybersecur. Priv. 1 (1), 140–163.
Ramya, C.M., Shanmugaraj, M., Prabakaran, R., 2011. Study on Zigbee Technology. 2011 3rd International Conference on Electronics Computer Technology, vol. 6. IEEE, pp. 297–301.
Rana, M.U., Ellahi, O., Alam, M., Webber, J.L., Mehbodniya, A., Khan, S., 2022. Offensive security: cyber threat intelligence enrichment with counterintelligence and counter-attack. IEEE Access 10, 760–108 774.
Riesco, R., Villagrá, V.A., 2019. Leveraging cyber threat intelligence for a dynamic risk framework: automation by using a semantic reasoner and a new combination of standards (STIX™, SWRL and OWL). Int. J. Inf. Secur. 18 (6), 715–739.
Robertson, J., Diab, A., Marin, E., Nunes, E., Paliath, V., Shakarian, J., Shakarian, P., 2017. Darkweb Cyber Threat Intelligence Mining. Cambridge University Press.
Sakthivelu, U., Vinoth Kumar, C., 2022. An approach on cyber threat intelligence using recurrent neural network. In: ICT Infrastructure and Computing: Proceedings of ICT4SD 2022. Springer, pp. 429–439.
Samtani, S., Chinn, K., Larson, C., Chen, H., 2016. Azsecure hacker assets portal: cyber threat intelligence and malware analysis. In: 2016 IEEE Conference on Intelligence and Security Informatics (ISI). IEEE, pp. 19–24.
Samtani, S., Chinn, R., Chen, H., Nunamaker Jr, J.F., 2017. Exploring emerging hacker assets and key hackers for proactive cyber threat intelligence. J. Manag. Inf. Syst. 34 (4), 1023–1053.
Samtani, S., Li, W., Benjamin, V., Chen, H., 2021. Informing cyber threat intelligence through dark web situational awareness: the azsecure hacker assets portal. Digit. Treats Res. Pract. 2 (4), 1–10.
Sangher, K., Singh, A., Pandey, H.M., Kumar, V., 2023b. Towards safe cyber practices: developing proactive cyber threat intelligence system for dark web forums content by employing deep learning approaches. Inf. Sci. 14 (6).
Sangher, K.S., Singh, A., Pandey, H.M., Kumar, V., 2023a. Towards safe cyber practices: developing a proactive cyber-threat intelligence system for dark web forum content by identifying cybercrimes. Information 14 (6), 349.
Sanjeev, K., Janet, B., Eswari, R., 2020. Automated cyber threat intelligence generation from honeypot data. In: Inventive Communication and Computational Technologies: Proceedings of ICICCT 2019. Springer, pp. 591–598.
Sarhan, I., Spruit, M., 2021. Open-CyKG: an open cyber threat intelligence knowledge graph. Knowl.-Based Syst. 233, 107524.
Schlette, D., Caselli, M., Pernul, G., 2021b. A comparative study on cyber threat intelligence: the security incident response perspective. IEEE Commun. Surv. Tutor. 23 (4), 2525–2556.
Schlette, D., Vielberth, M., Pernul, G., 2021a. CTI-SOC2M2–the quest for mature, intelligence-driven security operations and incident response capabilities. Comput. Secur. 111, 102482.
Serketzis, N., Katos, V., Ilioudis, C., Baltatzis, D., Pangalos, G., 2019. Improving forensic triage efficiency through cyber threat intelligence. Future Internet 11 (7), 162.
Settanni, G., Shovgenya, Y., Skopik, F., Graf, R., Wurzenberger, M., Fiedler, R., 2017. Acquiring cyber threat intelligence through security information correlation. In: 2017 3rd IEEE International Conference on Cybernetics (CYBCONF). IEEE, pp. 1–7.
Shin, Y., Lim, C., Park, M., Cho, S., Han, I., Oh, H., Lee, K., 2019. Alert correlation using diamond model for cyber threat intelligence. In: Proceedings of the European Conference on Cyber Warfare and Security. Academic Conferences International Limited Oxfordshire, UK, pp. 444–450.
Sistu, S., Liu, Q., Ozcelebi, T., Dijk, E., Zotti, T., 2019. Performance evaluation of thread protocol based wireless mesh networks for lighting systems. In: 2019 International Symposium on Networks, Computers and Communications (ISNCC). IEEE, pp. 1–8.
Song, B., Chen, R., Liu, B., Jiang, Z., Wang, X., 2022a. Time series attention based transformer neural Turing machines for diachronic graph embedding in cyber threat intelligence. In: International Conference on Computational Science. Springer, pp. 17–30.
Song, B., Chen, R., Liu, B., Jiang, Z., Wang, X., 2022b. The hyperbolic temporal attention based differentiable neural Turing machines for diachronic graph embedding in cyber threat intelligence. In: 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, pp. 1353–1359.
Sontowski, S., Gupta, M., Laya Chukkapalli, S.S., Abdelsalam, M., Mittal, S., Joshi, A., Sandhu, R., 2020. Cyber attacks on smart farming infrastructure. In: 2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC), pp. 135–143.
Spyros, A., Papoutsis, A., Koritsas, I., Mengidis, N., Iliou, C., Kavallieros, D., Tsikrika, T., Vrochidis, S., Kompatsiaris, I., 2022. Towards continuous enrichment of cyber threat

intelligence: a study on a honeypot dataset. In: 2022 IEEE International Conference on Cyber Security and Resilience (CSR). IEEE, pp. 267–272.
Suhasini, S., Tatini, N.B., Arslan, F., et al., 2023. Smart explainable artificial intelligence for sustainable secure healthcare application based on quantum optical neural network. Opt. Quantum Electron. 55, 887. https://doi.org/10.1007/s11082-023-05155-3 [Online].
Sun, T., Yang, P., Li, M., Liao, S., 2021a. An automatic generation approach of the cyber threat intelligence records based on multi-source information fusion. Future Internet 13 (2), 1–19.
Sun, T., Yang, P., Li, M., Liao, S., 2021b. An automatic generation approach of the cyber threat intelligence records based on multi-source information fusion. Future Internet 13 (2), 40.
Suryotrisongko, H., Ginardi, H., Ciptaningtyas, H.T., Dehqan, S., Musashi, Y., 2022a. Topic modeling for cyber threat intelligence (CTI). In: 2022 Seventh International Conference on Informatics and Computing (ICIC). IEEE, pp. 1–7.
Suryotrisongko, H., Musashi, Y., Tsuneda, A., Sugitani, K., 2022b. Robust botnet DGA detection: blending XAI and OSINT for cyber threat intelligence sharing. IEEE Access 10, 613–34 624.
Tekin, U., Yilmaz, E.N., 2021. Obtaining cyber threat intelligence data from Twitter with deep learning methods. In: 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE, pp. 82–86.
Tounsi, W., Rais, H., 2018. A survey on technical threat intelligence in the age of sophisticated cyber attacks. Comput. Secur. 72, 212–233. [Online]. Available https://www.sciencedirect.com/science/article/pii/S0167404817301839.
Truvé, S., 2016. Temporal analytics for predictive cyber threat intelligence. In: Proceedings of the 25th International Conference Companion on World Wide Web, pp. 867–868.
Tundis, A., Ruppert, S., Mühlhäuser, M., 2020. On the automated assessment of open-source cyber threat intelligence sources. In: Computational Science–ICCS 2020: 20th International Conference, Amsterdam, the Netherlands, June 3–5, 2020, Proceedings, Part II 20. Springer, pp. 453–467.
Vouros, G.A., 2022. Explainable deep reinforcement learning: state of the art and challenges. ACM Comput. Surv. 55 (5), 1–39.
Wagner, T.D., Palomar, E., Mahbub, K., Abdallah, A.E., 2018a. Towards an anonymity supported platform for shared cyber threat intelligence. In: Risks and Security of Internet and Systems: 12th International Conference. CRiSIS 2017, Dinard, France, September 19-21, 2017. In: Revised Selected Papers, vol. 12. Springer, pp. 175–183.
Wagner, T.D., Palomar, E., Mahbub, K., Abdallah, A.E., 2018b. A Novel Trust Taxonomy for Shared Cyber Threat Intelligence. Security and Communication Networks, vol. 2018.
Wang, A., An, N., Xia, Y., Li, L., Chen, G., 2014. A logistic regression and artificial neural network-based approach for chronic disease prediction: a case study of hypertension. In: 2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom). IEEE, pp. 45–52.
Wang, M., Yang, L., Lou, W., 2022a. A comprehensive dynamic quality assessment method for cyber threat intelligence. In: 2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W). IEEE, pp. 178–181.
Wang, T., Chow, K.P., 2019. Automatic tagging of cyber threat intelligence unstructured data using semantics extraction. In: 2019 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, pp. 197–199.
Wang, X., He, S., Xiong, Z., Wei, X., Jiang, Z., Chen, S., Jiang, J., 2022b. APTNER: a specific dataset for ner missions in cyber threat intelligence field. In: 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, pp. 1233–1238.
Wang, X., Liu, R., Yang, J., Chen, R., Ling, Z., Yang, P., Zhang, K., 2022c. Cyber threat intelligence entity extraction based on deep learning and field knowledge engineering. In: 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, pp. 406–413.
Wheelus, C., Bou-Harb, E., Zhu, X., 2016. Towards a big data architecture for facilitating cyber threat intelligence. In: 2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS). IEEE, pp. 1–5.
Wickramasinghe, C.S., Amarasinghe, K., Marino, D.L., Rieger, C., Manic, M., 2021. Explainable unsupervised machine learning for cyber-physical systems. IEEE Access 9, 824–131 843.
Williams, R., Samtani, S., Patton, M., Chen, H., 2018. Incremental hacker forum exploit collection and classification for proactive cyber threat intelligence: an exploratory study. In: 2018 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, pp. 94–99.
Xu, Y., Yang, Y., He, Y., 2020. A representation of business oriented cyber threat intelligence and the objects assembly. In: 2020 10th International Conference on Information Science and Technology (ICIST). IEEE, pp. 105–113.
Yazdinejad, A., Zolfaghari, B., Azmoodeh, A., Dehghantanha, A., Karimipour, H., Fraser, E., Green, A.G., Russell, C., Duncan, E., 2021. A review on security of smart farming and precision agriculture: security aspects, attacks, threats and countermeasures. Appl. Sci. 11 (16), 7518.
Yeboah-Ofori, A., Islam, S., Yeboah-Boateng, E., 2019. Cyber threat intelligence for improving cyber supply chain security. In: 2019 International Conference on Cyber Security and Internet of Things (ICSIoT). IEEE, pp. 28–33.
Yoo, S., Lee, T., 2023. A study of the ordinal scale classification algorithm for cyber threat intelligence based on deception technology. Electronics 12 (11), 2474.
Yu, Z., Wang, J., Tang, B., Lu, L., 2022. Tactics and techniques classification in cyber threat intelligence. Computer J., bxac048.
Zhang, H., Shen, G., Guo, C., Cui, Y., Jiang, C., 2021a. Ex-action: automatically extracting threat actions from cyber threat intelligence report based on multimodal learning. Secur. Commun. Netw. 2021, 1–12.
Zhang, N., Ebrahimi, M., Li, W., Chen, H., 2022b. Counteracting dark web text-based captcha with generative adversarial learning for proactive cyber threat intelligence. ACM Trans. Manag. Inf. Syst. 13 (2), 1–21.
Zhang, P., Ya, J., Liu, T., Shi, J., 2021b. Mining open-source cyber threat intelligence with distant supervision from the web. In: 2021 IEEE Sixth International Conference on Data Science in Cyberspace (DSC). IEEE, pp. 76–82.
Zhang, S., Li, S., Chen, P., Wang, S., Zhao, C., 2022a. Generating network security defense strategy based on cyber threat intelligence knowledge graph. In: International Conference on Emerging Networking Architecture and Technologies. Springer, pp. 507–519.
Zhao, J., Yan, Q., Li, J., Shao, M., He, Z., Li, B., 2020a. Timiner: automatically extracting and analyzing categorized cyber threat intelligence from social data. Comput. Secur. 95, 101867.
Zhao, J., Yan, Q., Liu, X., Li, B., Zuo, G., 2020b. Cyber threat intelligence modeling based on heterogeneous graph convolutional network. In: 23rd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2020), pp. 241–256.
Zhao, L., Li, J., Li, Q., Li, F., 2021. A federated learning framework for detecting false data injection attacks in solar farms. IEEE Trans. Power Electron. 37 (3), 2496–2501.
Zhou, Y., Tang, Y., Yi, M., Xi, C., Lu, H., 2022. CTI view: APT threat intelligence analysis system. Secur. Commun. Netw. 2022, 1–15.
Zuo, J., Gao, Y., Li, X., Yuan, J., 2022. An end-to-end entity and relation joint extraction model for cyber threat intelligence. In: 2022 7th International Conference on Big Data Analytics (ICBDA). IEEE, pp. 204–209.

Dr Hang Thanh Bui is a researcher at Charles Sturt University, School of Computing, Mathematics and Engineering. She has 5 years of experience working on industrial projects in the application of digitalisation and cutting-edge technology. Her research interest is in cyber security, and the application of machine learning in interdisciplinary, distributed computing.

Dr Hamed Aboutorab is a cybersecurity research fellow at Charles Sturt University. He completed his PhD at the University of New South Wales. His research is centred around the Integration of cybersecurity and artificial intelligence, with a current focus on security automation and orchestration. His research has been published in prestigious international journals, including IEEE Transactions on Services Computing, Expert Systems with Applications, Journal of Network and Computer Applications, and Future Generation Computer Systems.

Dr Arash Mahboubi is a deputy leader of the Cyber Security Research Group at Charles Sturt University. He is an esteemed expert in the realm of information security, having earned his PhD from Queensland University of Technology. With a deep-rooted passion for cybersecurity, Dr. Mahboubi's research extensively delves into areas such as the interplay of artificial intelligence and machine learning with cybersecurity, secure automation in cyber environments, and adaptive defence strategies against intricate cyber-attacks.

Dr Yansong Gao is a Research Scientist at CSIRO's Data61. He received his M.Sc degree from the University of Electronic Science and Technology of China in 2013 and a Ph.D. degree from the University of Adelaide, Australia, in 2017. His work has appeared in prestigious conferences (such as NDSS, ACSAC, DSN, and ACM AsiaCCS), as well as journals (such as Nature Electronics, IEEE Transactions on Dependable and Secure Computing, and IEEE Transactions on Information Forensics and Security). His current research interests are AI security and privacy, system security, and hardware security.

Dr Nazatual Sultan is a Research Scientist at CSIRO Data61 and a member of the Distributed System Security (DSS) group. Prior to CSIRO, he was working as a Senior Consultant with one of the Big 4 consulting firms in the Security Architecture Team in building cyber resilience critical infrastructures. He worked as a postdoctoral researcher for 3 years combined at the University of Newcastle & CSIRO Data61 in Australia and Telecom SudParis in France. He has published his papers in top venues like ESORICS, SRDS, IEEE TCC, IEEE TSC.

Dr Aufeef Chauhan is a researcher at The University of Adelaide, focuses on Autonomous Systems, Robots, Autonomous Decision Making, Knowledge Sharing, and Cyber Security Data Orchestration. His work extends to software systems' architecture for Distributed and Cloud Computing, covering areas like IoT and the Confidentiality-Integrity-Availability (CIA) of systems. Combining industry and academic experience, he has collaborated with institutions like the Technical University of Denmark (DTU) and IT University of Copenhagen (ITU), contributing to financial application development. He received a Degree in Computer Science from IT University of Copenhagen, Denmark.

Dr Mohammad Zavid Parvez earned his PhD in Computer Science from Charles Sturt University, Australia, in May 2016, specializing in "Epileptic Seizure Detection and Prediction by Analyzing EEG Signals". Since 2007, his research spans Biomedical Engineering, AI, Data Science, BCI, and Software Engineering. Contributing to high-impact journals and conferences, he has 37 publications, including IEEE Transactions, Neurocomputing, and Chaos, Solitons and Fractals. Proficient in MATLAB and Python, Dr. Parvez

conducted postdoctoral research at the ISI Foundation in Italy, focusing on AI, machine learning, and medical signal/image processing.

Dr Michael Bewong is a Senior Lecturer in Computing specializing in Data Science, AI, and Cyber Security at Charles Sturt University, is dedicated to applying his expertise in data analytics to tackle real-world issues collaboratively. Previously a research fellow at the Data to Decisions Cooperative Research Centre (D2D CRC), he participated in projects like "Beat the News," creating data mining models to predict socially disruptive events. Additionally, his work on "Predicting Cyber Security Exploits" involved developing innovative machine learning techniques to address challenges in forecasting cyber vulnerabilities and exploits.

Dr Rafiqul Islam is an Associate Professor at Charles Sturt University, Australia, specializes in Cybersecurity. His research focuses on malware analysis, authentication, cloud security, and IoT privacy. Leading the Cybersecurity research team, Dr. Islam has a robust leadership and collaborative research record. With 180+ peer-reviewed papers, he received national and international acclaim, winning the 2021 Cyber Security Researcher of the Year Award. Acknowledged through various awards, he stands at the forefront of national and international cybersecurity research priorities. Dr. Islam is a co-recipient of over 10 external grants, exceeding $5 million, contributing significantly to successful projects within the Cybersecurity CRC.

Dr Zahid Islam is a Professor of Computer Science in Charles Sturt University's School of Computing, Mathematics, and Engineering, Australia. Currently, he holds the position of Associate Dean (Research) in the Faculty of Business, Justice, and Behavioural Sciences since September 2023. From January 2019 to September 2023, he directed the Data Science and Engineering Research Unit (DSERU) and served as Theme Lead and CSU Academic Lead for the Cyber Security CRC from January 2021 to September 2023. His extensive research interests span Data Mining, Classification algorithms, Privacy Preserving Data Mining, Cyber Security, and real-life applications.

Dr Seyit Camtepe is a principal research scientist at CSIRO's Data61 leading the Autonomous Security and Software Security team. He is passionate about discovering unusual solutions to challenging cybersecurity problems with a specific focus on pervasive security. He was among the first to inform society about Android malware outbreak, and to realise the model-to-data paradigm in computing to enable research on data in captivity. From 2007 to 2013, he was with the TU-Berlin, Germany, as a senior researcher and research group leader in security. Dr Seyit worked for five years as an ECARD lecturer at the QUT, Australia.

Dr Praveen Gauravaram is a Senior Scientist at Tata Consultancy Services Ltd. (TCS), ANZ. Leading TCS's Co-Innovation (COIN) Research and Development Partnership with Cyber Security CRC, he actively contributes to national cyber security development, playing a key role in initiatives like the Australian Cyber Security Strategy 2020 and Security Legislation Amendment (Critical Infrastructure Protection) Act 2022. Recognized with the Young Elite Researcher Award in 2010, he holds honorary academic titles, including Adjunct Associate Professor at the University of New South Wales. With over 50 research publications, he advises on cyber security and serves as an ICT Curriculum Advisory Group Advisor at Southern Cross University.

Dinesh Singh is an interdisciplinary researcher specializing in the impact of digital technologies on agriculture, water, fisheries, and livestock. His expertise extends to cross-domain research, encompassing hypothesis formulation, experiment design, pilot setup, data collection, analysis, and results interpretation. Dinesh excels in orchestrating a seamless progression from research endeavors to the practical application, guiding the process towards productization. His commitment to understanding the intricate intersections of various domains and translating research findings into tangible outcomes underscores his dedication to driving innovation and sustainable solutions in agriculture and related sectors.

Dr M. Ali Babar is a Professor of Software Engineering at the University of Adelaide, Australia, also holds an Associate Professorship at IT University of Copenhagen, Denmark. Previously, he served as a Reader in Software Engineering at Lancaster University, UK. With extensive experience as a researcher and project leader in Ireland and Australia, he has authored/co-authored over 140 peer-reviewed research papers and co-edited a book on software architecture knowledge management. Prof. Ali Babar has guest-edited special issues for various journals and served on program committees for international conferences. Notably, he was the founding general chair of Nordic-Baltic Symposium on Cloud Computing and Internet Technologies (NordiCloud) in 2012.

Dr Shihao Yan received his Ph.D degree in Electrical Engineering from the University of New South Wales (UNSW), Sydney, Australia, in 2015. His impactful research secured $480k funding, including prestigious fellowships. With 2 book chapters, 65 journal articles, 38 conference papers, and 8 patents, he serves as an Editor and chaired the Technical Program Committee for IEEE GlobeCOM Workshop in 2018. Recognized for expertise, he was a Panel Member for IEEE Vehicular Technology Conference, Local Arrangements Chair for the Australian Communications Theory Workshop, and received certificates for top 2% Exemplary Reviewers from IEEE Transactions on Communications in 2017 and 2019.