A Review of Hybrid Cyber Threats Modelling and Detection Using Artificial Intelligence in IIoT

Yifan Liu
Cardiff University

Shancang Li
Xi'an Jiaotong University

Xinheng Wang
Xi'an Jiaotong University

Li Xu
Old Dominion University, lxu@odu.edu

Original Publication Citation

Liu, Y. F., Li, S. C., Wang, X. H., & Xu, L. (2024). A review of hybrid cyber threats modelling and detection using artificial intelligence in IIoT. *Computer Modeling in Engineering & Sciences*, *140*(2), 1233-1261. https://doi.org/10.32604/cmes.2024.046473

Computer Modeling in
Engineering & Sciences

Tech Science Press

DOI: 10.32604/cmes.2024.046473

REVIEW

A Review of Hybrid Cyber Threats Modelling and Detection Using Artificial Intelligence in IIoT

Yifan Liu[1], **Shancang Li**[1,*], **Xinheng Wang**[2] **and Li Xu**[3]

[1]School of Computer Science and Informatics, Cardiff University, Cardiff, CF24 4AG, UK

[2]Department of Mechatronics and Robotics, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China

[3]Department of Information Technology & Decision Sciences, Old Dominion University, Norfolk, 23529, USA

*Corresponding Author: Shancang Li. Email: shancang.li@ieee.org

Received: 02 October 2023 Accepted: 21 December 2023 Published: 20 May 2024

ABSTRACT

The Industrial Internet of Things (IIoT) has brought numerous benefits, such as *improved efficiency*, *smart analytics*, and *increased automation*. However, it also exposes connected devices, users, applications, and data generated to cyber security threats that need to be addressed. This work investigates hybrid cyber threats (HCTs), which are now working on an entirely new level with the increasingly adopted IIoT. This work focuses on emerging methods to model, detect, and defend against hybrid cyber attacks using machine learning (ML) techniques. Specifically, a novel ML-based HCT modelling and analysis framework was proposed, in which L_1 regularisation and Random Forest were used to cluster features and analyse the importance and impact of each feature in both individual threats and HCTs. A grey relation analysis-based model was employed to construct the correlation between IIoT components and different threats.

KEYWORDS

Cyber security; Industrial Internet of Things; artificial intelligence; machine learning algorithms; hybrid cyber threats

1 Introduction

The Internet of Things (IoT) refers to a network of interconnected objects using wired and wireless communication schemes [1]. The Industrial Internet of Things (IIoT) is a paradigm of IoT application in industry. It combines emerging technologies such as 5G, big data, cloud computing, and Artificial Intelligence (AI), enabling industrial devices, control systems, and other production components to exchange and analyse massive amounts of data. The information-driven IIoT promotes the digitisation and automation of industries, significantly boosting the efficiency and economy of production. It has been increasingly adopted by all industrial sectors, especially manufacturing, healthcare, and transportation [2]. While the IIoT considerably contributes to industrial production, it raises a number of challenges in terms of cyber security concerns [3].

Heterogeneity and interconnectivity are critical features in an IIoT system that enable different components to communicate with each other effectively. However, the attack surfaces are broadened, and the security capabilities of heterogeneous hardware and software in an IIoT system have assorted levels. Attackers may utilise vulnerable objects to control IIoT devices and launch further malicious activities. For instance, the Mirai botnet was constructed by exploiting firmware vulnerabilities. It was used to conduct large-scale Distributed Danial of Service (DDoS) attacks [4]. Insecure network protocols could lead to lethal ramifications. For example, attackers could use Modbus to gain unauthorized access [5]. Furthermore, different industrial scenarios may have various resources and security requirements. Existing authentication, access control, and other security measures may be unsatisfactory in IIoT environment due to constrained resources [6–8], which makes IIoT system protection more challenging against evolving threats.

Besides the challenges introduced by IIoT, current adversaries tend to be more skilled, increasing the landscape of cyber threats [9]. Attackers increasingly combine different attack vectors and techniques, such as *social engineering, malware, network intrusion*, and *Denial of Service (DoS) attacks*, to compromise targets and achieve their goals [9]. Furthermore, experienced hackers usually develop various strategies which involve multiple stages to evade detection. For instance, a series of designed stealth attacks can serve as the basis for further malicious activities. Low-aggressive attacks may bypass conventional IDS, thus obscuring the complete attack path and raising the complexity of investigation and forensics [10]. Especially within the IIoT environment, interconnected components could be scaffolding for attackers, enabling more covert and versatile attack chains. HCTs arise from this intricate circumstance, which requires a more comprehensive approach to defend IIoT systems.

AI-enabled methods have shown a promising capability for data analysis, and they offer an avenue for studies to explore IIoT security solutions from various perspectives. A sheer of AI-based threat detection and attack analysis methods have been proposed to secure IIoT systems and achieved impressive performance, such as email filtering, code analysis, vulnerability scanning, intrusion detection, and attack path analysis. However, novel attack detection, false alarm mitigation and persuasive security incident correlation are still challenging tasks [4]. To adapt threat detection for complicated circumstances of HCTs in IIoT, combining multiple security countermeasures from both technical and temporal dimensions could be essential. The technical dimension refers to different cybersecurity measures, such as Intrusion Detection Systems (IDSs), threat intelligence, attack path modelling and alert correlation analysis. The temporal dimension refers to the related stages of these different security measures. This survey reviews the application of AI methods in various security schemes and discusses their combination scheme by constructing a technical framework. Section 2 reviews surveys related to IIoT security. Section 3 briefly introduced the IIoT structure, security requirements and HCTs. Section 4 investigates the widely used AI methods in threat detection. Section 5 discusses the AI-based schemes, which could be employed to assemble multiple methods to analyse HCTs. The HCT detection framework is discussed in this section. Section 6 presents the conclusion and gives several challenges and potential future works in IIoT HCT detection.

This survey aims to investigate AI-enabled security schemes and their practical solutions for IIoT. This paper also introduces a technical framework for HCT detection, which can be a reference for future research. The main contributions are summarised as:

1) This work proposed an HCT analysis framework in the IIoT context by combining the characteristics of IIoT systems to address emerging HCTs.

2) Recent advances in AI-based security methods and their application in IIoT are reviewed in this paper.

3) Threat modelling methods, L_1 *regularisation, random forest (RF)*, and *grey relational analysis (GRA)* are introduced in detail and integrated with promising technologies to construct a framework for HCT detection.

4) Challenges and research trends of HCT analysis in IIoT were discussed in this paper.

2 Related Works

Many surveys have discussed IIoT security issues from the viewpoint of system structure. Xu et al. [11] proposed a panoramic view of IIoT, in which the service-oriented architecture (SOA) and critical embedding technologies, application sectors and common issues of IIoT were introduced. Dhirani et al. [12] reviewed general security standards and communication protocols of IoT in industry, and current challenges were presented. There is also a sheer of surveys concentrated on security issues in IoT. Jayalaxmi et al. [13] gave an overview of the security framework studies while Panchal et al. [14], and Tsiknas et al. [3] analysed various threats and countermeasures in detail from the perspective of different layers.

Some other surveys discussed IIoT security from a view of threats. Gao et al. [15] discussed the difference between IoT and IIoT, provided an extensive investigation into cyber security from the CIA triad perspective and refined the security requirement and challenges in the IIoT context. Moreover, this work introduced the potential applications of fog computing in IIoT security. Shah et al. [16] gave a taxonomy of the IIoT device based on usage scenarios and listed potential security issues and requirements according to the category of devices. The authors discussed countermeasures of several common attacks as well. Similarly, Sengupta et al. [17] analysed specific threats of different IIoT components and provided some case studies in real scenarios. Furthermore, this survey comprehensively investigated blockchain applications on IIoT security. Makhdoom et al. [18] thoroughly investigated IoT threats and proposed a security framework for IoT.

Some surveys focus on specific security problems in IIoT. Jiang et al. [19] discussed the application of differential privacy for IIoT in detail. They claimed that the extensive utilization of big data technology in IIoT may introduce new challenges to privacy protection. Vishwakarma et al. [20] conducted an exhaustive investigation on DDoS attacks and countermeasures in IoT networks. Kim et al. [21] emphasized the insider threat of IoT in their survey, in which they conducted a comprehensive investigation on internal attacks and internal threat mitigation solutions.

Rakas et al. [22] focused on the security issues in supervisory control and data acquisition systems and network-based IDS. Knowles et al. [23] discussed the Industrial Control System (ICS) security standard and management solutions. Al-Mhiqani et al. [24] proposed a taxonomy of threats in IIoT, which categorizes threats from four perspectives, i.e., attack method, impact, intention and event type. Yaacoub et al. [25] highlighted the security issues of cyber-physical systems.

There is some research discussing IoT or IIoT security from a methodology perspective. Al-Garadi et al. [4] studied the ML and Deep Learning (DL) based security solutions for IoT. Similarly, Hussain et al. [1] systematically investigated general IoT issues and ML-based mitigation schemes. Tatam et al. [26] investigated the threat modelling methods for sophisticated attacks. Navarro et al. [27] and Kotenko et al. [28] focused on the correlation analysis approaches, given a comprehensive survey on multi-step attack detection methods.

While there are plenty of surveys in the IIoT threat detection area, they often provide summaries of various methods corresponding to the different problems with the IIoT environment, lacking an integrative perspective. Compared to other surveys, this survey identified the IIoT security requirements and HCTs characteristics. AI-based methods and their application in IIoT security have been investigated. In addition, a technical framework is proposed to converge various methodological dimensions onto the main issue, namely, HCT detection in IIoT.

3 IIoT and HCTs

An IIoT system usually involves a large number of heterogeneous and interconnected objects that collect, process, and transmit sensitive data. It could be divided into different layers according to diverse functions and technologies of its components, including *perception layer, network layer*, and *application layer*. Attackers can exploit vulnerabilities in one or more of these devices to gain unauthorized access to sensitive data, compromise system integrity, and damage critical infrastructure [29].

An IIoT system may have different layers according to the diverse function of components, including *perception layer, network layer*, and *application layer*, which highlight the technical features of IIoT, e.g., the extensibility, interoperability and robustness [11]. *Perception Layer*. The perception layer is a crucial component of IIoT systems directly associated with the production site, encompassing various sensors for collecting environmental data. The perception layer typically also includes actuators, such as LEDs and buzzers that can be used to indicate equipment status, and mechatronic devices that can execute linear, rotational, or even more complex movements [30]. *Network Layer*. In IIoT systems, numerous heterogeneous devices and various application software are interconnected through the network layer, which plays a crucial role in IIoT systems [31]. In recent years, with the development of wireless communication technology, wireless devices have received increasing attention, and low-power wireless communication protocols have gradually become popular in IIoT, such as *Bluetooth, LPWAN, ZigBee,* etc. *Processing Layer*. The processing layer is also called middle-ware layer in some literature, which primarily consists of services and applications designed to provide an efficient collaborative platform for complex IIoT components. One of its significant functions is to provide communication standards, such as APIs and protocols, allowing heterogeneous devices and software to exchange data seamlessly [4,11]. Therefore, access management for service objects should be considered at this layer. *Application Layer*. The application layer generally faces the user directly, and the solutions differ based on industry and even the business sector. This layer typically includes a large number of sensor logs, machine operation data and business process information [32]. Since the carrier of the application layer is often a conventional computing device, the security challenges faced could be more complex.

3.1 Security Requirements in IIoT

The diversity of IIoT components brings a complicated context for cyber security [33], which consists of various *devices, firmware, communication protocols, control software*, and *various software interfaces* together with expended attack surfaces [34]. The single vulnerable object may expose the system to cyber hybrid threats such as *Man in the middle, DoS, malicious code injection*, etc. Because of the interconnected nature of IIoT, once an attacker successfully accesses from an inevitable attack surface, it could be liable to move horizontally in the network to expand the attack range and even obtain advanced privileges. In addition, there are usually some constraints in the IIoT environment, for example, device energy consumption, low-latency communication, and data privacy. Moreover, with the development of edge computing and blockchain technology, decentralization has become a future trend, which will inevitably increase the complexity of IIoT security circumstances.

A security requirement model may prove advantageous in issue analysis and strategy design. For example, the CIA triad is an extensively used threat modelling method which defines the fundamental elements of information security: confidentiality, integrity and availability. Satisfying confidentiality means preventing data from unauthorized access and disclosure. Integrity emphasizes authenticity and non-repudiation of information. Availability is an attribute that evaluates the timeliness and availability of data. However, when discussing security in the context of IIoT environments, it is customary to meticulously consider security requirements, which are precisely described as follows:

Availability: It refers to the ability of a system or application to remain operational and accessible to authorized users at all times. For example, DoS attacks are common means of compromising system availability [20]. Since the IIoT system usually contains resource-constrained components, it could be more vulnerable to this attack [35]. As downtime or disruption to services may cause numerous risks for IIoT, such as production delays, safety hazards, and financial losses, availability could be a vital property of the IIoT system.

Authentication: It could be a critical process of verifying the identity of a user, device, or software attempting to access an IIoT system or network. In an industrial setting, it is necessary to consider data security and ensure Quality of Service (QoS). For instance, low latency and limited computing resources [36]. Furthermore, the mobility, scalability and heterogeneity of IIoT objects make it more challenging to recognise suspicious devices [37]. It has been demonstrated that malicious devices could impose serious consequences, and a robust authentication mechanism should be established for the IIoT system.

Confidentiality: Data-driven could be one of the major characteristics of the IIoT system. A vast amount of data will be produced and transmitted between the IIoT devices and software [29]. Extremely sensitive information such as access key, commercial data, and device status may be included in different processes, and compromised confidentiality may neutralize access control and lead to severe ramifications.

Integrity: It is one of the critical pillars of IIoT security. Integrity ensures that data transmitted between IIoT devices and software remains unaltered. Compromised data may result in process failures, system shutdowns, and even security issues. Thus, reliable strategies should be introduced to discern unexpected modifications in the IIoT system.

Non-Repudiation: It guarantees a secure and dependable transmission of information by furnishing the intended recipient with proof of delivery and verifying the identity of the sender to the data source. Non-repudiation is not typically considered a crucial security feature for IIoT systems. Nevertheless, in specific contexts, such as payment systems, it plays a critical role in ensuring that both parties involved in a transaction cannot later deny their involvement in the payment [4].

3.2 HCTs in IIoT

Hybrid threats were originally coined to refer to adversarial activities that pose a national threat by employing various methods [38]. In the context of cyber security, the IIoT is encountering hybrid threats to a certain extent. Plenty of research efforts have claimed that industrial targets are increasingly attracting hackers, as summarised in Table 1. Attack surfaces in IIoT could be exploited by attackers as entry points to infiltrate the network using hybrid methods to achieve their ultimate goal [37,39,40], such as stealing data, hijacking, or destroying industrial facilities. Makhdoom et al. [18] indicated that 70% of devices accessing the internet are prone to various attacks. Navarro et al. [27] discussed multi-step and single attacks and noted that single-step attack, such as a DoS attack, should not be confounded with a multi-step attack because there are usually no associated actions subsequent to the launch of the former. HCTs could be more applicable to describe the emerging sophisticated threats in IIoT.

Table 1: Cyber hybrid threats/attacks in IIoT

Phases	Cyber hybrid threats	Associated layers	Security property
Reconnaissance	1) Generic Scanning, the initial step of attack that listens to the ports to collect information such as the *Operation system, available services*, and *software version,* etc.; 2) Vulnerability Scanning, using automated tools to detect vulnerabilities according to discovered information and Common Vulnerabilities and Exposures (CVE); 3) Fuzzing, which refers to injecting a payload into target software to discover potential useful information by analyzing the response messages; 4) Discovering Resources, which could be conducted when misconfigurations exist in the target machine, e.g., *unrestricted file download* and *remote code execution*.	Perception layer Network layer Processing layer Application layer	Authentication Confidentiality
Weaponization	1) Brute-force Attack, which refers to the enumerating of simple user names and passwords to gain unauthorized access; 2) Dictionary-attack, which aims to crack the target authentication mechanism using a dictionary which may contain real accounts and passwords; 3) Malicious insiders, which could be any employee who is disgruntled or defrauded by social engineering attacks who usually have access to the facility, which means that they can conduct malicious activity stealthily and easily evade abnormal detection.	Perception layer Processing layer Application layer	Authentication Confidentiality

(Continued)

Table 1 (continued)

Phases	Cyber hybrid threats	Associated layers	Security property
Exploitation	1) Reverse shell, which is a technology that allows the target host to connect to and receive commands from the adversarial machine; 2) Man in the middle attack, an attacker may establish a fraud node between two interconnected devices to intercept their communication.	Perception layer Network layer Processing layer Application layer	Authentication Confidentiality Integrity
Lateral movement	Attacks aim to find more critical information about the target IIoT system to expand threat scales. It usually utilises communication protocol vulnerabilities such as MQTT, Modbus and TCP.	Network layer	Authentication
Command & control	This step could be important for APTs to take control of the target system.	Perception layer Application layer	Authentication Confidentiality
Exfiltration	It refers to abnormal data extraction activities conducted by attackers.	Processing layer Application layer	Authentication Confidentiality Integrity
Tampering	Attackers may compromise system configuration, event logs and other sensitive data to conceal their activities.	Perception layer Processing layer Application layer	Integrity
DoS	DoS is a broad concept and one of the most prevalent forms of cyber attack. This kind of attack is usually conducted by overloading the target machine or software to disable normal access to the service. Ransomware could be considered a kind of DoS attack to some extent.	Application layer	Availability Confidentiality

One typical example of HCT is APT, which is usually carried out by adequately financed hacker teams employed by consortium establishments or governing bodies that aim to obtain vital intelligence of their target or to damage the infrastructures of adversaries [41]. In [20], Vishwakarma et al. claimed that DDoS attack has begun to attempt to achieve their goal by amalgamating multiple layers of attacks. Ma et al. [42] proposed a stealthy attack method which is based on the vulnerability of redundant controllers in industrial CPS.

Al-Hawawreh et al. [43] proposed a comprehensive dataset which includes various data sources, scenarios, and hybrid attack methods in the IIoT context. The author defined a taxonomy of attacks according to the target of certain malicious activities. Typical HCTs faced by IIoT systems include: (1) *Botnets*. Botnets are networks of infected devices that can be controlled by attackers to carry out a range of malicious activities, including distributed denial-of-service (DDoS) attacks, spamming, and data theft. IoT devices are particularly vulnerable to botnet attacks, as they often have limited security features and are connected to the internet. (2) *Malware*. Malware is a common threat to IoT systems, as it can be used to compromise device security, steal sensitive data, or carry out other malicious activities. Malware can be introduced through a variety of vectors, including phishing emails, infected websites, and vulnerabilities in device firmware. (3) *Physical attacks*. IoT devices can be physically vulnerable to attacks, such as tampering, theft, or destruction. Physical attacks can compromise the security and privacy of the data being processed by the device and can cause disruptions to critical infrastructure. (4) *Insider threats*. Insider threats involve individuals with authorized access to IoT devices or systems using that access to carry out malicious activities, such as stealing sensitive data or introducing malware. Insider threats can be particularly difficult to detect and prevent, as the individuals involved often have legitimate access to the system. (5) *Social engineering*. Social engineering attacks involve attackers using psychological manipulation to trick users into revealing sensitive information or performing actions that compromise system security. Social engineering attacks can be used to gain unauthorized access to IoT devices or systems or to trick users into downloading malware or giving away sensitive information.

Miller et al. [44] provide a review of cyber attacks against ICS over several decades. It has been pointed out that threat actors have become increasingly organized and large-scale in recent years. Social engineering, such as phishing, has become a primary means of initiating attacks to gain unauthorized access to victim systems. Subsequent actions mainly rely on exploiting existing vulnerabilities. To address these HCTs, ICS systems require various security measures, including strong encryption and authentication mechanisms, regular updates and patches, and appropriate access controls.

4 Machine Learning Algorithms for HCTs Detection

Machine learning (ML) is a data-driven technology that allows computers to train mathematical models with immense data to learn from previous information to resolve specific tasks [45]. According to the given task and training approach, ML algorithms could be divided into three types: *supervised, unsupervised*, and *reinforcement learning* [46]. *Supervised learning* refers to building a mathematical model to learn the correspondence between given features and labels of training examples to predict the output of new samples. In contrast, *unsupervised learning algorithms* do not have prior knowledge of training data. Machine learning algorithms usually need to cluster samples in the dataset automatically. Different from other algorithms, *reinforcement learning (RL)* achieves goals by interacting with a virtual environment to formulate optimal policies, usually referring to a set of states and corresponding actions. To formulate optimal policies, RL models use a trial and error strategy to maximize the reward within the specified time step [47].

Deep learning (DL) is a subset of ML that has developed tremendously in the last decade [48]. DL algorithms are generally structured as multiple-layer networks with a series of computing units, which are more capable of automatically abstracting features from high dimensional data [47,49]. In shallow ML, which is also mentioned as traditional ML, feature selection could be a sophisticated manual step that needs professional knowledge to select highly correlated factors from a dataset and may significantly impact models' performance. However, the advantage of deep learning depends on the enormous scale of data. Shallow learning algorithms could still be competitive in some scenarios that lack available data [4]. This section will briefly introduce common shallow and deep ML algorithms and discuss their application in cyber security.

4.1 Shallow Learning

Shallow learning contains various *supervised, unsupervised* and *reinforcement learning algorithms* that have been widely implemented in cyber threat detection in early research [50].

4.1.1 Supervised Algorithms

Decision tree (DT): It is a simple algorithm widely used for categorizing tasks. It uses flowcharts with tree architecture to conduct decision analysis, which is intuitive and easy to implement. DT generally contains root vertex, inner nodes and leaves. The decision process of a sample starts at the root vertex representing its feature set. Then, the internal nodes assert the next branch according to current feature values and finally route to a leaf node, which denotes a potential category [51]. Various outstanding algorithms evolved from DT, e.g., the *RF [52], Extra-trees (ET) [53], eXtreme Gradient Boosting (XGB) [54]* and *Light Gradient Boosting Machine (LightGBM) [55].*

Kasongo et al. [51] investigated several tree-based algorithms and proposed an IDS combined RF with Genetic Algorithm (GA) for IIoT, in which GA was used to optimize the RF structure and conduct feature engineering. Five models (i.e., *LR, DT, ET, XGB, RF*) were evaluated on the UNSW-NB15 dataset. GA generated ten and seven groups of features separately for binary and multiclass classification tasks. RF achieved the highest accuracy (i.e., *95.91%* and *87%* on average) for the binary classification tasks in the validation and test sets. In comparison, ET models could be more competitive for multi-categorization. Their average accuracy reached *82.74%* and *76.76%.*

Logistic regression (LR): LR is a statistical-based algorithm commonly used in various fields, including cyber security. The LR model calculates the probability of the occurrence of a particular result based on the given variables. The output is a value between 0 and 1, which is then converted to a binary outcome using a threshold which can be adjusted to control the trade-off between the accuracy and false positive rate. Besharati et al. [56] proposed a HIDS called LR-HIDS, which is based on logistic regression to identify threats in cloud scenarios that are usually running in a virtual environment.

Naive Bayesian (NB): NB is a kind of probabilistic model which is based on Bayes' theorem. It has been used for network intrusion detection in early research [57,58]. However, NB may fail to detect the potential clue among features [4], and it takes less advantage than other methods in the big data context [59]. In recent studies, it has been implemented as an assistant method. Gu et al. [60] proposed an IDS framework using NB for feature embedding to improve the quality of a dataset and then using SVM to recognize abnormal samples. They claimed that data quality could be a critical factor that may considerably impact the performance of ML-based IDS models, while it was rarely considered in existing studies. The result shows that SVM models trained on the fixed dataset achieved better accuracy.

Support vector machine (SVM): SVM classifies samples by constructing a hyperplane within the feature vector space. It performs effectively in both binary and multi-class scenarios while maintaining a low level of computational complexity. It has been widely employed in cybersecurity area, such as intrusion detection [61–63] and malware analysis [64,65].

4.1.2 Unsupervised Algorithms

***K*-means:** *K*-means is a simple unsupervised clustering algorithm based on Euclidean distance. It is primarily utilized for classifying or grouping objects into *K*-distinct groups based on their properties. *K* refers to an integer number that is required to be predetermined for the clustering algorithm. Li et al. utilized *K*-means to detect Sybil attacks in industrial wireless sensor networks [66]. Chang et al. proposed [67] a method using *K*-means and Autoencoder to identify anomalies for industrial control system.

4.1.3 Reinforcement Learning

Q-Learning: *Q*-Learning is an *RL* algorithm that is model-free and able to learn by interaction with the environment using an agent. It provides an optimum strategy for any finite Markov decision process (FMDP) by maximising the anticipated amount of the entire reward across consecutive steps. Given boundless exploring and a stochastic policy, Q-Learning can find the most effective route for any FMDP. The working principle of Q-Learning is shown in Fig. 1, in which A_n denotes the *n*th Action, S_n denotes the *n*th State, respectively. For action A_n in state S_n, there is a certain reward R_n, and the agent will calculate and update the Q value $Q(S_n, A_n)$ of each set of state and action in every step until the iteration is completed. Xiao et al. proposed to use *Q*-Learning for access control in wireless networks [48]. Li et al. proposed a method for DoS attack based on *Q*-Learning [68].

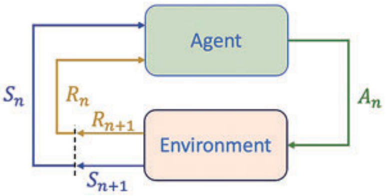


Figure 1: Q-Learning working principle

4.2 Deep Learning Algorithms

4.2.1 Supervised Algorithms

Convolutional neural network (CNN): CNN is developed to extract information from pixel matrix. It is commonly used in computer vision. CNN usually consists of an input layer, several hidden layers, and an output layer. The hidden layers of CNN contain at least one convolutional layer, and the pooling layer and fully connected layer are also included in general. CNN can reduce raw sample dimension and reserve key information through convolution operation, making it a powerful feature extraction tool. Chawa et al. [69] proposed a host-based intrusion detection method which used CNN to abstract the call sequence in a time window. Li et al. proposed [70] a fusion model using CNN to detect network intrusion of IIoT.

Recurrent neural network (RNN): RNN is a kind of neural network that has the ability to remember short-term sequential information. Fig. 2 presents key processes. Gated recurrent unit (GRU) is a commonly used RNN algorithm. RNN units learn the context information by considering the former input vector. For given sequential data $X = \{x_1, \ldots, x_n\}$, where x_n is orderly samples, *n* refers to order number. There will be two outputs for each iteration, $Y = \{y_1, \ldots, y_n\}$ and $C = \{c_1, \ldots, c_n\}$, where y_n is the output of the model, c_n is the saved information that will be forwarded to next calculation, the subscript of GRU means the iteration times. RNN and its variants such as GRU and LSTM has been successfully employed in network intrusion detection and exhibited favourable performance [71–73].

Because RNN have the capability to analyze positional information of sequential data, some research used it to identify acivities of advanced persistent threat [74,75].

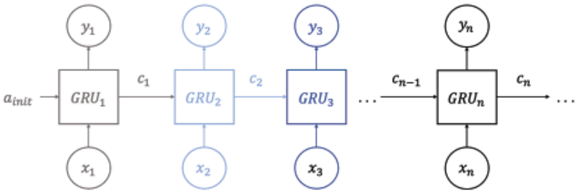


Figure 2: RNN working principle

Transformer. Similar to RNN, transformer model is proposed to analyse sequential samples. It has achieved remarkable success in natural language processing (NLP) problems. The transformer involves several sets of encoders and decoders, and its key concept is attention mechanism. The attention mechanism allows the model to analyse the contribution of a single sample in a given input sequence for the current target output unit of decoders. Fig. 3 shows an example of transformer. Firstly, the input vectors $x_1, x_2, \ldots, x_6$ were encoded to attach positional information. Then encoder extracted a similarity matrix as output which records the context information of the input sequence.

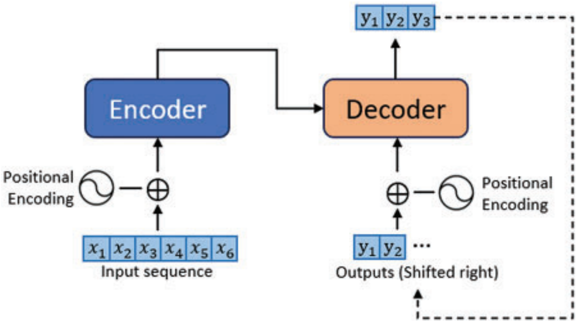


Figure 3: Transformer working principle

The decoder received output from the encoder and generated y_1 according to the context information. Then y_2 was generated using y_1 and input context as input of the decoder and attached after y_1. Further steps are similar until the output sequence is generated completely. Because of its powerful function to analyse long-term sequence data, it has been used to reveal the potential pattern among network traffic. Ho et al. [76] converted network flow to image sequences and proposed a transformer-based IDS which achieved 98.5% and 96.3% accuracy in binary classification on the UNSW-15 and CICIDS2017 datasets.

4.2.2 Unsupervised

Autoencoder (AE). AE is a kind of neural network which can extract efficient representations from raw data through an unsupervised training process to disregard noise information. As shown in Fig. 4, AE consists of an encoder and a decoder, both components include a hidden layer. After training, it can encode the input vector, and then attempt to decode and output the code. It is commonly used for dimension reduction. Lopez-Martin et al. [77] applied AE to recover the missing information of incomplete samples.

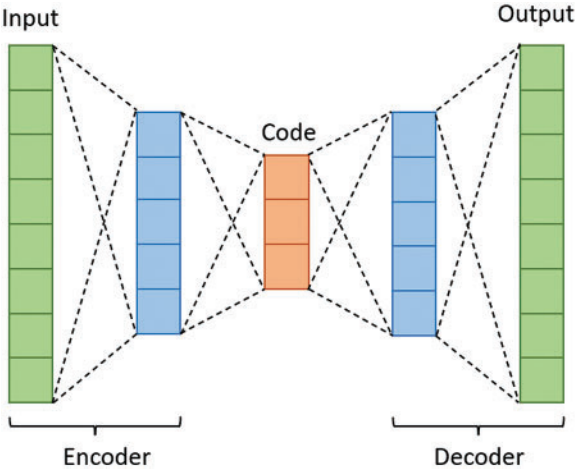


Figure 4: Autoencoder

Generative adversarial network (GAN). GANs are generative models based on game theory. The model consists of a generator and a discriminator. The generator is responsible for creating samples according to a given initial vector, while the discriminator is previously trained to distinguish between real and generated samples. The fundamental concept of its training process is to train the generator under the supervision of a discriminator, which is also a neural network that can update itself. The structure of GAN is shown in Fig. 5. Because of the nature of GAN model, it has been used to generate samples, which is effective in protecting privacy in sensitive scenarios [78], and functional for the problem of insufficient samples [79]. It could be a powerful method to improve the robustness of ML models and has been used to mitigate poisoning attacks in machine learning [80,81].

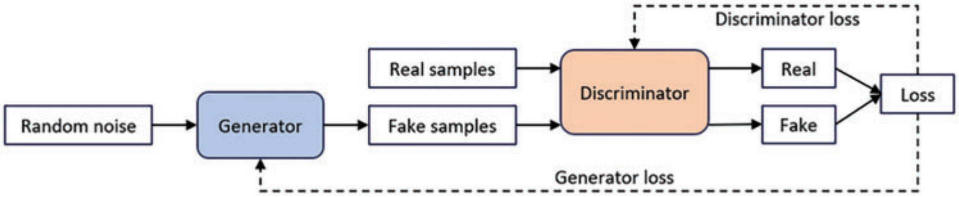


Figure 5: GAN working principle

4.2.3 Reinforcement Learning

Deep Q network (DQN). Since Q-learning is not able to calculate the most optimized action when states and actions are incredibly variable, DQN, a combination of Q-learning and DL, is developed to solve this problem. The concept of DQN is that using a function $F(Q(S_n, A_n), \omega)$ to represent $Q(S_n, A_n)$, where a neural network will calculate ω. Sethi et al. [82] proposed a DQN-based NIDS as shown in Fig. 6, in which network traffic features were abstracted to environment states, and attack types were regarded as action.

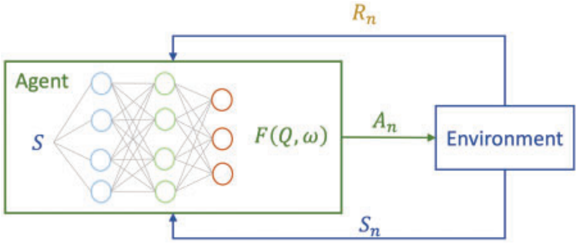


Figure 6: DQN working principle

5 HCT Modelling Schemes and Framework

Threat modelling could serve as a reference for threat detection and attack correlation to provide explainability [28] and improve accuracy [27]. However, little research has considered the integration of threat modelling and detection. This section investigated potential methods that could be used to establish an HCT model that could be integrated with threat detection. An integrating framework is proposed, and promising HCT detection methods are discussed according to the framework.

5.1 HCT Modelling

HCT modelling aims to provide the visibility of HCTs in IIoT by evaluating and abstracting the characteristics and potential impact of threat vectors. Existing threat modelling can be categorized into two types: manual modelling and mathematical modelling. The manual modelling schemes are usually presented as graphs, data flows, or tables. The mathematical modelling approaches are based on probability models, such as Markov chains and their variants [26]. However, the complexity of IIoT systems limits the practical application of these modelling schemes [26]. To utilize an HCT model in attack detection, two components could be involved: a threat model and a system environment model [83]. The threat model aims to describe the features of each threat vector. The system environment model refers to an abstraction of IIoT system status, such as software versions and existing vulnerabilities.

5.1.1 RF-Based HCT Modelling

Machine learning techniques, such as Lasso regularisation [84] and RF [85], are able to remove features with low variance and can be used in feature selection in HCT dependencies analysis. An HCT is usually a combination of multiple individual threats. Using clustering-based regularisation along with decision trees (DTs), the dependencies and features of individual threats could be identified for further attack event analysis. For the regression model, the residual sum of squares can be used as splitting criteria, as shown in Eq. (1).

$$\mathrm{RSS} = \sum_{\text{left}} \left(y_i - y_L^*\right)^2 + \sum_{\text{right}} \left(y_i - y_R^*\right)^2 \tag{1}$$

in which y_L^* denotes y value for the left node, y_R^* denotes y value for the right node.

DT is widely used in feature selection tasks because of its nature. It can be created using given features $\{x_1, x_2, \ldots, x_N\}$, entropy, and labelled classes. The RF consists of a collection of DTs, each of which is built using a subset of training data and randomly selected features. In the classification, *Gini* criterion is used as

$$Gini = N^L \sum_{k=1}^{K} p_k^L \left(1 - p_k^L\right) + N^R \sum_{k=1}^{K} p_k^R \left(1 - p_k^R\right) \tag{2}$$

in which p_k^L is the proportion of class k at left node, and p_k^R is the proportion of class k at right node.

Individual HCT vector usually has different contributions to an attack event. The importance of an individual threat T_m could be evaluated by impurity importance [86].

$$\sum_{m=1}^{p} \text{Imp}\left(T_m\right) = I\left(f_1, \ldots, f_p; A\right) \tag{3}$$

in which the f_i, $(i = 1, \ldots, p)$ are the features of HCT, and A is a potential hybrid attack of HCT.

Wali et al. [87] proposed an IDS based on RF, which can explain the correlation degree between attack and features.

5.1.2 Grey Relational Analysis (GRA) Based HCT Modelling

GRA is a modelling technique used to evaluate the correlation between variables or factors. It is insensitive to the quantity of samples, has no specific requirements on the data distribution pattern, and has a relatively small computational load.

Environment modelling. Existing threat analysis methods focus on the connection between object features and attacks at a macro level. Attack classification models are often established by analyzing the overall pattern of data samples under different states. For instance, common approaches are general machine learning models using network traffic, system calls, or physical signals to establish classifiers to detect diverse individual attacks. However, the network system structures are becoming increasingly complex, and the proliferation of heterogeneous hardware and software may render the identification of abnormal states more challenging. Furthermore, threat actors lean toward employing hybrid attack methods to achieve their objectives and ambiguous the clues to disrupt investigations. For example, attackers may utilize adversarial attacks to generate attack vectors to evade IDS detection, where the real attack chain can be concealed. Moreover, attacks may have distinct patterns in different network objects, potentially introducing disruptions to attack modelling. To adapt the HCT, this paper proposes a novel threat modelling concept by introducing HCT context modelling, considering the inherent status of different components (Including hardware, software and communication protocols) as the context of the attack model within the threat analysis process. There are two main steps to generate the model.

Data aggregation. Given a dataset $D = \{d_1, \ldots, d_h\}$, where h is the number of samples. $d_h = (f_h(1), \ldots, f_h(k))$, k is the number of features. Aggregate the data according to the type of device, the environment matrix $E = (e_1, \ldots, e_n)$, where n denotes the number of object types in the entire system. The device element vector $e_n = (v_n(1), \ldots, v_n(k))$, where v_n represents the attributes value of the device n, it can be the mean value of the device attributes in their normal state.

Given an attack set $A = (a_1, \ldots, a_m)$, in which m denotes the number of attack samples. The attack element vector $a_m = (v_m(k), \ldots, v_m(k))$, where v_m represents the average attributes value of the attack type m. To quantify the degree of differentiation between attacks and normal states across different targets, The initial context matrix can be defined by Eq. (4). Where s_{mn} refers to the probability of device e_n being attacked by attack a_m, S can be generated using Algorithm 1.

Algorithm 1: Generate matrix S

```
    Input $A$, $E$;
    Output $S$;
$i \leftarrow 1, j \leftarrow 1, q \leftarrow 1$
while $i \leq m$ do
    while $j \leq n$ do
        $s_{ij} \leftarrow 0$
        while $q \leq k$ do
            $s_{ij} \leftarrow s_{ij} + |a_i(k) - e_j(k)|$
        end while
    end while
end while
```

$$S = \begin{pmatrix} s_{11} & \dots & s_{1n} \\ \dots & \dots & \dots \\ s_{m1} & \dots & s_{mn} \end{pmatrix} \tag{4}$$

Because S may contain heterogeneous data sources, normalization is required before GRA with Eqs. (5) and (6), as

$$s_i(j) = \frac{s_i(j)}{\bar{s}_i} \tag{5}$$

$$\bar{s}_i = \frac{\sum_{j=1}^{n} s_i(j)}{n} \tag{6}$$

Normalization is used to eliminate the scale differences between different feature values, assuring that each feature contributes equally to the model and preventing certain features from retaining an excessive impact or being neglected.

Grey relational analysis (GRA). GRA is an effective method for correlation analysis, and it has achieved good performance in various areas. The GRA is defined as Eq. (7).

$$\zeta_i(j) = \frac{F_{min} + \rho \cdot F_{max}}{|s_0(j) - s_i(j)| + \rho \cdot F_{max}} \tag{7}$$

$$F_{min} = min_i min_j |s_0(j) - s_i(j)| \tag{8}$$

$$F_{max} = max_i max_j |s_0(j) - s_i(j)| \tag{9}$$

where $s_0(j) = (s_0(j), \dots, s_0(j))$ is a selected reference feature vector, which could be the max value of feature k selected from S. $s_i(j) = (s_i(1), \dots, s_i(j)), i = 1, 2, \dots, n$ are the alternative feature vector. The GRA $\zeta_i(j)$ indicates the distinguishability of attack i on device j. ρ is resolution factor, the smaller ρ means greater resolution. Generally $\rho = 0.5$. $\zeta_i(j)$ indicates the distinguishability of an attack on different components. The proposed model provides an environment abstraction for HCTs, describing the correlation of each attack and IIoT object using a matrix. It enables a more effective assessment of ambiguous attack nodes during the security information and event management (SIEM) phase. The Fig. 7 presents an example of HCT context modelling. The directed graph refers to a threat chain. $t_i, i = 1, 2, 3$ in Fig. 7 correspond to $Threat_i, i = 1, 2, 3$ in Table 2 which indicates the correlated degree of a threat with a device. $t_i(j), i = 1, 2, 3; j = 1, 2, \dots, 5$. Refers to $Threat_i$ occurs on device

D_j, $t_i(j,k), i = 1,2,3; j = 1,2,\ldots,5; k = 1,2,\ldots,5$. Refers to $Threat_i$ moves between Devices D_j and D_k.

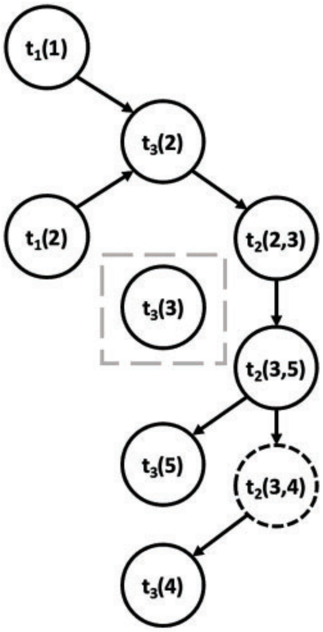


Figure 7: Graph-based model for HCT context

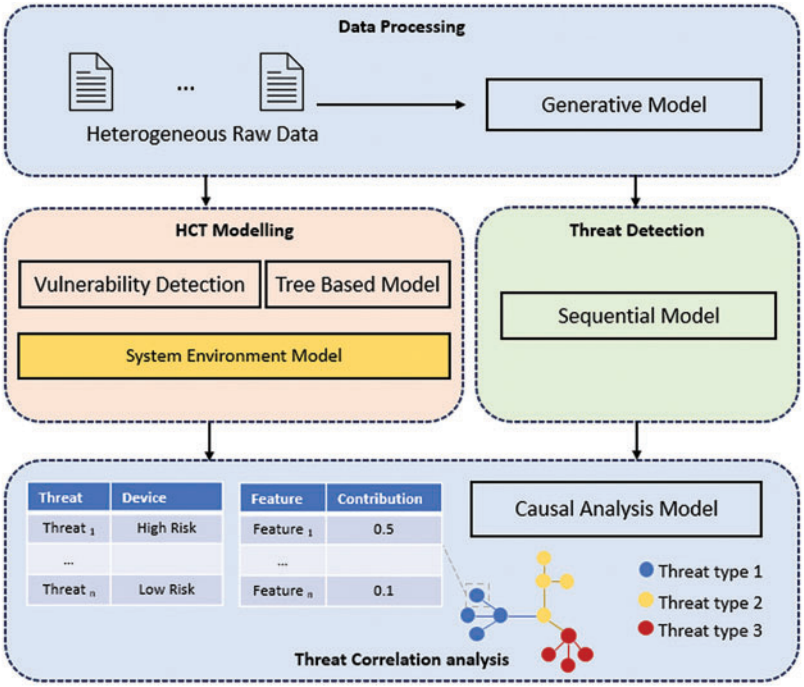
Table 2: HCT context modelling example

	D_1	D_2	D_3	D_4	D_5
$Threat_1$	High	Hig	Medium	Medium	Low
$Threat_2$	Medium	Medium	High	High	Low
$Threat_3$	Low	High	Low	High	Low

E.g., in Table 2, device D_1 has a high correlation with threat $Threat_1$, and a low correlation with threat $Threat_3$. The threat node in Fig. 7 indicates the threat type , source, and destination. In Fig. 7, threat node $t_3(2)$ indicates device D_2 is under $Threat_3$, threat node $t_2(3,5)$ refers $Threat_2$ moves from device D_3 to device D_5. The dotted-line nodes represent nodes added based on the HCT context model. For a given real threat chain $(t_2(3,5), t_2(3,4), t_3(4))$, node $t_2(3,4)$ is not detected by intrusion detection while $t_3(4)$ need a former node. Since $Threat_2$ shows a high correlation with device D_2 and device D_3, and its former node $t_2(3,5)$ and forward node $t_3(4)$ has been detected, the missing node $t_2(3,4)$ could be recovered by HCT environment model. Node $t_3(3)$ surrounded by grey rectangles is suggested to be a false positive because it is not involved in a threat chain and has a low correlation with device D_3 in the HCT environment model.

5.2 HCTs Detection Framework

Detecting HCTs in IoT requires a comprehensive and multi-faceted approach that involves a combination of techniques, including machine learning. As shown in Fig. 8, the framework consists of four main components: data processing, vulnerability analysis for HCTs, HCTs sequence detection and attack correlation analysis.



Threat	Device
Threat $_1$	High Risk
...	
Threat $_n$	Low Risk

Feature	Contribution
Feature $_1$	0.5
...	
Feature $_n$	0.1

Figure 8: HCTs detection and classification

5.2.1 Data Processing

The heterogeneity of IIoT determines that the generated data may be in various formats, numerical scales, and quantities. There are several methods for handling heterogeneous data. The regular approach is to extract heterogeneous data to a standardized format manually. Al-Hawawreh et al. [43] proposed a pre-processed dataset called X-IIoTID, which is represented as a Comma-separated values file with features extracted from network flow and system logs. They correlated data from different sources based on device and time. In addition to manually extracting features from multi-source data, some studies have employed ensemble learning methods to evaluate the extracted data for optimizing feature selection [88,89]. However, manual feature extraction requires interdisciplinary expertise and incurs high labour costs.

Some other research uses anomaly alerts from different components as inputs to the detection model. Namely, a series of anomaly detection models are created firstly for different devices or systems, and then attacks are further detected based on the information contained in the anomaly alerts [90,91]. Although alert-based schemes automate the feature selection, they highly rely on the accuracy of anomaly detection methods, which may limit their performance. Park et al. [92] proposed a generative model-based intrusion detection scheme that can abstract data from multiple sources and extract features automatically. Since generative models can abstract different information and generate data, mitigating the data imbalance issue, they could be promising solutions for IIoT data processing.

5.2.2 Vulnerability and Malware Analysis for HCTs

Vulnerability detection is a vital part of cyber security strategies, typically aimed at identifying potential risks within a system. Additionally, it can be utilized to improve intrusion detection [93]. Morin et al. [94] introduced a framework that combines IDS and vulnerability detection, effectively alleviating false alarms. Furthermore, it can be integrated with IDS to correlate alerts with detected vulnerabilities, providing interpretability for alarms from IDS, therefore automating the information security control process [95].

Mokhov et al. indicated [96] that NLP technologies are promising for vulnerability detection task. In previous research, a commonly used approach is to convert source code into abstract syntax tree and then employ deep learning methods [97,98], such as RNN, LSTM, CNN and their variants [99,100] to analyze semantic information. In recent years, with the rise of large-scale language models, an increasing number of researchers are attempting to apply them to the field of security. Ziems et al. developed [101] a vulnerability dataset for C programming language, and designed several deep learning models. The best model achieved over 93% accuracy. They claimed that maintaining contextual information is important for vulnerability hunting. Thapa et al. considered [102] the similarity between high-level programming languages and natural languages and evaluated the performance of large-scale language models based on transformers for vulnerability detection task. The results indicate that language models have better performance than conditional NLP methods such as LSTM.

HaddadPajouh et al. [103] proposed a malware detection method based on instruction machine code analysis using RNN. They collected malware and benign software designed for a 32-bit ARM processor and then decompiled the application to machine code sequences in each sample, which were then used to build a TF-IDF bag-of-words model. A bidirectional LSTM model was used to analyse the machine code sequence of the sample to obtain a malicious code detector. Their method reached 98% accuracy in detecting new malware. The malware detection or malicious code identification task was regarded as a sequence-to-one NLP problem. RNN has the ability to analyse the semantics of a piece of code according to the context of the machine code sequence so that it can effectively classify whether the sample contains malicious code. However, the dataset used in this work is relatively small, with less than 300 positive and negative samples, which could be insufficient in the real software environment. In practice, imbalanced samples are a persistent problem in current datasets for threat detection tasks because abnormal data is relatively difficult to obtain [104].

Nguyen et al. mentioned that the dynamic analysis can effectively monitor the network behaviour and system calls of malware [105]. Using the PSI-graph [105], a hybrid detection method was proposed to detect the botnet in the IIoT. The PSI-graph can extract the information related to the attack phase in the binary file based on the life-cycle of the botnet, such as *IP address, usernames, passwords*, and *malicious payload*. Since PSI-graph only focuses on specific steps in the botnet, it can significantly reduce computational complexity compared to conventional function call graphs. Different from [105,106], it adds the information obtained by dynamic analysis into PSI-graphs. The graphs were converted into feature vectors to train the ML model (DF, RF, SVM, KNN and bagging). The classifiers were evaluated on ARM and MIPS datasets, and the results showed that the hybrid method is superior to conventional methods in terms of time consumption and detection accuracy.

5.2.3 Automated HCT Modelling

An appropriate modelling scheme can integrate system status, vulnerability analysis and threat intelligence to enhance the efficiency and accuracy of attack detection [107]. Wang et al. proposed [108] a method based on CVSS score to construct the dependency relationships between exploits.

Existing modelling approaches usually require a manual process, and additional steps are required to adopt a threat model to attack analysis [26]. Some research has implemented automation at specific stages of modelling, and their combination may hold promise. Firstly, identified vulnerabilities could be regarded as features of IIoT components. Secondly, $L1$ regularization could be adapted to select proper features for GRA-based environment model, which can be employed to model the relevance degree between different devices and attacks [109]. Sun et al. proposed [110] a Bayesian based method to predict zero-day threat path by analysing system calls.

5.2.4 HCTs Sequence Detection

Sequence models take advantage of contextual awareness and have shown tremendous potential for threat detection in recent years. They can capture more potential patterns from continuous data, enabling better threat detection performance [111]. RNN is a classic sequence model that has been used for threat detection [69]. However, RNN and its variants are extremely time-consuming and computationally consuming models. The requirement for high-performance processors and stable power supply are usually not feasible in IoT devices. Wani et al. [112] developed a Software-defined networking (SDN)-based IDS for resource-constrained environments.

Li et al. [113] proposed an improved LSTM model, quasi-recurrent neural network (QRNN), to mitigate the time complexity. the proposed model combines the advantages of CNN and LSTM to preserve timing dependencies in parallel. Not only is it 16 times faster to train and test than ordinary LSTM, but it also has better accuracy. Al-Taleb et al. [114] proposed a hybrid neural network which combined QRNN and CNN to identify cyber threats in smart city environments. The one-dimensional convolutional layer and the one-dimensional pooling layer in this model are used to reduce the dimensionality of the features, and the QRNN is used to save the time series state. Furthermore, a dropout layer is added to prevent overfitting. This model achieves 99.99% accuracy on both the BoT-IoT and TONIoT datasets, and the FPR is 0.3% and 1%, respectively. In addition, the author compares the time taken by the model to use LSTM and QRNN. On the BoT-IoT dataset, using QRNN is about 23% faster, and on the TONIoT dataset, LSTM is 23% slower.

The CNN model could also be sequence-sensitive when structured as one dimension. It has relatively low computational complexity and is often used in combination with other models [115,116]. Kale et al. [117] proposed a framework that contains three algorithms working together to detect anomalies and specify the type of attack in network traffic. There are three main steps in this framework. Firstly, the K-means algorithm is used to identify significantly abnormal traffic samples. Then, those samples classified as benign will be filtered again by GANomaly to further select anomalies. Lastly, in the final stage, all detected abnormal samples would be classified by a trained CNN to distinguish different types of attacks. The methods used in the first two stages are based on unsupervised learning. CNN is trained using the labelled dataset to recognize the kinds of attacks. The performance of this framework is evaluated on NSL-KDD, CIC-IDS2018, and TONIoT. According to the shown result, the AUC scores of this proposed method in the first two stages on three datasets are 91.6%, 70.3%, and 91.8%. The Log-loss scores for CNN are 17.87%, 13.88%, and 1.3%.

Transformer-based models take advantage of their capability for parallel computing over long-term sequence input, achieving dramatic performance in anomaly detection tasks. Casaju-Setien et al. [118] proposed a simplified transformer model which uses single-head self-attention to learn the normal network flow pattern in a time window. The accuracy of this model reached 97.44% over the WUSTL IIoT dataset. However, anomaly detection methods based on time series are typically restricted to situations where the number of abnormal samples is much less than normal. Once anomalies occur in a dense manner, their performance may be adversely impacted. Xiao et al. [119] proposed a diffusion-based scheme to mitigate model bias when anomaly concentration arises. This method allows the model to choose samples flexibly based on changes in anomaly density. This method demonstrates superior stability and performance across five datasets. It achieved 94.19%, 97.66%, 97.95%, 96.95% and 92.75% in F_1 score over five datasets (MSL, SWaT, PSM, SMAP, SMD).

5.2.5 Attack Correlation Analysis

Effectively analyzing the correlations between attacks can identify the trajectory of attackers, enabling the timely discovery of potential threats [120]. Establishing a comprehensive view of attacks helps security teams understand new attack trends to prevent future attacks proactively. Moreover, accurately tracing the cause and effect of attacks can also reduce false positives [121].

Khosravi et al. [122] employed the Security Information Event and Management system to analyse causal relationships among detected attacks. The proposed scheme achieved 87.10% recall in their experiment. Jadidi et al. [120] proposed a causal anomaly detection method for the ICS. It discovers the correlation of malicious activities by analyzing ICS logs. This method reached 98% accuracy on attack causal diagnosis. Kumar et al. [123] proposed an advanced threat detection method for IIoT called RAPTOR, which is designed to recognize different attack phases and generate an attack graph by analyzing IDS alerts. Hu et al. [124] proposed a Markov chain-based attack correlation mining model, which integrates information from IDS, firewalls and other network components and automates the IDS alert correlation process. Ding et al. [40] developed an attribute correlation-based algorithm to recognize the hybrid threat in IoT systems. The core of the proposed approach is the grey correlation analysis algorithm, which is an effective theory for analyzing the impact of unknown factors on the system. The proposed work can successfully identify 92.3% of correlated attack routes in maximum.

6 Challenges and Future Directions

6.1 Challenges

According to a review of current research on threat modelling, detection and complex attack analysis, there have been notable advancements in IIoT security in the last decades. However, some challenges remain.

- **Increasing attack surfaces** could raise challenges for IIoT security. An HCT involves multiple cyber attack vectors, each of which may have diverse behavioural characteristics based on different attack surfaces and their targets. This circumstance poses challenges to threat detection and alert analysis, especially in IIoT environments where a multitude of attack surfaces may obscure attack chains. A significant amount of time and human resources for evidence collection, correlation, and source tracing to investigate and manage security events.
- **Dynamic and evolving attack tactics** could be a challenge from adversaries. Actors of HCTs usually adapt their tactics to evade detection. Kumari et al. [125] proposed a community deception method to prevent a node from being recognized by community detection algorithms. Detecting novel combinations of known attack vectors is very challenging, which may impact the attack correlation analysis. Moreover, HCTs involve insider involvement or insider threats, where authorised users or employees intentionally or unintentionally facilitate the attack. Compared to external adversaries, internal threat recognition requires a great deal of effort.
- **Lack of comprehensive data sharing**, effective detection of combined threats requires sharing threat intelligence and data across different organizations, sectors, and regions. However, there are challenges in sharing sensitive information due to legal, privacy, and competitive concerns,

limiting the collective ability to detect and respond to combined threats. In addition, the lack of unified hybrid attack vector signatures may pose challenges to HCT detection.

6.2 Future Directions

Addressing these challenges requires close collaboration between security organisations and communities. The future research directions in HCT detection include:

- **Explainable machine learning** techniques are being extensively explored to detect HCTs, which involves advanced detection algorithms to analyse large volumes of data, feature extraction approaches, etc. Using interpretable models to correlate and analyse system components, potential threats, and attack alerts may be a direction for the future. It can reveal attack surfaces to support security event investigations and benefit false positive mitigation to some extent.
- **Context-aware behaviour analysis**, HCTs often involve a series of distinct behaviours, such as reconnaissance and exploiting vulnerabilities. Behaviour, including application behaviour, user activity, and the interplay between different components, will play a key role in detecting suspicious activities and identifying HCTs. Context-aware behaviour analysis will detect and correlate events related to HCTs by considering the relationships, dependencies, and contextual information of different features.
- **HCTs modelling and simulation** involves feature extraction, attack pattern analysis and system environment emulation. An appropriate modelling scheme could effectively abstract complex HCT vectors from limited data.

It is important to develop innovative techniques to effectively detect and respond to the HCTs in the rapidly evolving IIoT.

7 Conclusion

HCT detection in IIoT systems is a critical and complex problem that requires integrated consideration. There are many promising methods which take advantage of the powerful data analysis capability of ML and can be used to protect the security of IIoT.

This survey discusses HCTs in IIoT from an integrated view. Firstly, his paper introduced the characteristics and security requirements of the IIoT system and discussed the emerging HCTs. Secondly, HCT modelling schemes are introduced in detail from threat and system environment aspects. A framework is proposed to integrate the HCT model and HCT detection methods. It involves data processing, HCT modelling, threat detection, and threat correlation analysis. The promising technologies are introduced based on the components of this framework, providing a handbook for future IIoT security solution development. Specifically, a test use case was introduced to demonstrate the effectiveness of the proposed HCT framework. Finally, this paper explains the challenges and provides several potential future directions for IIoT security.

Acknowledgement: The authors would like to acknowledge the anonymous reviewers and journal editors, their valuable comments significantly improved the quality of this paper.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Study conception and design: Yifan Liu, Shancang Li; Methodology: Yifan Liu, Shancang Li; Original draft manuscript

preparation: Yifan Liu, Shancang Li; Supervision: Shancang Li; Writing, review and editing: Xinheng Wang, Li Xu. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Hussain, F., Hussain, R., Hassan, S. A., Hossain, E. (2020). Machine learning in IoT security: Current solutions and future challenges. *IEEE Communications Surveys & Tutorials, 22(3),* 1686–1721. https://doi.org/10.1109/COMST.9739
2. Huo, R., Zeng, S., Wang, Z., Shang, J., Chen, W. et al. (2022). A comprehensive survey on blockchain in industrial internet of things: Motivations, research progresses, and future challenges. *IEEE Communications Surveys & Tutorials, 24(1),* 88–122. https://doi.org/10.1109/COMST.2022.3141490
3. Tsiknas, K., Taketzis, D., Demertzis, K., Skianis, C. (2021). Cyber threats to industrial IoT: A survey on attacks and countermeasures. *IoT, 2(1),* 163–186. https://doi.org/10.3390/iot2010009
4. Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I. et al. (2020). A survey of machine and deep learning methods for internet of things (IoT) security. *IEEE Communications Surveys & Tutorials, 22(3),* 1646–1685. https://doi.org/10.1109/COMST.9739
5. Sinha, A., Patel, S. S., Kumar, A., Vyas, O. (2021). Exploiting vulnerabilities in the scada modbus protocol: An ICT-reliant perspective. *International Conference on Advanced Network Technologies and Intelligent Computing*, pp. 94–108. Cham, Springer International Publishing.
6. Gao, H., Wang, X., Wei, W., Al-Dulaimi, A., Xu, Y. (2023). Com-DDPG: Task offloading based on multiagent reinforcement learning for information-communication-enhanced mobile edge computing in the internet of vehicles. *IEEE Transactions on Vehicular Technology, 73(1),* 348–361.
7. Mishra, N., Pandya, S. (2021). Internet of things applications, security challenges, attacks, intrusion detection, and future visions: A systematic review. *IEEE Access, 9,* 59353–59377. https://doi.org/10.1109/ACCESS.2021.3073408
8. Gao, H., Fang, D., Xiao, J., Hussain, W., Kim, J. Y. (2022). Camrl: A joint method of channel attention and multidimensional regression loss for 3D object detection in automated vehicles. *IEEE Transactions on Intelligent Transportation Systems, 24(8),* 8831–8845.
9. Chng, S., Lu, H. Y., Kumar, A., Yau, D. (2022). Hacker types, motivations and strategies: A comprehensive framework. *Computers in Human Behavior Reports, 5,* 100167. https://doi.org/10.1016/j.chbr.2022.100167
10. Mao, B., Liu, J., Lai, Y., Sun, M. (2021). MIF: A multi-step attack scenario reconstruction and attack chains extraction method based on multi-information fusion. *Computer Networks, 198,* 108340. https://doi.org/10.1016/j.comnet.2021.108340
11. Xu, L. D., He, W., Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics, 10(4),* 2233–2243. https://doi.org/10.1109/TII.2014.2300753
12. Dhirani, L. L., Armstrong, E., Newe, T. (2021). Industrial IoT, cyber threats, and standards landscape: Evaluation and roadmap. *Sensors, 21(11),* 3901. https://doi.org/10.3390/s21113901
13. Jayalaxmi, P., Saha, R., Kumar, G., Kumar, N., Kim, T. H. (2021). A taxonomy of security issues in industrial internet-of-things: Scoping review for existing solutions, future implications, and research challenges. *IEEE Access, 9,* 25344–25359. https://doi.org/10.1109/Access.6287639
14. Panchal, A. C., Khadse, V. M., Mahalle, P. N. (2018). Security issues in IIoT: A comprehensive survey of attacks on IIoT and its countermeasures. *2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN)*, pp. 124–130. Lonavala, India, IEEE.

15. Gao, H., Huang, J., Tao, Y., Hussain, W., Huang, Y. (2022). The joint method of triple attention and novel loss function for entity relation extraction in small data-driven computational social systems. *IEEE Transactions on Computational Social Systems, 9(6),* 1725–1735. https://doi.org/10.1109/TCSS.2022.3178416
16. Shah, Y., Sengupta, S. (2020). A survey on classification of cyber-attacks on IoT and IIoT devices. *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 0406–0413. New York City, New York, USA, IEEE.
17. Sengupta, J., Ruj, S., Bit, S. D. (2020). A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT. *Journal of Network and Computer Applications, 149,* 102481. https://doi.org/10.1016/j.jnca.2019.102481
18. Makhdoom, I., Abolhasan, M., Lipman, J., Liu, R. P., Ni, W. (2019). Anatomy of threats to the internet of things. *IEEE Communications Surveys & Tutorials, 21(2),* 1636–1675. https://doi.org/10.1109/COMST.9739
19. Jiang, B., Li, J., Yue, G., Song, H. (2021). Differential privacy for industrial internet of things: Opportunities, applications, and challenges. *IEEE Internet of Things Journal, 8(13),* 10430–10451. https://doi.org/10.1109/JIOT.2021.3057419
20. Vishwakarma, R., Jain, A. K. (2020). A survey of ddos attacking techniques and defence mechanisms in the iot network. *Telecommunication Systems, 73(1),* 3–25. https://doi.org/10.1007/s11235-019-00599-z
21. Kim, A., Oh, J., Ryu, J., Lee, K. (2020). A review of insider threat detection approaches with IoT perspective. *IEEE Access, 8,* 78847–78867. https://doi.org/10.1109/Access.6287639
22. Rakas, S. V. B., Stojanović, M. D., Marković-Petrović, J. D. (2020). A review of research work on network-based scada intrusion detection systems. *IEEE Access, 8,* 93083–93108. https://doi.org/10.1109/Access.6287639
23. Knowles, W., Prince, D., Hutchison, D., Disso, J. F. P., Jones, K. (2015). A survey of cyber security management in industrial control systems. *International Journal of Critical Infrastructure Protection, 9,* 52–80. https://doi.org/10.1016/j.ijcip.2015.02.002
24. Al-Mhiqani, M. N., Ahmad, R., Yassin, W., Hassan, A., Abidin, Z. Z. et al. (2018). Cyber-security incidents: A review cases in cyber-physical systems. *International Journal of Advanced Computer Science and Applications, 9(1),* 499–508.
25. Yaacoub, J. P. A., Salman, O., Noura, H. N., Kaaniche, N., Chehab, A. et al. (2020). Cyber-physical systems security: Limitations, issues and future trends. *Microprocessors and Microsystems, 77,* 103201. https://doi.org/10.1016/j.micpro.2020.103201
26. Tatam, M., Shanmugam, B., Azam, S., Kannoorpatti, K. (2021). A review of threat modelling approaches for APT-style attacks. *Heliyon, 7(1),* e05969. https://doi.org/10.1016/j.heliyon.2021.e05969
27. Navarro, J., Deruyver, A., Parrend, P. (2018). A systematic survey on multi-step attack detection. *Computers & Security, 76,* 214–249. https://doi.org/10.1016/j.cose.2018.03.001
28. Kotenko, I., Gaifulina, D., Zelichenok, I. (2022). Systematic literature review of security event correlation methods. *IEEE Access, 10,* 43387–43420. https://doi.org/10.1109/ACCESS.2022.3168976
29. Serror, M., Hack, S., Henze, M., Schuba, M., Wehrle, K. (2021). Challenges and opportunities in securing the industrial internet of things. *IEEE Transactions on Industrial Informatics, 17(5),* 2985–2996. https://doi.org/10.1109/TII.9424
30. Latif, S., Driss, M., Boulila, W., Jamal, S. S., Idrees, Z. et al. (2021). Deep learning for the industrial internet of things (IIoT): A comprehensive survey of techniques, implementation frameworks, potential applications, and future directions. *Sensors, 21(22),* 7518. https://doi.org/10.3390/s21227518
31. Khan, W. Z., Rehman, M., Zangoti, H. M., Afzal, M. K., Armi, N. et al. (2020). Industrial internet of things: Recent advances, enabling technologies and open challenges. *Computers & Electrical Engineering, 81,* 106522. https://doi.org/10.1016/j.compeleceng.2019.106522

32. Chen, B., Wan, J., Shu, L., Li, P., Mukherjee, M. et al. (2018). Smart factory of Industry 4.0: Key technologies, application case, and challenges. *IEEE Access, 6,* 6505–6519. https://doi.org/10.1109/ACCESS.2017.2783682

33. Khraisat, A., Gondal, I., Vamplew, P., Kamruzzaman, J., Alazab, A. (2019). A novel ensemble of hybrid intrusion detection system for detecting internet of things attacks. *Electronics, 8(11),* 1210. https://doi.org/10.3390/electronics8111210

34. Berger, S., Bürger, O., Röglinger, M. (2020). Attacks on the industrial internet of things–development of a multi-layer taxonomy. *Computers & Security, 93,* 101790. https://doi.org/10.1016/j.cose.2020.101790

35. Abomhara, M., Køien, G. M. (2015). Cyber security and the internet of things: Vulnerabilities, threats, intruders and attacks. *Journal of Cyber Security and Mobility, 4,* 65–88. https://doi.org/10.13052/jcsm2245-1439.414

36. Loske, M., Rothe, L., Gertler, D. G. (2019). Context-aware authentication: State-of-the-art evaluation and adaption to the IIoT. *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pp. 64–69. Limerick, Ireland, IEEE.

37. Meidan, Y., Bohadana, M., Shabtai, A., Ochoa, M., Tippenhauer, N. O. et al. (2017). Detection of unauthorized IoT devices using machine learning techniques. arXiv preprint arXiv:1709.04647.

38. Treverton, G. F., Thvedt, A., Chen, A. R., Lee, K., McCue, M. (2018). *Addressing hybrid threats*. Försvarshögskolan (FHS).

39. Lajevardi, A. M., Amini, M. (2019). A semantic-based correlation approach for detecting hybrid and low-level apts. *Future Generation Computer Systems, 96,* 64–88. https://doi.org/10.1016/j.future.2019.01.056

40. Ding, Z., Wang, Y. (2022). Multi-step attack threat recognition algorithm based on attribute association in internet of things security. *Wireless Networks, 28,* 1–12.

41. Alshamrani, A., Myneni, S., Chowdhary, A., Huang, D. (2019). A survey on advanced persistent threats: Techniques, solutions, challenges, and research opportunities. *IEEE Communications Surveys & Tutorials, 21(2),* 1851–1877. https://doi.org/10.1109/COMST.9739

42. Ma, R., Cheng, P., Zhang, Z., Liu, W., Wang, Q. et al. (2019). Stealthy attack against redundant controller architecture of industrial cyber-physical system. *IEEE Internet of Things Journal, 6(6),* 9783–9793. https://doi.org/10.1109/JIoT.6488907

43. Al-Hawawreh, M., Sitnikova, E., Aboutorab, N. (2022). X-IIoTID: A connectivity-agnostic and device-agnostic intrusion data set for industrial internet of things. *IEEE Internet of Things Journal, 9(5),* 3962–3977. https://doi.org/10.1109/JIOT.2021.3102056

44. Miller, T., Staves, A., Maesschalck, S., Sturdee, M., Green, B. (2021). Looking back to look forward: Lessons learnt from cyber-attacks on industrial control systems. *International Journal of Critical Infrastructure Protection, 35,* 100464. https://doi.org/10.1016/j.ijcip.2021.100464

45. Alpaydin, E. (2014). *Introduction to machine learning*, pp. 1–20. Cambridge, MA: MIT Press.

46. Janiesch, C., Zschech, P., Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets, 31(3),* 685–695. https://doi.org/10.1007/s12525-021-00475-2

47. Dasgupta, D., Akhtar, Z., Sen, S. (2022). Machine learning in cybersecurity: A comprehensive survey. *The Journal of Defense Modeling and Simulation, 19(1),* 57–106. https://doi.org/10.1177/1548512920951275

48. Xiao, L., Li, Y., Han, G., Liu, G., Zhuang, W. (2016). Phy-layer spoofing detection with reinforcement learning in wireless networks. *IEEE Transactions on Vehicular Technology, 65(12),* 10037–10047. https://doi.org/10.1109/TVT.2016.2524258

49. LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature, 521(7553),* 436–444. https://doi.org/10.1038/nature14539

50. Drucker, H., Wu, D., Vapnik, V. N. (1999). Support vector machines for spam categorization. *IEEE Transactions on Neural networks, 10(5),* 1048–1054. https://doi.org/10.1109/72.788645

51. Kasongo, S. M. (2021). An advanced intrusion detection system for IIoT based on ga and tree based algorithms. *IEEE Access, 9,* 113199–113212. https://doi.org/10.1109/ACCESS.2021.3104113
52. Pavaiyarkarasi, R., Manimegalai, T., Satheeshkumar, S., Dhivya, K., Ramkumar, G. (2022). A productive feature selection criterion for Bot-IoT recognition based on random forest algorithm. *2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 539–545. Indore, India, IEEE.
53. Alsariera, Y. A., Adeyemo, V. E., Balogun, A. O., Alazzawi, A. K. (2020). Ai meta-learners and extra-trees algorithm for the detection of phishing websites. *IEEE Access, 8,* 142532–142542. https://doi.org/10.1109/Access.6287639
54. Bhati, B. S., Chugh, G., Al-Turjman, F., Bhati, N. S. (2021). An improved ensemble based intrusion detection technique using xgboost. *Transactions on Emerging Telecommunications Technologies, 32(6),* e4076. https://doi.org/10.1002/ett.v32.6
55. Al-Kasassbeh, M., Abbadi, M. A., Al-Bustanji, A. M. (2020). LightGBM algorithm for malware detection. In: Arai, K., Kapoor, S., Bhatia, R. (eds.), *Intelligent computing*, vol. 1230. Cham: Springer International Publishing.
56. Besharati, E., Naderan, M., Namjoo, E. (2019). LR-HIDS: Logistic regression host-based intrusion detection system for cloud environments. *Journal of Ambient Intelligence and Humanized Computing, 10(9),* 3669–3692. https://doi.org/10.1007/s12652-018-1093-8
57. Panda, M., Patra, M. R. (2007). Network intrusion detection using naive bayes. *International Journal of Computer Science and Network Security, 7(12),* 258–263.
58. Mukherjee, S., Sharma, N. (2012). Intrusion detection using naive bayes classifier with feature reduction. *Procedia Technology, 4,* 119–128. https://doi.org/10.1016/j.protcy.2012.05.017
59. Ng, A., Jordan, M. (2001). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, Cambridge, MA, USA, MIT Press.
60. Gu, J., Lu, S. (2021). An effective intrusion detection approach using SVM with naïve bayes feature embedding. *Computers & Security, 103,* 102158. https://doi.org/10.1016/j.cose.2020.102158
61. Gu, J., Wang, L., Wang, H., Wang, S. (2019). A novel approach to intrusion detection using SVM ensemble with feature augmentation. *Computers & Security, 86,* 53–62. https://doi.org/10.1016/j.cose.2019.05.022
62. Ponmalar, A., Dhanakoti, V. (2022). An intrusion detection approach using ensemble support vector machine based chaos game optimization algorithm in big data platform. *Applied Soft Computing, 116,* 108295. https://doi.org/10.1016/j.asoc.2021.108295
63. Jing, D., Chen, H. B. (2019). SVM based network intrusion detection for the UNSW-NB15 dataset. *2019 IEEE 13th International Conference on ASIC (ASICON)*, pp. 1–4. Chongqing, China, IEEE.
64. Wadkar, M., di Troia, F., Stamp, M. (2020). Detecting malware evolution using support vector machines. *Expert Systems with Applications, 143,* 113022. https://doi.org/10.1016/j.eswa.2019.113022
65. Han, H., Lim, S., Suh, K., Park, S., Cho, S. J. et al. (2020). Enhanced android malware detection: An SVM-based machine learning approach. *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 75–81. Busan, South Korea, IEEE.
66. Li, Q., Zhang, K., Cheffena, M., Shen, X. (2017). Channel-based sybil detection in industrial wireless sensor networks: A multi-kernel approach. *GLOBECOM 2017–2017 IEEE Global Communications Conference*, pp. 1–6. Singapore, IEEE.
67. Chang, C. P., Hsu, W. C., Liao, I. (2019). Anomaly detection for industrial control systems using K-means and convolutional autoencoder. *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pp. 1–6. Split, Croatia, IEEE.
68. Li, Y., Quevedo, D. E., Dey, S., Shi, L. (2016). Sinr-based dos attack on remote state estimation: A game-theoretic approach. *IEEE Transactions on Control of Network Systems, 4(3),* 632–642.

69. Chawla, A., Lee, B., Fallon, S., Jacob, P. (2019). Host based intrusion detection system with combined cnn/rnn model. In: *ECML PKDD 2018 workshops*, pp 149–158. Dublin, Ireland, Springer.

70. Li, Y., Xu, Y., Liu, Z., Hou, H., Zheng, Y. et al. (2020). Robust detection for network intrusion of industrial IoT based on multi-CNN fusion. *Measurement, 154,* 107450. https://doi.org/10.1016/j.measurement.2019.107450

71. Park, S. H., Park, H. J., Choi, Y. J. (2020). RNN-based prediction for network intrusion detection. *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Fukuoka, Japan, IEEE.

72. Kim, J., Kim, J., Thu, H. L. T., Kim, H. (2016). Long short term memory recurrent neural network classifier for intrusion detection. *2016 International Conference on Platform Technology and Service (PlatCon)*, pp. 1–5. Jeju, South Korea, IEEE.

73. Roy, B., Cheung, H. (2018). A deep learning approach for intrusion detection in internet of things using bi-directional long short-term memory recurrent neural network. *2018 28th International Telecommunication Networks and Applications Conference (ITNAC)*, pp. 1–6. Sydney, Australia, IEEE.

74. Sai Charan, P., Gireesh Kumar, T., Mohan Anand, P. (2019). Advance persistent threat detection using long short term memory (LSTM) neural networks. *Emerging Technologies in Computer Engineering: Microservices in Big Data Analytics: Second International Conference*, Jaipur, India, Springer.

75. Do Xuan, C., Dao, M. H. (2021). A novel approach for APT attack detection based on combined deep learning model. *Neural Computing and Applications, 33,* 13251–13264. https://doi.org/10.1007/s00521-021-05952-5

76. Ho, C. M. K., Yow, K. C., Zhu, Z., Aravamuthan, S. (2022). Network intrusion detection via flow-to-image conversion and vision transformer classification. *IEEE Access, 10,* 97780–97793. https://doi.org/10.1109/ACCESS.2022.3200034

77. Lopez-Martin, M., Carro, B., Sanchez-Esguevillas, A., Lloret, J. (2017). Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in IoT. *Sensors, 17(9),* 1967. https://doi.org/10.3390/s17091967

78. Li, W., Meng, P., Hong, Y., Cui, X. (2020). Using deep learning to preserve data confidentiality. *Applied Intelligence, 50,* 341–353. https://doi.org/10.1007/s10489-019-01515-3

79. Yilmaz, I., Masum, R., Siraj, A. (2020). Addressing imbalanced data problem with generative adversarial network for intrusion detection. *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 25–30. Las Vegas, Nevada, IEEE.

80. Taheri, R., Shojafar, M., Alazab, M., Tafazolli, R. (2021). FED-IIoT: A robust federated malware detection architecture in industrial IoT. *IEEE Transactions on Industrial Informatics, 17(12),* 8442–8452. https://doi.org/10.1109/TII.2020.3043458

81. Xing, M., Ding, W., Zhang, T., Li, H. (2023). STCGCN: A spatio-temporal complete graph convolutional network for remaining useful life prediction of power transformer. *International Journal of Web Information Systems, 19(2),* 102–117. https://doi.org/10.1108/IJWIS-02-2023-0023

82. Sethi, K., Sai Rupesh, E., Kumar, R., Bera, P., Venu Madhav, Y. (2020). A context-aware robust intrusion detection system: A reinforcement learning-based approach. *International Journal of Information Security, 19(6),* 657–678. https://doi.org/10.1007/s10207-019-00482-7

83. Martins, G., Bhatia, S., Koutsoukos, X., Stouffer, K., Tang, C. et al. (2015). Towards a systematic threat modeling approach for cyber-physical systems. *2015 Resilience Week (RWS)*, pp. 1–6. IEEE.

84. ElSayed, M. S., Le-Khac, N. A., Albahar, M. A., Jurcut, A. (2021). A novel hybrid model for intrusion detection systems in sdns based on cnn and a new regularization technique. *Journal of Network and Computer Applications, 191,* 103160. https://doi.org/10.1016/j.jnca.2021.103160

85. Li, X., Chen, W., Zhang, Q., Wu, L. (2020). Building auto-encoder intrusion detection system based on random forest feature selection. *Computers & Security, 95,* 101851. https://doi.org/10.1016/j.cose.2020.101851

86. Najar, A. A., Manohar Naik, S. (2022). DDoS attack detection using MLP and random forest algorithms. *International Journal of Information Technology, 14(5),* 2317–2327. https://doi.org/10.1007/s41870-022-01003-x

87. Wali, S., Khan, I. (2021). Explainable AI and random forest based reliable intrusion detection system. *TechRxiv*.

88. Lin, Y. D., Wang, Z. Y., Lin, P. C., Nguyen, V. L., Hwang, R. H. et al. (2022). Multi-datasource machine learning in intrusion detection: Packet flows, system logs and host statistics. *Journal of Information Security and Applications, 68,* 103248. https://doi.org/10.1016/j.jisa.2022.103248

89. Sahu, A., Mao, Z., Wlazlo, P., Huang, H., Davis, K. et al. (2021). Multi-source multi-domain data fusion for cyberattack detection in power systems. *IEEE Access, 9,* 119118–119138. https://doi.org/10.1109/ACCESS.2021.3106873

90. Zhou, P., Zhou, G., Wu, D., Fei, M. (2021). Detecting multi-stage attacks using sequence-to-sequence model. *Computers & Security, 105,* 102203. https://doi.org/10.1016/j.cose.2021.102203

91. Cheng, Z., Sun, D., Wang, L., Lv, Q., Wang, Y. (2022). MMSP: A LSTM based framework for multi-step attack prediction in mixed scenarios. *2022 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1–6. Rhodes, Greece, IEEE.

92. Park, C., Lee, J., Kim, Y., Park, J. G., Kim, H. et al. (2023). An enhanced AI-based network intrusion detection system using generative adversarial networks. *IEEE Internet of Things Journal, 10(3),* 2330–2345. https://doi.org/10.1109/JIOT.2022.3211346

93. Hubballi, N., Suryanarayanan, V. (2014). False alarm minimization techniques in signature-based intrusion detection systems: A survey. *Computer Communications, 49,* 1–17. https://doi.org/10.1016/j.comcom.2014.04.012

94. Morin, B., Mé, L., Debar, H., Ducassé, M. (2002). M2D2: A formal data model for IDS alert correlation. *International Workshop on Recent Advances in Intrusion Detection*, pp. 115–137. Berlin, Heidelberg, Springer.

95. Montesino, R., Fenz, S., Baluja, W. (2012). Siem-based framework for security controls automation. *Information Management & Computer Security, 20(4),* 248–263. https://doi.org/10.1108/09685221211267639

96. Mokhov, S. A., Paquet, J., Debbabi, M. (2014). The use of NLP techniques in static code analysis to detect weaknesses and vulnerabilities. In: *Advances in artificial intelligence*, pp. 326–332. Montréal, QC, Canada, Springer.

97. Wang, S., Liu, T., Tan, L. (2016). Automatically learning semantic features for defect prediction. *Proceedings of the 38th International Conference on Software Engineering*, pp. 297–308. New York, NY, USA, Association for Computing Machinery. https://doi.org/10.1145/2884781.2884804

98. Lin, G., Zhang, J., Luo, W., Pan, L., Xiang, Y. (2017). POSTER: Vulnerability discovery with function representation learning from unlabeled projects. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 2539–2541. New York, NY, USA, Association for Computing Machinery. https://doi.org/10.1145/3133956.3138840

99. Li, Z., Zou, D., Xu, S., Ou, X., Jin, H. et al. (2018). Vuldeepecker: A deep learning-based system for vulnerability detection. arXiv preprint arXiv:1801.01681.

100. Tang, G., Meng, L., Wang, H., Ren, S., Wang, Q. et al. (2020). A comparative study of neural network techniques for automatic software vulnerability detection. *2020 International Symposium on Theoretical Aspects of Software Engineering (TASE)*, Hangzhou, China.

101. Ziems, N., Wu, S. (2021). Security vulnerability detection using deep learning natural language processing. *IEEE INFOCOM 2021–IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 1–6. IEEE.

102. Thapa, C., Jang, S. I., Ahmed, M. E., Camtepe, S., Pieprzyk, J. et al. (2022). Transformer-based language models for software vulnerability detection. *Proceedings of the 38th Annual Computer Security Applications Conference*, New York, NY, USA, Association for Computing Machinery. https://doi.org/10.1145/3564625.3567985
103. HaddadPajouh, H., Dehghantanha, A., Khayami, R., Choo, K. K. R. (2018). A deep recurrent neural network based approach for internet of things malware threat hunting. *Future Generation Computer Systems, 85,* 88–96. https://doi.org/10.1016/j.future.2018.03.007
104. Le, T. T. H., Oktian, Y. E., Kim, H. (2022). XGBoost for imbalanced multiclass classification-based industrial internet of things intrusion detection systems. *Sustainability, 14(14),* 8707. https://doi.org/10.3390/su14148707
105. Nguyen, H. T., Ngo, Q. D., Le, V. H. (2020). A novel graph-based approach for IoT botnet detection. *International Journal of Information Security, 19(5),* 567–577. https://doi.org/10.1007/s10207-019-00475-6
106. Nguyen, T. N., Ngo, Q. D., Nguyen, H. T., Nguyen, G. L. (2022). An advanced computing approach for IoT-botnet detection in industrial internet of things. *IEEE Transactions on Industrial Informatics, 18(11),* 8298–8306. https://doi.org/10.1109/TII.2022.3152814
107. Jia, Y., Gu, Z., Du, L., Long, Y., Wang, Y. et al. (2023). Artificial intelligence enabled cyber security defense for smart cities: A novel attack detection framework based on the mdata model. *Knowledge-Based Systems, 276,* 110781. https://doi.org/10.1016/j.knosys.2023.110781
108. Wang, L., Jajodia, S., Singhal, A., Cheng, P., Wang, L. et al. (2017). Refining CVSS-based network security metrics by examining the base scores. *Network Security Metrics,* 25–52.
109. Ye, C., Shi, W., Zhang, R. (2021). Research on gray correlation analysis and situation prediction of network information security. *EURASIP Journal on Information Security, 2021(1),* 1–6.
110. Sun, X., Dai, J., Liu, P., Singhal, A., Yen, J. (2018). Using bayesian networks for probabilistic identification of zero-day attack paths. *IEEE Transactions on Information Forensics and Security, 13(10),* 2506–2521. https://doi.org/10.1109/TIFS.2018.2821095
111. Zhong, M., Zhou, Y., Chen, G. (2021). Sequential model based intrusion detection system for IoT servers using deep learning methods. *Sensors, 21(4),* 1113. https://doi.org/10.3390/s21041113
112. Wani, A., Khaliq, R. (2021). SDN-based intrusion detection system for IoT using deep learning classifier (IDSIoT-SDL). *CAAI Transactions on Intelligence Technology, 6(3),* 281–290. https://doi.org/10.1049/cit2.v6.3
113. Li, S. (2019). Zero trust based internet of things. *EAI Endorsed Transactions on Internet of Things, 5(20),* 165168.
114. Al-Taleb, N., Saqib, N. A. (2022). Towards a hybrid machine learning model for intelligent cyber threat identification in smart city environments. *Applied Sciences, 12(4),* 1863. https://doi.org/10.3390/app12041863
115. Cao, B., Li, C., Song, Y., Qin, Y., Chen, C. (2022). Network intrusion detection model based on cnn and gru. *Applied Sciences, 12(9),* 4184. https://doi.org/10.3390/app12094184
116. Azizjon, M., Jumabek, A., Kim, W. (2020). 1D CNN based network intrusion detection with normalization on imbalanced data. *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 218–224. Fukuoka, Japan.
117. Kale, R., Lu, Z., Fok, K. W., Thing, V. L. L. (2022). A hybrid deep learning anomaly detection framework for intrusion detection. *2022 IEEE 8th International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing, (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS)*, pp. 137–142. Jinan, China.
118. Casaju-Setien, J., Bielza, C., Larranaga, P. (2023). Anomaly-based intrusion detection in IIoT networks using transformer models. *2023 IEEE International Conference on Cyber Security and Resilience (CSR)*, pp. 72–77. Venice, Italy.

119. Xiao, C., Gou, Z., Tai, W., Zhang, K., Zhou, F. (2023). Imputation-based time-series anomaly detection with conditional weight-incremental diffusion models. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2742–2751. New York, NY, USA, Association for Computing Machinery. https://doi.org/10.1145/3580305.3599391
120. Jadidi, Z., Hagemann, J., Quevedo, D. (2022). Multi-step attack detection in industrial control systems using causal analysis. *Computers in Industry, 142,* 103741. https://doi.org/10.1016/j.compind.2022.103741
121. Melo, R., Macedo, D. (2019). A cloud immune security model based on alert correlation and software defined network. *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, pp. 52–57. Napoli, Italy.
122. Khosravi, M., Ladani, B. T. (2020). Alerts correlation and causal analysis for APT based cyber attack detection. *IEEE Access, 8,* 162642–162656. https://doi.org/10.1109/ACCESS.2020.3021499
123. Kumar, A., Thing, V. L. (2023). Raptor: Advanced persistent threat detection in industrial IoT via attack stage correlation. arXiv preprint arXiv:2301.11524.
124. Hu, H., Liu, Y., Zhang, H., Zhang, Y. (2018). Security metric methods for network multistep attacks using amc and big data correlation analysis. *Security and Communication Networks, 2018,* 1–14.
125. Kumari, S., Yadav, R. J., Namasudra, S., Hsu, C. H. (2021). Intelligent deception techniques against adversarial attack on the industrial system. *International Journal of Intelligent Systems, 36(5),* 2412–2437. https://doi.org/10.1002/int.v36.5