Artificial Intelligence Review (2022) 55:1029–1053
https://doi.org/10.1007/s10462-021-09976-0

Artificial intelligence in cyber security: research advances, challenges, and opportunities

Zhimin Zhang[1] · Huansheng Ning[1,2] · Feifei Shi[1] · Fadi Farha[1] · Yang Xu[1] · Jiabo Xu[3] · Fan Zhang[1] · Kim-Kwang Raymond Choo[4]

Published online: 13 March 2021

Abstract
In recent times, there have been attempts to leverage artificial intelligence (AI) techniques in a broad range of cyber security applications. Therefore, this paper surveys the existing literature (comprising 54 papers mainly published between 2016 and 2020) on the applications of AI in user access authentication, network situation awareness, dangerous behavior monitoring, and abnormal traffic identification. This paper also identifies a number of limitations and challenges, and based on the findings, a conceptual human-in-the-loop intelligence cyber security model is presented.

Keywords Cyber Security · Artificial Intelligence · Security Methods · Human-in-the-Loop

1 Introduction

As our society becomes more connected and technologically advanced, the role of security solutions and mitigation strategies will be more important. The challenge of securing our systems and society (that relies on these systems) is, however, compounded by the constantly evolving threat landscape (Xiao et al. 2018b; Guan and Ge 2018; Sliti et al. 2018).[1] Hence, designing more efficient and effective cyber security solutions is a topic of ongoing interest.

[1] Cloud adoption risk report 2019 (pdf). https://mscdss.ds.unipi.gr/wp-content/uploads/2018/10/Cloud-Adoption-Risk-Report-2019.pdf (2019).

✉ Huansheng Ning
ninghuansheng@ustb.edu.cn

[1] School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

[2] Beijing Engineering Research Center for Cyberspace Data Analysis and Applications, Beijing 100083, China

[3] School of Information Engineering, Xinjiang Institute of Engineering, Xinjiang, China

[4] Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, TX 78249-0631, USA

Cyber security refers to the use of various measures, methods, and means to ensure that systems are protected from threats and vulnerabilities, and to provide users with correct services efficiently. Therefore, the cyber security mentioned in this paper includes threats from outside[2] and within systems (known as network security in some studies). These threats will have a severe impact on the regular operation of the systems, so the goal of cyber security is to protect threats as much as possible, and to timely and effectively meet the requirements of detection before the accident, handling in the accident, and recovery after the accident.

In recent years, there have been attempts to design artificial intelligence (AI)-based solutions for a broad range of cyber security applications, partly due to the growing understanding of organizations in the importance of AI in mitigation cyber threats.[3] For example, AI-based approaches to model nonlinear problems have been shown to perform well in nonlinear classification (Ozsen et al. 2009), which can also be used to facilitate cyber threat classification. Interests in AI-based solutions are also partly driven by advances in computing capabilities. For example, according to Stanford University's *AI Index 2019 Report*,[4] the time required to train large-scale image classification system on cloud infrastructure decreases from approximately three hours in October 2017 to about 88 seconds in July 2019. Computing power for AI-based approaches is also reportedly doubling every three months or so, surpassing Moore's law. Such capabilities can be utilized to improve AI-based cyber security solutions' performance.[3,4] Examples of AI-based solutions include those developed by MIT and PatternEx (Veeramachaneni et al. 2016b), Darktrace (which uses AI to build an enterprise immune system),[5] DeepArmor (AI driven system against adversarial attacks) (Ji et al. 2019a), X by Invincea (which uses deep learning to understand and detect security threats),[6] Cognigo's DataSense (which uses machine learning algorithms to distinguish and protect sensitive data from non-sensitive data).[7] However, it is also known that machine intelligence cannot totally replace human intelligence, and the next generation of AI will most probably combine both human and machine intelligence (Kowert 2017; Zhang et al. 2020) (also referred to as human-in-the-loop).

Therefore, this paper surveys and summarizes key AI-based approaches for cyber security applications in user access authentication, network situation awareness, dangerous behavior monitoring, and abnormal traffic identification. Specifically, the following academic platforms are mainly searched: Google Scholar, ACM Digital Library, IEEE Xplore, SpringerLink, and ScienceDirect, as well as archival sites: ResearchGate, using the keywords and Boolean operators such as:

– ("artificial intelligence" *OR* "AI" *OR* "machine learning") *AND* ("access authentication" *OR* "mode authentication" *OR* "biometric authentication"),

[2] What's the difference between network security & cyber security? https://www.ecpi.edu/blog/whats-difference-between-network-security-cyber-security (2020).

[3] Ai in cybersecurity-capgemini worldwide. https://www.capgemini.com/news/ai-in-cybersecurity/ (2020).

[4] Ai index 2019 report (pdf). https://hai.stanford.edu/sites/g/files/sbiybj10986/f/ai_index_2019_report.pdf (2020).

[5] Enterprise immune system-darktrace. https://www.darktrace.com/en/products/enterprise/ (2019).

[6] Invincea launches x-as-a-service managed security. https://www.eweek.com/security/invincea-launches-x-as-a-service-managed-security (2020).

[7] Congnigo-infosecurity magazine. https://www.infosecurity-magazine.com/directory/cognigo/ (2019).

- ("artificial intelligence" *OR* "AI" *OR* "machine learning") *AND* ("situation awareness" *OR* "security situation awareness"),
- ("artificial intelligence" *OR* "AI" *OR* "machine learning") *AND* ("dangerous monitoring" *OR* "attacks"),
- ("artificial intelligence" *OR* "AI" *OR* "machine learning") *AND* ("traffic identification" *OR* "traffic analysis"),
- ("artificial intelligence" *OR* "AI" *OR* "machine learning") *AND* ("cyber security" *OR* "network security").

We located over 150 articles, and we used the following inclusion criteria that resulted in the selection of 54 articles to be discussed in this paper.

- The article has data, comparative experiments, or a detailed feasibility analysis of some proposed framework.
- The subject of the article aligns with the topic of our survey.
- The article was published in a peer-reviewed journal or a conference.
- The article was published within the last five years.

In addition, the paper located a number of related literature review and survey articles. Table 1 explains how our paper differs from existing literature review and survey articles (*Note:* the column of *Number of articles discussed* only counts the related methods and frameworks).

The remaining part of this paper is organized as follows. In the next two sections, the paper briefly reviews the key advantages and limitations of utilizing AI in the four cyber security applications (i.e., user access authentication, network situation awareness, dangerous behavior monitoring, and abnormal traffic identification). In the fourth section, the conceptual human-in-the-loop cyber security model is presented. Finally, the last section concludes this paper.

2 Potential applications of AI in cyber security applications

This section reviews related literature on AI-based solutions for user access authentication, network situation awareness, dangerous behavior monitoring, and abnormal traffic identification in Sects. 2.1–2.4, prioring to summarizing the discussion in Sect. 2.5.

2.1 User access authentication

2.1.1 User access authentication requirements

As the first defense line of cyber security, the system needs to strengthen the management of user access authentication, accurately identifies all kinds of camouflage behaviors, and realizes the detection of illegal or malicious objects. Before operation, the system should ensure that users are authenticated. At the same time, the user data should be confidential to prevent other risk events such as malicious collection of user information. Figure 1 shows that in the current authentication process, one of the research focuses on adding

Table 1 Existing literature review and survey articles on AI-based cyber security solutions

Existing review/survey	Review/survey cut-off period	Number of articles discussed	Focus
Mahmood and Afzal (2013)	Dec 2013	13	Real-time network streams monitoring and surveillance
Abdallah (2016)	2015	37	Insider threat detection and prediction algorithms
Jenab and Moslehpour (2016)	Apr 2015	139	Specific attacks, passwords, and intrusion detection
Benias and Markopoulos (2017)	May 2017	14	Cyber security in Industry 4.0
Yavanoglu and Aydos (2017a)	2017	34	Datasets for analyzing network traffic and detecting abnormalities
Taylor et al. (2019)	Apr 2018	42	Blockchain applications
Kruse et al. (2016)	Jul 2018	31	Applications in healthcare
Lu and Xu (2019)	Sep 2018	38	Cyber security applications in industries

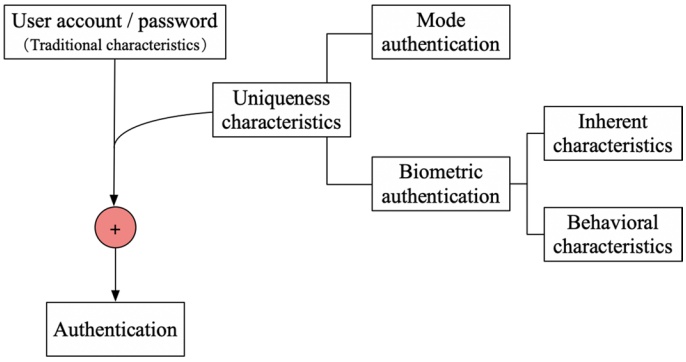


Fig. 1 User access authentication research focuses

other features to enhance the uniqueness of password matching process, so as to minimize the probability of others passing off as legitimate users.

2.1.2 Cases of mode authentication

How to match passwords and add other user characteristics to ensure the security of dual authentication is a challenge that needs to be solved in mode authentication. For example, current ATMs only use PIN codes for identity verification. This single mode does not guarantee the security of authentication (Adekunle et al. 2019). Based on the shortcomings of one-time authentication, multi authentication technology such as Shoufan (2017a) has been considered, it used Random Forest to achieve this goal. Korkmaz (2016) not only did password matching in the password authentication system, but also trained the user's keyboard using some styles through neural network. These styles included the user's typing speed and the typing style, key combination, and other aspects. Wang and Fang (2019) designed a kernel function with both global and local functions, and they built a mobile communication network security authentication mechanism based on Support Vector Regression (SVR), but less data was used in simulation. Chang et al. (2016) used One-Class Support Vector Machine (One-Class SVM) to realize keystroke dynamics pattern recognition, and this pattern has received widespread attention due to AI (Qiu 2017). Lu et al. (2020) used Convolutional Neural Network (CNN), reinforcement learning, and transfer learning to construct a physical authentication scheme. It aimed at mobile edge computing, and was used to resist rogue edge attacks.

2.1.3 Cases of biometric authentication

Compared with mode authentication, biometric authentication has been widely concerned because of its uniqueness, non-replicability, heredity, and invariance. McIntire et al. (2009) pointed out that to ensure the network security and stability of cooperation, it was necessary to determine whether the other party is an AI or a human user. Therefore, it was necessary to use "reverse Turing test" (a group of problems that can be solved by humans but not by computers). After determining whether it is a machine or a human, in order to prevent others from passing off, humans need to be verified. At present, the identification

is mainly based on the inherent characteristics of the human body (such as fingerprint, iris, etc.) and behavioral characteristics (such as voice, gait, etc.), and the powerful self-learning ability of AI that can effectively use them.

In the aspect of fingerprint recognition, Singh et al. (2017b) have proposed a fingerprint recognition method based on sparse proximity. Hariyanto et al. (2015b) have proposed a fingerprint feature point matching algorithm based on Artificial Neural Network (ANN) and compared the distance between feature points; the training process was accelerated by hardware. However, the paper missed performance evaluation. Saeed et al. (2018a) proposed a new fingerprint classification method based on modified Histograms of Oriented Gradients (HOG) descriptor, and this system used Extreme Learning Machine (ELM) with RBF kernel. Bakhshi and Veisi (2019a) put forword an end to end recognition model based on CNN without extracting features. In face recognition, Ding and Tao (2018) proposed a framework based on CNN. The features extracted from the clear and fuzzy pictures were shared, and the triple-state loss function was improved. Salyut and Kurnaz (2018b) proposed an ANN based on local binary mode to realize contour face recognition. Verma et al. (2019) used the hybrid genetic feature learning network in facial expression recognition.

In the aspect of iris recognition, Păvăloi and Niţă (2018a) used some distance measures and Scale Invariant Feature Transform (SIFT). Zhang et al. (2019) proposed a new method that uses dilated convolution to extract extra iris features, and several evaluation methods were used to test the model. Gangwar and Joshi (2016a) used the Deep Convolution Neural Network (DCNN) for iris recognition. Another technology combining AI and feature extraction technology, namely genetic and evolutionary feature extraction technology, was used in Shelton et al. (2016) to extract the most significant features in the images (small sizes). In the aspect of finger vein, multi-layer ELM (Yang et al. 2019), multi-layer CNN (Liu et al. 2017; Zhang et al. 2019; Hong et al. 2017), Fully CNN (FCN) (Zeng et al. 2020), Transfer Learning (Fairuz et al. 2018), and other methods could be used to achieve a recognition.

Amberkar et al. (2018b) studied the important role of Recurrent Neural Networks (RNNs) in the field of voice recognition. Some researchers introduced ladder networks to speech recognition(Parthasarathy and Busso 2019b),[8] and achieved good results. Han and Wang (2019a) proposed a new speech recognition method; it used Deep Belief Networks (DBN) to extract features and Proximal SVM to achieve recognition. Gait, as an important part of behavioral characteristics, has also attracted many researchers. For instance, Uddin et al. (2017a) firstly extracted features from depth silhouettes, and then used CNN to train and recognize. Deng et al. (2019) combined the three methods of RNN, CNN, and Radial Basis Function Neural Network (RBFNN) to eliminate the influence of perspective on gait recognition and achieved good results in the experiment. C4.5 decision tree (Thongsook et al. 2019a), HOG (Sugandhi and Raju 2019), and DCNN (Nithyakani et al. 2019) also performed well in gait recognition.

[8] Speech emotion recognition using semi-supervised learning with ladder networks. In: 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), pp. 1–5 (2018).

2.2 Network situation awareness

2.2.1 Network situation awareness requirements

In the process of network construction, the network designers may not find the vulnerability and insecurity in the network topology. In the process of network use, the non-uniform flow of data, which exposes the position of the network, perceives the weak link of the network in advance, provides the basis for network adjustment, needs to use network situation awareness. In the process of network situation awareness, complex networks need to be modeled, analyze the security situation of the network, and finally give the quantitative results of network situation awareness. To achieve this process, it is required that the situation awareness model has a strong knowledge base, from which it can quickly detect and match the network situation. At the same time, the model needs to have the ability to extract features, aim at never appearing in the network situation. Besides, reasoning can be realized to give reliable perception results.

2.2.2 Cases of network situational awareness combined with AI

Multi-entity Bayesian networks (MEBN) performs well in situational awareness, but there are some problems such as complex, so the idea of human-aided was used (Young Park et al. 2016b). Fuzzy Neural Network (FNN) could also play an important role in situation assessment (Li and Li 2017a), machine learning method combined with fuzzy theory could better reflect the change of network state. Yunhu Jin et al. (2016b) proposed a assessment model based on Random Forest. Every tree in the forest used independent samples and participated in the classification together, making the final result more objective. Li et al. (2018b) proposed an information fusion model based on time-space network situation awareness mechanism. This model used RBFNN for situation prediction. Yang et al. (2019) proposed a new calculating security indexes method based on CNNs. These indexes can help assess the network situation.

Shi et al. (2017) proposed a security situational awareness model. It is based on immune system and grey prediction theory. Dongmei and Jinxing (2018) used Wavelet Neural Network (WNN) based on particle swarm algorithm to achieve network situational awareness. They also designed a new algorithm to reduce data attributes. This research was committed to meeting the requirements of situation awareness in big data environment. Naderpour et al. (2014) used dynamic Bayesian network as situation assessment component, and used a fuzzy risk estimation method to generate results. In this design, the idea of human-in-the-loop was also well reflected. Bao et al. (2019b) aimed at the background of big data and AI to optimize the design of information security situation awareness system, including optimizing system hardware configuration, standardizing the synchronous operation mechanism of AI in multiple data security perception, improving the information security situation inference algorithm, designing the system software structure, and adding comparative repair steps based on security characteristic parameters.

2.3 Dangerous behavior monitoring

While new technologies such as big data and cloud computing continue to emerge, hackers' offensive methods are also constantly developing. With the rapid growth of data

volume and increasing access to the Internet, hackers are committed to find "lethal points" of the network and launch attacks on the network at any time. The original intrusion detection systems have been unable to adapt to the characteristics of the network. However, the high-speed flow of data is also conducive to find traces left by hacking activities, and has become important evidence for taking security precautions in advance. In order to achieve cyber security with accurate methods, it is necessary to monitor dangerous behaviors and their types in time. Otherwise, there will be a situation of "emergency medical treatment", which effectively protects the network but it wastes a lot of resources. To this end, researchers have begun to improve and innovate on the basis of the original intrusion detection systems to make the current network requirements of the intrusion detection systems as scalable as possible.

Marir et al. (2018) pointed out a new distributed large-scale network abnormal behavior detection method. It combined the deep feature extraction and multi-layer integrated Support Vector Machine (SVM) and used the distributed DBN to reduce the dimension of large-scale network traffic dataset to find abnormal behaviors. Kanimozhi and Jacob (2019b) proposed an AI-based hyper-parameter optimized network intrusion detection. The system used ANN technology to detect botnet attacks and abled to deploy on multiple machines. Aljamal et al. (2019b) proposed a hybrid intrusion detection system using machine learning in a cloud computing environment. The system fused the K-Means clustering algorithm and the SVM classification algorithm. Pandeeswari and Kumar (2016) proposed a hypervisor-based anomaly detection system, in which the main technology was a neural network based on fuzzy C-means algorithm. In the cloud computing environment, the system showed good performance under low frequency attack.

Some systems focused on monitoring a single dangerous behavior, such as Distributed Denial of Service (DDoS). Jyothi et al. (2016b) proposed a complete detection framework for DDoS, and achieved good results in the experiment. It used K-Means for behavior clustering and SVM for classification. Yuan et al. (2017a) proposed a deep learning-based detection method for DDoS. The whole system consisted of CNNs, RNNs, and fully-connected layers. Hsieh and Chan (2016a) divided the entire system into five parts, namely: data collector, Hadoop-HPFS, format converter, data processing device, and neural network detection module. This system could analyze high-speed, high-traffic network systems, and neural networks could also effectively identify data packet characteristics. The advantages of AI can play a significant role in mitigating a variety of specific attacks on the network (Jenab and Moslehpour 2016).

With the advent of the 5G era, some scholars have started to study the anomaly detection of 5G technologies. For example, an adaptive deep learning based 5G network anomaly detection system was proposed in Fernández Maimó et al. (2018). In this framework, two layers of deep learning models were used; one was focused on the method of using network flow aggregation detection to quickly search for abnormal signs, it mainly uses Deep Neural Network (DNN) for processing; the other one was based on the relationship between the timeline and related symptoms to identify network anomalies, and directly communicated with the monitoring and diagnosis module after finding the anomalies. The Long Short-Term Memory (LSTM) was implemented to handle time series well.

2.4 Abnormal traffic identification

Any network has a certain carrying capacity. Within normal threshold, network can play a significant role in and provide users with high-quality services. Hackers will deliberately

inject a large amount of illegal data into the network structure, which makes the network nodes and links unable to bear and cause accidents, unable to provide services for users, and even lead to serious problems such as information loss. How to provide an important basis for network situational awareness through analysis of network traffic, timely detection of high-risk behaviors on the cyberspace, and effective measures are of great significance for enhancing network response and maintaining overall cyber security.

According to research results by Ahmed et al. (2015a), abnormal flow detection methods could be divided into four categories, which were detection methods based on classification, statistics, clustering, and information theory. Aljurayban and Emam (2015a) proposed an intrusion detection system framework in cloud computing. This framework could be integrated on different cloud levels and could capture traffic then sent it to ANN. Zhang et al. (2019) proposed a Parallel Cross Convolutional Neural Network (PCCN) based on deep learning to implement traffic anomaly detection in multi-class imbalanced networks. It was mainly composed of two parallel CNNs and used multiple feature fusion methods. Zeng et al. (2019) considered the current development trend of network traffic encryption and proposed an end-to-end network traffic recognition framework based on deep learning. The framework had a two-layer structure; it used CNN to extract features and LSTM to record time characteristics. Kong's team is dedicated to the combination of abnormal traffic identification and AI. They compared the performance of K-means (unsupervised) and SVM (supervised) methods in abnormal traffic (Kong et al. 2018a), and built a system based on SVM to identify and classify multiple attack traffic (Kong et al. 2017b). Besides, they proposed to use parallel computing to accelerate the training of the model (Kong et al. 2018b).

2.5 Summary

The aforementioned four subsections respectively introduced the AI in cyber security from different aspects. This subsection mainly summarizes the relevant technologies used in various aspects as shown in Table 2.

By summarizing these articles, it is found that most of the proposed methods are realized through the transformation of the basic methods of AI as shown in Table 2. Among them, 24% of the methods used CNN, 15% of the methods used SVM, and 12% of the methods used ANN, which are the most frequent used basic methods (refer to Fig. 2a for detailed usage proportion). These basic methods provide the basis and reflect the feasibility and superiority for the applications of cyber security.

But at the same time, the field of cyber security has its own characteristics, so these articles combine the characteristics of the research direction to improve the basic methods, mainly including: methods fusion (using two or more basic methods in the model), features selection (selecting new features or expressions to improve the identification ability), and models optimization (used to speed up the parameter update speed or better finding the optimal solution). From this point of view, we classify the articles and get Table 3.

In order to more clearly describe the use of basic methods in the four research aspects, Table 3 is drawn with pie chart, as shown in Fig. 2b. For user access authentication, more researches focused on features selection. Network situation awareness and dangerous behavior monitoring focused on the research of models optimization and methods fusion. Models optimization was regarded as the focus of abnormal traffic identification. For different research aspects, researchers can choose to determine the means of using the methods, and finally get the purpose of achieving new breakthroughs in technology.

Table 2 Basic methods used in references

Serial no.	Basic methods	References
1.	Artificial neural network (ANN) (including deep ANN (DNN))	Korkmaz (2016), Hariyanto et al. (2015b), Salyut and Kurnaz (2018b), Kanimozhi and Jacob (2019b), Pandeeswari and Kumar (2016), Hsieh and Chan (2016a), Aljurayban and Emam (2015a) and Fernández Maimó et al. (2018)
2.	Convolutional neural network (CNN) (including dilated convolution, deep CNN (DCNN), fully CNN (FCN))	Bakhshi and Veisi (2019a), Ding and Tao (2018), Zhang et al. (2019), Gangwar and Joshi (2016a), Yuan et al. (2017a), Zhang et al. (2019), Zeng et al. (2019), Yang et al. (2019), Liu et al. (2017), Zhang et al. (2019), Hong et al. (2017), Zeng et al. (2020), Uddin et al. (2017a), Deng et al. (2019), Nithyakani et al. (2019) and Lu et al. (2020)
3.	Support vector machine (SVM) (including one-class SVM, support vector regression (SVR), proximal SVM (PSVM))	Wang and Fang (2019), Bao et al. (2019b), Marir et al. (2018), Aljamal et al. (2019b), Kong et al. (2017b), Kong et al. (2018a), Kong et al. (2018b), Han and Wang (2019a), Chang et al. (2016) and Jyothi et al. (2016b)
4.	Extreme learning machine (ELM)	Saeed et al. (2018a) and Yang et al. (2019)
5.	Long short-term memory (LSTM)	Fernández Maimó et al. (2018) and Zeng et al. (2019)
6.	Recurrent neural network (RNN)	Yuan et al. (2017a), Amberkar et al. (2018b) and Deng et al. (2019)
7.	Radial basis function neural network (RBFNN)	Li et al. (2018b) and Deng et al. (2019)
8.	Ladder networks	Parthasarathy and Busso (2019b)[a]
9.	Genetic algorithm	Verma et al. (2019) and Shelton et al. (2016)
10.	Wavelet neural network (WNN)	Dongmei and Jinxing (2018)
11.	Deep belief network (DBN)	Marir et al. (2018) and Han and Wang (2019a)
12.	Fuzzy neural network (FNN)	Li and Li (2017a)
13.	Bayesian network (including multi-entity Bayesian networks (MEBN))	Naderpour et al. (2014) and Young Park et al. (2016b)
14.	Artificial immune	Shi et al. (2017)
15.	Transfer learning	Fairuz et al. (2018) and Lu et al. (2020)
16.	Reinforcement learning	Lu et al. (2020)
17.	K-Means	Aljamal et al. (2019b), Kong et al. (2018a) and Jyothi et al. (2016b)
18.	Sparse representation based classification (SRC)	Singh et al. (2017b)
19.	Scale-invariant feature transform (SIFT)	Păvăloi and Niţă (2018a)
20.	Random forest	Shoufan (2017a) and Yunhu Jin et al. (2016b)
21.	Histograms of oriented gradients (HOG)	Saeed et al. (2018a) and Sugandhi and Raju (2019)

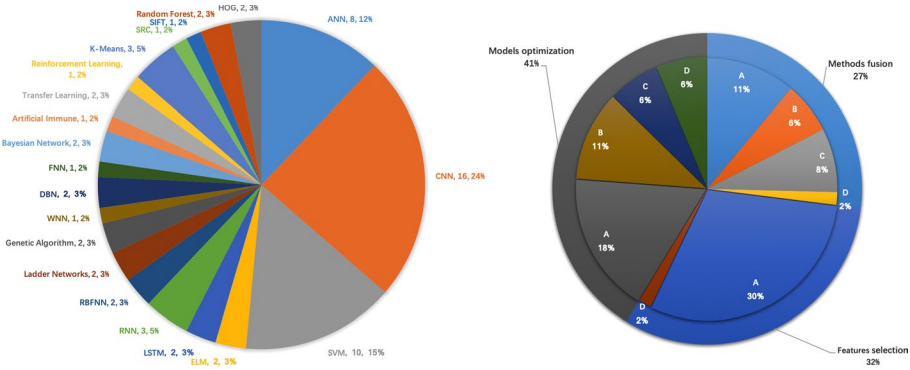
[a]Speech emotion recognition using semi-supervised learning with ladder networks. In: 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), pp. 1–5 (2018)

Figure 3 shows a model that summarizes most of the research ideas in the field of cyber security. This model deals with security issues through four steps, including data selection and acquisition, data feature extraction, model construction, and specific applications. To this end, the entire model is divided into four levels as follows:

- Data layer: data selection is the most basic work, and the quality of data selection directly affects the performance of the model. For the four research aspects, the data used in the experiments include general datasets and self-collecting datasets. In mode authentication and network situation awareness, all the articles mentioned in this paper (except Dongmei and Jinxing (2018)) used self-collecting datasets, such as operators behaviors (Shoufan 2017a), cyber security reports (Yunhu Jin et al. 2016b), specific network traffic data (Li and Li 2017a; Li et al. 2018b), etc. Using self-collecting data can enrich the diversity of data, but it causes some difficulties for the accuracy of single model estimation and the comparison of different models. On the contrary, a small number of articles in the remaining research perspectives collected data on their own (e.g. Chang et al. 2016; Lu et al. 2020; Shoufan 2017a; Korkmaz 2016), and most of them used general datasets. The general datasets they have covered are given in Table 4.
- Feature layer: effective feature extraction is an important factor in determining security issues accurately. The unified processing of data is a necessary step to do before starting data extraction, especially when using self collecting datasets [e.g. (Wang and Fang 2019)]. Some methods integrated feature extraction in model construction and representation, but others performed separate feature extraction to enhance the ability to express data (refer to Table 3).
- Intelligent layer: This layer is implemented in two steps, namely modeling and evaluation. The construction of the model is an essential step to embody AI and the core content of the general model (for the basic methods and usages involved in the model, refer to Tables 2 and 3, respectively). The effectiveness of the model was judged by the evaluation methods. The main used methods were accuracy rate, followed by the equal error rate (EER). Besides, some studies used specific evaluation methods for specific problems, such as response time (e.g. Lu et al. 2020; Hariyanto et al. 2015b; Salyut and Kurnaz 2018b), receiver operating characteristic (ROC) curve (e.g. Zhang et al. 2019; Ding and Tao 2018; Shelton et al. 2016; Fairuz et al. 2018), cumulative match characteristic (CMC) curve (e.g. Shelton et al. 2016), etc.
- Application layer: After construction, these models either provided solutions for problems, or deployed them in combination with specific scene. The theme of the applications was consistent because of using AI to ensure cyber security.

In addition, this paper also summarizes some of innovative methods mentioned in Table 5. These summaries include the datasets, features and their extraction methods, classification models, and maximum accuracy of methods. At the same time, timeliness and complexity are also used to compare the various methods. These two indicators can reflect the effectiveness of the methods, which also meet the processing requirements of cyber security issues.

In the field of cyber security, AI can play an important role, but at the same time, it needs to be adjusted to make this technology more suitable for the use requirements of this field. How to achieve fast detection, improve detection accuracy, and mine data characteristics are the focus of the current research in this field.



(a) Proportion of basic methods **(b)** Proportion of basic methods used

Fig. 2 Proportion of basic methods and their used

3 Limitations of AI-based approaches

Can AI detect all uncertain events? The answer is no. As a "double-edged sword", this new technology has its own shortcomings as well as a good performance. This section discusses the factors that make the AI model dishonest in the field of cyber security.

3.1 Interference of confusing data

How much interference can cheat AI? Maybe one pixel is enough. Su et al's experiment 2019 showed that only changing one pixel in the image can lead to the misclassification of neural network. Kolosnjaji et al. (2018a) modified a few bytes in the malicious sample software tightly, which leaded to neural network classification error. Hu and Tan (2017) used the Generative Adversarial Network (GAN) to obtain malware samples, which could bypass the detection system. As can be seen from the these examples, once the data is "infected", there is a chance to cheat the AI system, resulting in the unsafe state of the network.

3.2 Maliciously modified model

The implementation of AI model is a program, which may have some vulnerabilities. These vulnerabilities may be due to the designer's unreasonable and careless design of the logical structure of the model. They may come from specific high-level language, hardware specific problems, or the back door embedded in the model. Gu et al. (2017b) implemented the backdoor in the neural network, which made the performance of the neural network in the specific attacker sample very poor. These shortcomings also reflect from the side that the given answers by the program are not necessarily accurate.

Table 3 Classification of basic methods used

Basic methods used	References	Categories
Methods fusion	Verma et al. (2019), Lu et al. (2020), Zhang et al. (2019), Zeng et al. (2020), Han and Wang (2019a) and Deng et al. (2019)[a]	A
	Li et al. (2018b), Li and Li (2017a), Naderpour et al. (2014) and Bao et al. (2019b)	B
	Marir et al. (2018), Aljamal et al. (2019b), Yuan et al. (2017a), Jyothi et al. (2016b), Fernández Maimó et al. (2018)	C
	Zeng et al. (2019)	D
Features selection	Korkmaz (2016), Shoufan (2017a), Chang et al. (2016), Singh et al. (2017b), Hariyanto et al. (2015b), Saeed et al. (2018a), Lu et al. (2020), Ding and Tao (2018), Salyut and Kurnaz (2018b), Verma et al. (2019), Păvăloi and Niţă (2018a), Shelton et al. (2016), Yang et al. (2019), Hong et al. (2017), Parthasarathy and Busso (2019b), Han and Wang (2019a), Uddin et al. (2017a), Nithyakani et al. (2019) and Sugandhi and Raju (2019)	A
	–	B
	–	C
	Zhang et al. (2019)	D
Models optimization	Wang and Fang (2019), Bakhshi and Veisi (2019a), Ding and Tao (2018), Zhang et al. (2019), Gangwar and Joshi (2016a), Liu et al. (2017), Zhang et al. (2019), Fairuz et al. (2018), Parthasarathy and Busso (2019b), Lu et al. (2020) and Thongsook et al. (2019a)	A
	Li et al. (2018b), Yang et al. (2019), Shi et al. (2017), Young Park et al. (2016b), Yunhu Jin et al. (2016b), Dongmei and Jinxing (2018) and Bao et al. (2019b)	B
	Marir et al. (2018), Kanimozhi and Jacob (2019b), Pandeeswari and Kumar (2016) and Hsieh and Chan (2016a)	C
	Aljurayban and Emam (2015a), Zhang et al. (2019), Kong et al. (2017b) and Kong et al. (2018b)	D

A for user access authentication, B for network situation awareness, C for dangerous behavior monitoring, D for abnormal traffic identification

[a]Speech emotion recognition using semi-supervised learning with ladder networks. In: 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), pp. 1–5 (2018)

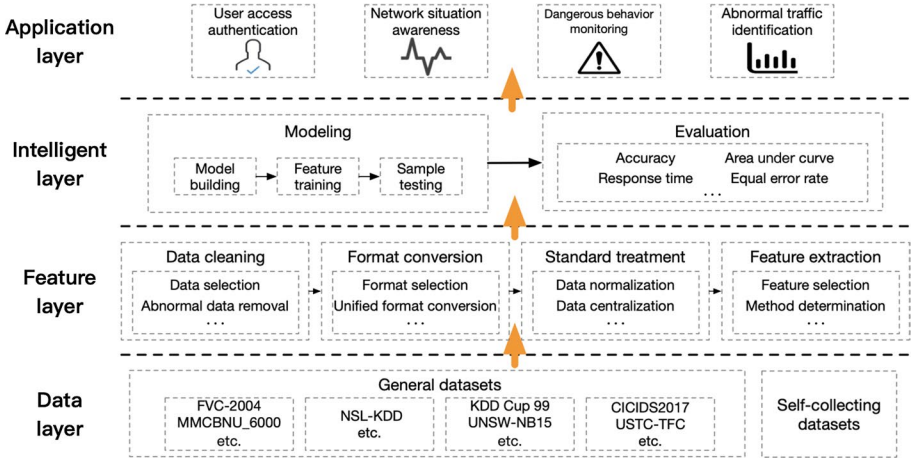


Fig. 3 A general model for cyber security

3.3 Lack of transparency in AI decision-making process

In the decision-making process of AI, all the participants including programmers, do not know why the AI model gives the final decision results i.e., the decision-making process of AI lacks transparency. The AI model is similar to a black box. In the process of its creation and self-improvement, it can realize the automatic configuration and adjustment of parameters without too much intervention of staff, and thus, saves human resources. Nevertheless, at the same time, the problem is that its decision-making process is difficult to explain clearly. Although the AI model can achieve high accuracy, the tests are all implemented in the test set. Therefore, when facing unknown events, whether the AI model can achieve such a high accuracy remains to be verified. When there are objections to the decision-making results given by the AI model, it is difficult to explain the decision-making process, so some people will be skeptical of the decision-making results. That will not be conducive to the rapid judgment of the network situation, or even cause irreversible consequences. Some research teams have begun to conduct in-depth research on this issue (Schlegel et al. 2019).

3.4 High data requirements

At present, the AI models basically need a lot of data to complete the training. Before using data, they may do operations which mainly includes a series of steps such as data noise reduction, normalization, missing value filling, etc. If a supervised method is used, it is necessary to label the data manually.[9] However, due to the strong heterogeneity of cyberspace, different cyber structures may produce different high-risk events, and these events have the characteristics of sudden. Therefore, each possible high-risk event cannot be estimated in advance before the design of the models, nor can these high-risk events be analyzed and labeled in advance. Meanwhile AI models have a high demand for data, which may not be able to make a timely judgment.

[9] Knowledge-directed artificial intelligence reasoning over schemas (kairos). https://www.darpa.mil/program/knowledge-directed-artificial-intelligence-reasoning-over-schemas (2020).

Table 4 Summary of general datasets

Research perspective		General datasets
Mode authentication		–
Biometric authentication	Fingerprint	FCV-2000 (Singh et al. 2017b), FCV-2002 (Singh et al. 2017b; Bakhshi and Veisi 2019a), FCV-2004 (Singh et al. 2017b; Saeed et al. 2018a)
	Face	COX Face (Ding and Tao 2018), YouTube Faces (Ding and Tao 2018), PaSC (Ding and Tao 2018), CK+ (Verma et al. 2019), AFEW (Verma et al. 2019), MUG (Verma et al. 2019), OULU (Verma et al. 2019)
	Iris	UBIRIS (Păvăloi and Niţă 2018a), UPOL (Păvăloi and Niţă 2018a), CASIA-4i (Zhang et al. 2019), UBIRIS.v2 (Zhang et al. 2019), ND-IRIS-0405 (Zhang et al. 2019; Gangwar and Joshi 2016a), ND-CrossSensor-Iris-2013 (Gangwar and Joshi 2016a), Multispectral Iris Dataset (Shelton et al. 2016)
	Finger vein	SDUMLA (Yang et al. 2019; Liu et al. 2017; Zhang et al. 2019; Hong et al. 2017; Zeng et al. 2020), UTFVP (Yang et al. 2019), MMCBNU-6000 (Yang et al. 2019; Zhang et al. 2019; Zeng et al. 2020), FV-USM (Zhang et al. 2019), HKPU (Zhang et al. 2019; Zeng et al. 2020)
	Voice	IEMOCAP (Parthasarathy and Busso 2019b)[a], MSP-IMPROV (Parthasarathy and Busso 2019b), MSP-Podcast (Parthasarathy and Busso 2019b)
	Gait	CASIA-B (Deng et al. 2019; Sugandhi and Raju 2019), TUM-GAID (Nithyakani et al. 2019)
Network situation awareness		NSL-KDD (Dongmei and Jinxing 2018)
Dangerous behavior monitoring		KDD Cup 99 (Marir et al. 2018; Pandeeswari and Kumar 2016), CICIDS2017 (Marir et al. 2018), UNSW-NB15 (Marir et al. 2018; Aljamal et al. 2019b),NSL-KDD (Marir et al. 2018), CSE-CIC-IDS2018 (Kanimozhi and Jacob 2019b), ISCX2012 (Yuan et al. 2017a), LLDOS1.0 (Hsieh and Chan 2016a), CTU (Fernández Maimó et al. 2018)
Abnormal traffic identification		CICIDS2017 (Zhang et al. 2019), USTC-TFC (Zeng et al. 2019), ISCX VPN-nonVPN (Zeng et al. 2019), KDD Cup 99 (Kong et al. 2018a, 2017b, 2018b)

[a]Speech emotion recognition using semi-supervised learning with ladder networks. In: 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), pp. 1–5 (2018)

Table 5 Summary of some innovate methods

Name	Category	Datasets	Features	Extraction methods	Models	Accuracy (max, %)	Timeliness	Complexity
Chang et al. (2016)	A (mode)	Self-collecting	Keystroke dynamics	One-class SVM	One-class SVM	98.00	High	Low
Korkmaz (2016)	A (mode)	Self-collecting	Password and keyboard styles	ANN	ANN	85.00	Mid	Low
TBE-CNN (Ding and Tao 2018)	A (biometric)	COX face, YouTube faces, and PaSC	Blur-robust face representations	TBE-CNN	TBE-CNN	94.96 ± 0.31	Mid	High
FD-UNet (Zhang et al. 2019)	A (biometric)	CASIA-4i, UBIRIS.v2, and ND-IRIS-0405	Iris segmentation	Dilated convolution	FD-UNet	97.36	Mid	Mid
Deng et al. (2019)	A (biometric)	CASIA-B	Gait silhouettes	RBFNN	CNN RNN	92.00	High	High
SASS (Naderpour et al. 2014)	B	–	Situation awareness	Risk indicators	Dynamic Bayesian network	–	High	High
Marir et al. (2018)	C	KDD Cup 99, CICIDS2017, UNSW-NB15, and NSL-KDD	Abnormal behavior from network traffic data	Distributed DBN	Multi-layer SVM	99.82	Mid	High
Fernández Maimó et al. (2018)	C	CTU	5G anomaly symptom and network	DNN LSTM	LSTM	95.00	High	Mid
PCNN (Zhang et al. 2019)	D	CICIDS2017	Flow features	PCNN	PCNN	99.10	Mid	Mid

A for user access authentication, B for network situation awareness, C for dangerous behavior monitoring, D for abnormal traffic identification

4 Conceptual human-in-the-loop cyber security model

4.1 The development of human-in-the-loop

The design and implementation of AI, especially neural networks, try to emulate the human brain. Its purpose was to achieve the same way of thinking as humans by using connecting neurons. Unfortunately, AI needs a large number of samples to achieve learning, without reasoning ability, the final model after training is complicated for us to understand how it makes decisions. Although some attempts to seek the explanation of AI (Schlegel et al. 2019; Wang et al. 2020), this work is still in the initial stage. Therefore, to achieve efficient use of intelligence, it is far from enough to rely on these tools without human participation (Nunes et al. 2015).

AI plays an important role in the prevention and detection of high-risk network behaviors, but there are some bad factors that will interfere with its correct judgment. The AI is to assist the security specialists in this field, not to replace them. Therefore, it is still necessary for relevant specialists to intervene and use relevant network knowledge to make professional judgment on the current network form.

At present, a new type of AI is being developed, that is human-in-the-loop. In 2017, defense advanced research projects agency (DARPA) designed *the DARPA robotics challenge (DRC)*, the idea of human-machine teamwork was embodied.[10] In January 2019, DARPA released the AI project named *KAIROS* to implement a system that could identify events and attract humans attention. In May, this administration announced the launch of *ACE* project to develop air combat capability of human-machine collaborative dogfighting.[11] In November, the U.S. Department of Defense received a report about the mechanical fighters and human-machine integration describing Cyborg fighters to be built for future wars.[12] In the military field, the research of this new technology has begun, which also reflects its importance.

Human-in-the-loop could combine human wisdom and machine intelligence, which is an important methodology to realize the complementary advantages of human and machine. AI can process a large number of data quickly and has a good recognition effect for specific scenes. But it may be disturbed and may not judge the new situation accurately. Compared with machines, human beings are more flexible, and can give a more rapid judgment in the face of new changes in the network, but also need machines to provide an auxiliary role. Interactive machine learning, which was used in AI (Holzinger et al. 2016), has also been embodied in cyber security. For example, the use of human-in-the-loop in the monitoring of software side-channel vulnerabilities could effectively improve the detection ability (Santhanam et al. 2017). Adding this idea to situational awareness, while achieving visualization, also enhanced the reliability of the system (Tyworth et al. 2013b). Thus, the use of AI and human-in-the-loop in cyber security will further enhance the models' capabilities.

[10] Darpa robotics challenge (DRC) using human-machine teamwork to perform disasterresponse with a humanoid robot. https://apps.dtic.mil/docs/citations/AD1027886 (2020).

[11] Training ai to win a dogfight. https://www.darpa.mil/news-events/2019-05-08 (2020).

[12] Cyborg super soldiers: Us army report reveals vision for deadly 'machine humans' with infrared sight, boosted strength and mind-controlled weapons by 2050. https://www.dailymail.co.uk/sciencetech/article-7738669/US-Military-scientists-create-plan-cyborg-super-soldier-future.html (2019).

4.2 Model design of cyber security based on human-in-the-loop

AI technology has significant advantages in the applications of cyber security, but also has its own shortcomings. Based on this fact, this paper proposes a new model based on human-in-the-loop named Human-in-the-Loop Cyber Security Model (HLCSM). The model design is shown in Fig. 4. HLCSM is mainly divided into two sub modules: Machine Detection Module (MDM) and Manual Intervention Module (MIM). The two sub modules interact with each other to prevent and detect the cyber uncertain events.

4.2.1 Machine detection module (MDM)

In HLCSM, MDM plays the "leading role". When high-risk events arrive, MDM will preprocess the data, which may include data cleaning, data normalization, and other operations. When the data are regular, it is very important to extract features from the data. The data contain the key information of locating event type and other information with little correlation. With the increase of data volume, it is necessary to select features and reduce dimensions in order to complete tasks quickly. After the data features are extracted, they are sent to the recognition method, which is the key link of MDM. Only when the selected method meets the requirements, the recognition accuracy will be higher. After the recognition result is given, the running result will be judged by the confidence level module (CLM), and its value will determine whether the final result is based on MDM.

In MDM, two identification methods are used. Based on the fact that neural networks can deploy backdoors (Gu et al. 2017b), the current use of a single recognition method does not guarantee complete reliability of the result. Therefore, in MDM design, two recognition methods are used to generate the judgment results in parallel, and the knowledge base provides the judgment basis. For these two methods, the used judgment techniques should be as different as possible to increase the diversity of judgment methods. By using two recognition methods, the difficulty of elusion can be increased and the accuracy of the result can be further improved. Both of the two produced results will be handled by CLM.

4.2.2 Manual intervention module (MIM)

In HLCSM, MIM plays the "auxiliary role". When the result of MDM is unsatisfactory, the processing power should be given to MIM. After the safety specialists receive the information feedback, the event will be handled according to the experiential knowledge. The final result of determining whether the event is safe will be given directly by the safety specialists, and will no longer be intervened by MDM.

Due to the uncertainty in the types of network events, and in order to further expand the processing capabilities of MDM, it is necessary to expand MDM with the results of MIM processing. After giving the final result, specialists also need to perform data calibration. Through features extraction, new type of event is added to the knowledge base to achieve the role of expanding MDM.

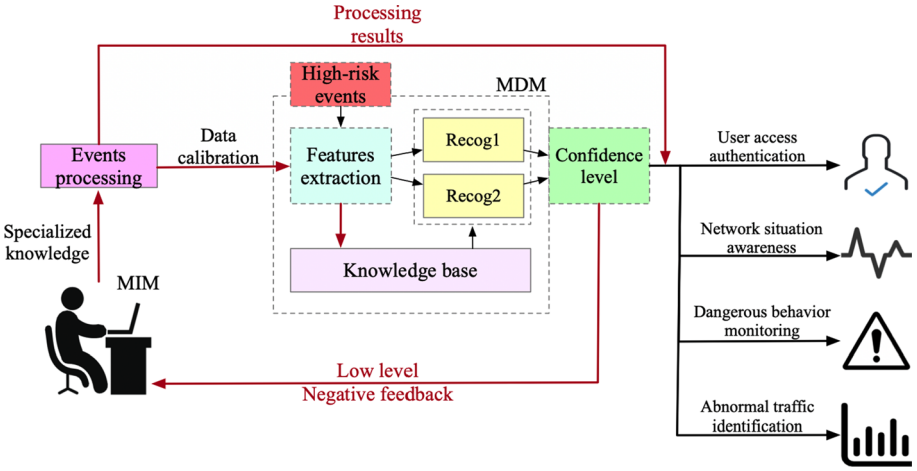


Fig. 4 Human-in-the-Loop Cyber Security Model (HLCSM)

4.2.3 Confidence level module (CLM)

In order to connect MDM and MIM and realize the cooperation of the two modules, CLM is introduced to the design. The main function of CLM is to determine whether MIM needs to be called to complete the event processing. When the recognition methods of MDM give two processing results, CLM integrates the results and gives the confidence level. When the confidence level is high, the final result will be given directly by MDM, which can not only save manpower, but also reduce the identification time and meet the requirements of cyber events processing. However, when the confidence level is low, the feedback will arrive at MIM and be processed by specialists to minimize errors.

There are two cases of low confidence level. First, MDM gives two completely different judgment results. This will not determine whether the event is safe or not. Secondly, the results of the two methods are the same, but they do not reach the theoretical threshold under the indexes such as accuracy, so it can be considered that the credibility of the results is not high, and they will not be regarded as the final judgment result.

4.3 Flow chart of HLCSM

In order to better express the relationship between the parts of HLCSM, Fig. 5 shows the flow chart of the model. It can be seen from the flow chart that some follow-up work should be carried out after the judgment of event i.e., if it is a high-risk event, quarantine and other means should be taken; if it is a safety event, the operation should be ended. This will constitute a complete network event defense model.

In HLCSM, the idea of human-in-the-loop is embodied. It is achieved through the cooperation between MDM and MIM judged by LCM. Based on this idea, despite the rapid development of AI, it will still be disturbed. The emergence of AI should assist human beings, rather than completely replace them. In this model, there is no need to use a large number of security specialists, because the main work is done by MDM. In this way, it can

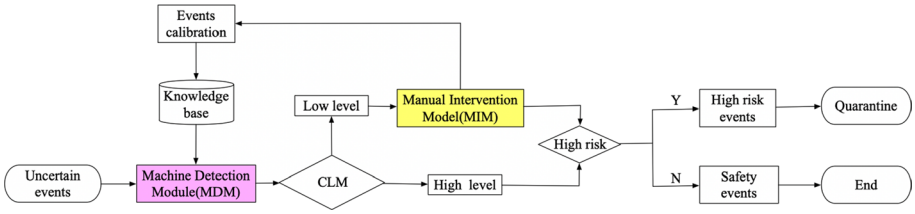


Fig. 5 Flow chart of HLCSM

reduce personnel allocation and expenditure cost. However, due to the particularity of the used environment, this work has to have security specialists, so as to more effectively resist the rapid changes of the current network environment.

By expanding the knowledge base, MDM can handle more event types. This expansion process needs MIM auxiliary implementation. When the knowledge base reaches a certain scale, it needs to be redesigned to meet the requirements of rapid search. This design is not considered in this model.

Using two recognition methods, on the one hand, can solve the problem of unreliable single recognition method. On the other hand, it will not cause too much burden to CLM module because of using too many recognition methods. At present, the recognition method is more complex, so it requires higher computing power for the local machine. If many recognition algorithms are used at the same time, they will also cause load to the operation of the machine. But at the same time, although cloud computing and other technologies can reduce this burden, in order to prevent insecurity, it is not recommended that the methods training be carried out in a non local way.

In order to realize the cooperation between human and machine, the intermediate "translation" work is also essential. LCM takes this responsibility and plays the role of human-machine interface. The key of human-machine integration is to realize the complementary advantages of both of them, which is also the original intention of our model. If a model wants to achieve a high degree of coordination between them, it needs to have full "communication". Through the confidence level, the bridge between human and machine communication is an important link which can not be obtained by human-machine interaction. The realization of this link will directly affect the mutual cooperation between the human and the machine.

4.4 Comparison

In order to more intuitively express the difference between the model in this paper and the general model, both models are roughly evaluated from 10 aspects in Table 6, including:

- Scalability: in MDM, two recognition methods are used to complement each other. Users can choose specific technical methods according to specific scenarios, or expand the number of recognition methods. At the same time, the knowledge base can be expanded and get greater advantages in dealing with new security issues.
- Maintainability: compared with the general model, HLCSM adds two new modules (MIM and CLM), which need to realize the cooperation between the modules. In case of failure, the maintenance difficulty is higher than the general model.

Table 6 Comparison of models

Characteristics	The general model	HLCSM
Scalability	Mid	High
Maintainability	Easy	Hard
Closeness	Low	High
Integration	Mid	High
Participatory	Low	High
Interpretability	Low	High
Reliability	Low	High
Complexity	Mid	High
Security	Mid	High
Deployability	Easy	Complex

- Closedness: HLCSM incorporated the concept of closed loop in the design, taking the high integration of human and machine as the final goal. After processing the result, it is necessary to further determine whether this result can be used as the final result.
- Integration: the general model can divide the processing of cyber security issues into four modules, which are reflected in MDM. In addition, the three major modules in HLCSM are highly integrated.
- Participatory: the main idea of HLCSM is human-in-the-loop, so the safety specialists can participate in the decision.
- Interpretability: CLM is an important bridge between AI and safety specialists, which can intervene the results with low confidence and solve the problem of poor interpretability of AI to a certain extent.
- Reliability: the experimental results are given by two recognition methods, and processed by the safety specialists. The reliability of the model is significantly improved compared with the general model.
- Complexity: the processing of the general model is only covered in MDM, which shows that the complexity of HLCSM is higher.
- Security: both models have the characteristics of security, but the use of HLCSM will further reduce the interference of data forgery and other factors, so the security is improved.
- Deployability: due to the larger system size of HLCSM, it is more difficult to deploy. Considering the fact that AI can be implanted in the backdoor (Gu et al. 2017b), it is recommended to deploy recognition methods separately in the cloud and on-premises to increase security.

To sum up, regardless that HLCSM performs poorly in maintainability and deployment, it has obvious advantages in other respects compared to the current general model. Therefore, HLCSM can be used as a new research idea to solve the shortcomings of the current use of AI in cyber security.

5 Conclusion

As AI has significant potential in cyber security applications, it is important for the researcher and practitioner community to understand the current state-of-play and the associated challenges. Hence, this paper reviewed articles focusing on the applications of AI in four cyber security domains, namely: user access authentication, network situation awareness, dangerous behavior monitoring, and abnormal traffic identification.

In addition to identifying research challenges and opportunities, the paper posited the importance of human-in-the-loop and proposed a conceptual model and explained how it can be utilized. Hence, a future follow-up work is to implement and evaluate the proposed conceptual model in collaboration with an organization.

Acknowledgements This work was funded by the National Natural Science Foundation of China (Grant No. 61872038). This work of K.-K. R. Choo was supported only by the Cloud Technology Endowed Professorship.

References

Adekunle YA, Okolie SO, Adebayo AO, Ebiesuwa S, Ehiwe DD (2019) Holistic exploration of gaps vis-à-vis artificial intelligence in automated teller machine and internet banking. In: International journal of applied information systems (IJAIS), vol 12

Abdallah AE (2016) Detection and prediction of insider threats to cyber security: a systematic literature review and meta-analysis. Big Data Anal 1(1):6

Ahmed M, Mahmood AN, Hu J (2015) A survey of network anomaly detection techniques. J. Netw. Comput. Appl. 60:19–31

Aljamal I, Tekeoğlu A, Bekiroglu K, Sengupta S (2019) Hybrid intrusion detection system using machine learning techniques in cloud computing environments. In: 2019 IEEE 17th international conference on software engineering research, management and applications (SERA), pp 84–89

Aljurayban NS, Emam A (2015) Framework for cloud intrusion detection system service. In: 2015 2nd world symposium on web applications and networking (WSWAN), pp 1–5

Amberkar A, Awasarmol P, Deshmukh G, Dave P (2018) Speech recognition using recurrent neural networks. In: 2018 international conference on current trends towards converging technologies (ICCTCT), pp 1–4

Bakhshi B, Veisi H (2019) End to end fingerprint verification based on convolutional neural network. In: 2019 27th Iranian conference on electrical engineering (ICEE), pp 1994–1998

Bao H, He H, Liu Z, Liu Z (2019) Research on information security situation awareness system based on big data and artificial intelligence technology. In: 2019 international conference on robots intelligent system (ICRIS), pp 318–322

Benias N, Markopoulos AP (2017) A review on the readiness level and cyber-security challenges in industry 4.0. In: 2017 south eastern European design automation, computer engineering, computer networks and social media conference (SEEDA-CECNSM), pp 1–5

Chang C, Eude T, Obando Carbajal LE (2016) Biometric authentication by keystroke dynamics for remote evaluation with one-class classification. In: Khoury R, Drummond C (eds) Advances in artificial intelligence. Springer, Cham, pp 21–32

Deng M, Yang H, Cao J, Feng X (2019) View-invariant gait recognition based on deterministic learning and knowledge fusion. In: 2019 international joint conference on neural networks (IJCNN), pp 1–8

Ding C, Tao D (2018) Trunk-branch ensemble convolutional neural networks for video-based face recognition. IEEE Trans Pattern Anal Mach Intell 40(4):1002–1014

Dongmei Z, Jinxing L (2018) Study on network security situation awareness based on particle swarm optimization algorithm. Comput Ind Eng 125:764–775. https://doi.org/10.1016/j.cie.2018.01.006

Fairuz S, Habaebi MH, Elsheikh EMA (2018) Finger vein identification based on transfer learning of alexnet. In: 2018 7th international conference on computer and communication engineering (ICCCE), pp 465–469

Fernández Maimó L, Perales Gómez AL, García Clemente FJ, Gil Pérez M, Martínez Pérez G (2018) A self-adaptive deep learning-based system for anomaly detection in 5g networks. IEEE Access 6:7700–7712

Gangwar A, Joshi A (2016) Deepirisnet: deep iris representation with applications in iris recognition and cross-sensor iris recognition. In: 2016 IEEE international conference on image processing (ICIP), pp 2301–2305

Gu T, Dolan-Gavitt B, Garg S (2017) BadNets: identifying vulnerabilities in the machine learning model supply chain. ArXiv e-prints arXiv:1708.06733

Guan Y, Ge X (2018) Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks. IEEE Trans Signal Inf Process Netw 4(1):48–59

Han Z, Wang J (2019) Speech emotion recognition based on deep learning and kernel nonlinear PSVM. In: 2019 Chinese control and decision conference (CCDC), pp. 1426–1430

Hariyanto, Sudiro SA, Lukman S (2015) Minutiae matching algorithm using artificial neural network for fingerprint recognition. In: 2015 3rd international conference on artificial intelligence, modelling and simulation (AIMS), pp 37–41

Holzinger A, Plass M, Holzinger K, Crişan GC, Pintea CM, Palade V (2016) Towards interactive machine learning (IML): applying ant colony algorithms to solve the traveling salesman problem with the human-in-the-loop approach. In: Buccafurri F, Holzinger A, Kieseberg P, Tjoa AM, Weippl E (eds) Availability, reliability, and security in information systems. Springer, Cham, pp 81–95

Hong H, Lee M, Park K (2017) Convolutional neural network-based finger-vein recognition using nir image sensors. Sensors (Switzerland) 17:1297. https://doi.org/10.3390/s17061297

Hsieh C, Chan T (2016) Detection ddos attacks based on neural-network using apache spark. In: 2016 international conference on applied system innovation (ICASI), pp 1–4

Hu W, Tan Y (2017) Generating adversarial malware examples for black-box attacks based on gan. CoRR. http://arxiv.org/abs/1702.05983

Jenab K, Moslehpour S (2016) Cyber security management: a review. Soc. Bus. Manag. Dyn. 5(11):16–39

Ji Y, Bowman B, Huang HH (2019) Securing malware cognitive systems against adversarial attacks. In: 2019 IEEE international conference on cognitive computing (ICCC), pp 1–9

Jyothi V, Wang X, Addepalli SK, Karri R (2016) Brain: behavior based adaptive intrusion detection in networks: Using hardware performance counters to detect ddos attacks. In: 2016 29th international conference on VLSI design and 2016 15th international conference on embedded systems (VLSID), pp 587–588

Sugandhi K, Raju G (2019) An efficient hog-centroid descriptor for human gait recognition. In: 2019 amity international conference on artificial intelligence (AICAI), pp 355–360

Kanimozhi, V, Jacob TP (2019) Artificial intelligence based network intrusion detection with hyperparameter optimization tuning on the realistic cyber dataset cse-cic-ids2018 using cloud computing. In: 2019 international conference on communication and signal processing (ICCSP), pp 0033–0036

Kolosnjaji B, Demontis A, Biggio B, Maiorca D, Giacinto G, Eckert C, Roli F (2018) Adversarial malware binaries: evading deep learning for malware detection in executables. In: 2018 26th European signal processing conference (EUSIPCO), pp 533–537

Kong L, Huang G, Wu K (2017) Identification of abnormal network traffic using support vector machine. In: 2017 18th international conference on parallel and distributed computing, applications and technologies (PDCAT), pp 288–292

Kong L, Huang G, Wu K, Tang Q, Ye S (2018) Comparison of internet traffic identification on machine learning methods. In: 2018 international conference on big data and artificial intelligence (BDAI), pp 38–41

Kong L, Huang G, Zhou Y, Ye J (2018) Fast abnormal identification for large scale internet traffic. In: Proceedings of the 8th international conference on communication and network security, ICCNS 2018. Association for Computing Machinery, New York, pp 117–120 (2018). https://doi.org/10.1145/3290480.3290498

Korkmaz Y (2016) Developing password security system by using artificial neural networks in user log in systems. In: 2016 electric electronics, computer science, biomedical engineerings' meeting (EBBT), pp 1–4

Kowert W (2017) The foreseeability of human-artificial intelligence interactions. Texas Law Rev 96:181–204

Kruse C, Frederick B, Jacobson T, Monticone D (2016) Cybersecurity in healthcare: a systematic review of modern threats and trends. Technol Health Care 25:1–10. https://doi.org/10.3233/THC-161263

Li C, Li XM (2017) Cyber performance situation awareness on fuzzy correlation analysis. In: 2017 3rd IEEE international conference on computer and communications (ICCC), pp 424–428
Li X, Zhang X, Wang D (2018) Spatiotemporal cyberspace situation awareness mechanism for backbone networks. In: 2018 4th international conference on big data computing and communications (BIGCOM), pp 168–173
Liu W, Li W, Sun L, Zhang L, Chen P (2017) Finger vein recognition based on deep learning. In: 2017 12th IEEE conference on industrial electronics and applications (ICIEA), pp 205–210
Lu X, Xiao L, Xu T, Zhao Y, Tang Y, Zhuang W (2020) Reinforcement learning based PHY authentication for Vanets. IEEE Trans Veh Technol 69(3):3068–3079
Lu Y, Xu LD (2019) Internet of things (IoT) cybersecurity research: a review of current research topics. IEEE Internet Things J 6(2):2103–2115
Mahmood T, Afzal U (2013) Security analytics: big data analytics for cybersecurity: a review of trends, techniques and tools. In: 2013 2nd national conference on information assurance (NCIA), pp 129–134
Marir N, Wang H, Feng G, Li B, Jia M (2018) Distributed abnormal behavior detection approach based on deep belief network and ensemble SVM using spark. IEEE Access 6:59657–59671
McIntire JP, McIntire LK, Havig PR (2009) A variety of automated turing tests for network security: Using ai-hard problems in perception and cognition to ensure secure collaborations. In: 2009 international symposium on collaborative technologies and systems, pp 155–162
Naderpour M, Lu J, Zhang G (2014) An intelligent situation awareness support system for safety-critical environments. Decis Support Syst 59:325–340. https://doi.org/10.1016/j.dss.2014.01.004
Nithyakani P, Shanthini A, Ponsam G (2019) Human gait recognition using deep convolutional neural network. In: 2019 3rd international conference on computing and communications technologies (ICCCT), pp 208–211
Nunes DS, Zhang P, Sá Silva J (2015) A survey on human-in-the-loop applications towards an internet of all. IEEE Commun Surv Tutor 17(2):944–965
Ozsen S, Gunes S, Kara S, Latifoglu F (2009) Use of kernel functions in artificial immune systems for the nonlinear classification problems. IEEE Trans Inf Technol Biomed 13(4):621–628
Pandeeswari N, Kumar G (2016) Anomaly detection system in cloud environment using fuzzy clustering based ANN. Mob Netw Appl 21(3):494–505
Parthasarathy S, Busso C (2019) Semi-supervised speech emotion recognition with ladder networks. IEEE/ACM Trans Audio, Speech, Lang Process 28:2697–2709
Păvăloi L, Niţă CD (2018) Iris recognition using sift descriptors with different distance measures. In: 2018 10th international conference on electronics, computers and artificial intelligence (ECAI), pp 1–4
Qiu M (2017) Keystroke biometric systems for user authentication. J Signal Process Syst 86(2–3):175–190
Saeed F, Hussain M, Aboalsamh HA (2018) Classification of live scanned fingerprints using histogram of gradient descriptor. In: 2018 21st Saudi computer society national computer conference (NCC), pp 1–5
Salyut J, Kurnaz C (2018) Profile face recognition using local binary patterns with artificial neural network. In: 2018 international conference on artificial intelligence and data processing (IDAP), pp 1–4
Santhanam GR, Holland B, Kothari S, Ranade N (2017) Human-on-the-loop automation for detecting software side-channel vulnerabilities. In: Shyamasundar RK, Singh V, Vaidya J (eds) Information systems security. Springer, Cham, pp 209–230
Schlegel U, Arnout H, El-Assady M, Oelke D, Keim DA (2019) Towards a rigorous evaluation of xai methods on time series. In: 2019 IEEE/CVF international conference on computer vision workshop (ICCVW), pp 4197–4201
Shelton J, Jenkins J, Roy K (2016) Micro-dimensional feature extraction for multispectral iris recognition. SoutheastCon 2016:1–5
Shi Y, Li T, Renfa L, Peng X, Tang P (2017) An immunity-based iot environment security situation awareness model. J Comput Commun 5:182–197. https://doi.org/10.4236/jcc.2017.57016
Shoufan A (2017) Continuous authentication of uav flight command data using behaviometrics. In: 2017 IFIP/IEEE international conference on very large scale integration (VLSI-SoC), pp 1–6
Singh K, Kumar J, Tripathi G, Chullai GA (2017) Sparse proximity based robust fingerprint recognition. In: 2017 international conference on computing, communication and automation (ICCCA), pp 1025–1028
Sliti M, Abdallah W, Boudriga N (2018) Jamming attack detection in optical uav networks. In: 2018 20th international conference on transparent optical networks (ICTON), pp 1–5
Su J, Vargas DV, Sakurai K (2019) One pixel attack for fooling deep neural networks. IEEE Trans Evolut Comput 23(5):828–841
Taylor PJ, Dargahi T, Dehghantanha A, Parizi RM, Choo KKR (2019) A systematic literature review of blockchain cyber security. Digit Commun Netw 6(2):147–156

Thongsook A, Nunthawarasilp T, Kraypet P, Lim J, Ruangpayoongsak N (2019) C4.5 decision tree against neural network on gait phase recognition for lower limp exoskeleton. In: 2019 1st international symposium on instrumentation, control, artificial intelligence, and robotics (ICA-SYMP), pp 69–72

Tyworth M, Giacobe NA, Mancuso VF, McNeese MD, Hall DL (2013) A human-in-the-loop approach to understanding situation awareness in cyber defence analysis. EAI End Trans Secur Saf. https://doi.org/10.4108/trans.sesa.01-06.2013.e6

Uddin MZ, Khaksar W, Torresen J (2017) A robust gait recognition system using spatiotemporal features and deep learning. In: 2017 IEEE international conference on multisensor fusion and integration for intelligent systems (MFI), pp 156–161

Veeramachaneni K, Arnaldo I, Korrapati V, Bassias C, Li K (2016) Ai$\hat{2}$: Training a big data machine to defend. In: 2016 IEEE 2nd 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 49–54

Verma M, Vipparthi SK, Singh G (2019) Hinet: hybrid inherited feature learning network for facial expression recognition. IEEE Lett Comput Soc 2(4):36–39

Wang Z, Fang B (2019) Application of combined kernel function artificial intelligence algorithm in mobile communication network security authentication mechanism. J Supercomput 75(9):5946–5964

Wang ZJ, Turko R, Shaikh O, Park H, Das N, Hohman F, Kahng M, Chau DH (2020) CNN explainer: learning convolutional neural networks with interactive visualization. IEEE Trans Vis Comput Gr. https://doi.org/10.1109/TVCG.2020.3030418

Xiao R, Zhu H, Song C, Liu X, Dong J, Li H (2018) Attacking network isolation in software-defined networks: New attacks and countermeasures. In: 2018 27th international conference on computer communication and networks (ICCCN), pp 1–9

Yang H, Jia Y, Han WH, Nie YP, Li SD, Zhao XJ (2019) Calculation of network security index based on convolution neural networks, pp 530–540. https://doi.org/10.1007/978-3-030-24271-8_47

Yang W, Wang S, Hu J, Zheng G, Yang J, Valli C (2019) Securing deep learning based edge finger vein biometrics with binary decision diagram. IEEE Trans Ind Inform 15(7):4244–4253

Yavanoglu O, Aydos M (2017) A review on cyber security datasets for machine learning algorithms. In: 2017 IEEE international conference on big data (big data), pp 2186–2193

Young Park C, Blackmond Laskey K, Costa PCG, Matsumoto S (2016) A process for human-aided multi-entity bayesian networks learning in predictive situation awareness. In: 2016 19th international conference on information fusion (FUSION), pp 2116–2124

Yuan X, Li C, Li X (2017) Deepdefense: identifying ddos attack via deep learning. In: 2017 IEEE international conference on smart computing (SMARTCOMP), pp 1–8

Yunhu Jin, Shen Y, Zhang G, Hua Zhi (2016) The model of network security situation assessment based on random forest. In: 2016 7th IEEE international conference on software engineering and service science (ICSESS), pp 977–980

Zeng J, Wang F, Deng J, Qin C, Zhai Y, Gan J, Piuri V (2020) Finger vein verification algorithm based on fully convolutional neural network and conditional random field. IEEE Access 8:65402–65419

Zeng Y, Qi Z, Chen W, Huang Y, Zheng X, Qiu H (2019) Test: an end-to-end network traffic examination and identification framework based on spatio-temporal features extraction. CoRR. http://arxiv.org/abs/1908.10271

Zhang W, Lu X, Gu Y, Liu Y, Meng X, Li J (2019) A robust iris segmentation scheme based on improved u-net. IEEE Access 7:85082–85089

Zhang Y, Chen X, Guo D, Song M, Teng Y, Wang X (2019) PCCN: Parallel cross convolutional neural network for abnormal network traffic flows detection in multi-class imbalanced network traffic flows. IEEE Access 7:119904–119916

Zhang Y, Li W, Zhang L, Ning X, Sun L, Lu Y (2019) Adaptive learning Gabor filter for finger-vein recognition. IEEE Access 7:159821–159830

Zhang Z, Shi F, Wan Y, Xu Y, Zhang F, Ning H (2020) Application progress of artificial intelligence in military confrontation. Chin J Eng 42(9):1106–1118. https://doi.org/10.13374/j.issn2095-9389.2019.11.19.001