Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Review:

Artificial intelligence algorithms for cyberspace security applications: a technological and status review[#]

Jie CHEN[†‡1,2], Dandan WU[†2], Ruiyun XIE[2]

[1]*School of Cybersecurity, Northwestern Polytechnical University, Xi'an 710000, China*

[2]*China Electronics Technology Network Information Security Co., Ltd., Chengdu 610000, China*

[†]E-mail: chenjie1900@mail.nwpu.edu.cn; wudd@cetcsc.com

Received July 21, 2022; Revision accepted Dec. 11, 2022; Crosschecked May 15, 2023

Abstract: Three technical problems should be solved urgently in cyberspace security: the timeliness and accuracy of network attack detection, the credibility assessment and prediction of the security situation, and the effectiveness of security defense strategy optimization. Artificial intelligence (AI) algorithms have become the core means to increase the chance of security and improve the network attack and defense ability in the application of cyberspace security. Recently, the breakthrough and application of AI technology have provided a series of advanced approaches for further enhancing network defense ability. This work presents a comprehensive review of AI technology articles for cyberspace security applications, mainly from 2017 to 2022. The papers are selected from a variety of journals and conferences: 52.68% are from Elsevier, Springer, and IEEE journals and 25% are from international conferences. With a specific focus on the latest approaches in machine learning (ML), deep learning (DL), and some popular optimization algorithms, the characteristics of the algorithmic models, performance results, datasets, potential benefits, and limitations are analyzed, and some of the existing challenges are highlighted. This work is intended to provide technical guidance for researchers who would like to obtain the potential of AI technical methods for cyberspace security and to provide tips for the later resolution of specific cyberspace security issues, and a mastery of the current development trends of technology and application and hot issues in the field of network security. It also indicates certain existing challenges and gives directions for addressing them effectively.

Key words: Artificial intelligence (AI); Machine learning (ML); Deep learning (DL); Optimization algorithm; Hybrid algorithm; Cyberspace security

https://doi.org/10.1631/FITEE.2200314 **CLC number:** TP393.08; TP18

1 Introduction

Cyberspace security faces three technical problems: the first is the timeliness and accuracy of detecting attacks, the second is the credibility assessment and prediction of the security situation, and the third is the effectiveness of security strategy decision-making. With an increasing number of network attacks, cyberspace security is becoming a crucial risk for any enterprise. Advances in artificial intelligence (AI) technology, especially machine learning (ML) and deep learning (DL), can support the detection of threats and advise network analysts, pointing the way for cyberspace security experts to deal with evolving cyber threats (Bresniker et al., 2019; Zeadally et al., 2020). It has been shown that flexible cyberspace security practices and approaches are needed to address the current problems facing aviation, noted in the National Aviation Safety Strategy Report released in December 2018 and the Government Accountability Office Aviation

‡ Corresponding author

[#] Electronic supplementary materials: The online version of this article (https://doi.org/10.1631/FITEE.2200314) contains supplementary materials, which are available to authorized users

ORCID: Jie CHEN, https://orcid.org//0000-0002-5643-193X; Dandan WU, https://orcid.org/0000-0001-5214-387X

Cyberspace Security Report released in October 2020. In addition, when designing solutions, if you do not explore the feasibility of using AI-driven cyberspace security tools, you will ignore the potential advantages that these technologies can offer (Choi et al., 2020; Sun YY et al., 2020). A new study shows that AI algorithms can extract the best feature representation from the attack and defense big data of cyberspace, which can effectively help security technical analysts research network threat detection, analysis and prediction, evaluation, scheme optimization, decision scheme generation, and other cyberspace security applications (Garcia et al., 2021). DL methods are widely used in the field of cyberspace security (Salih et al., 2021) to solve specific problems such as security-oriented program analysis, protection of return-oriented programming attacks, implementation of control flow integrity, defense against network attacks, malware classification, anomaly detection based on system events, memory forensics, and blurring of software security (Berman et al., 2019).

Research on AI algorithms in the field of cyberspace security is gradually developing. Related surveys (Nguyen TTT and Armitage, 2008; Wu and Banzhaf, 2010; Buczak and Guven, 2016; Torres et al., 2019) have described ML applications to cyberspace security issues but have not mentioned DL. Others analyze AI technology with cyberspace security, but these technical applications have limited cyberspace security. Existing research focuses mainly on AI algorithms applied to solve cyberspace security issues, such as attack detection, prediction, and analysis. The issues refer only to the accuracy under various types of network attack, such as those mentioned in Berman et al. (2019) for 12 kinds of cyberspace security issues of attack detection algorithm models, including malware, botnet detection, drive-by download attacks, network intrusion detection, tile type identification, network traffic identification, spam identification, insider threat detection, border gateway anomaly detection, verification if keystrokes were typed by a human, user authentication, and false data injection attack detection. Apruzzese et al. (2018) and Xin et al. (2018) paid specific attention to network attacks, including intrusion detection. They targeted weaknesses in datasets and the field of surveys for further progress. Salih et al. (2021) paid their attention to research on cyberspace security attack datasets with AI techniques. They provided a detailed comparison of techniques and field execution to achieve cyberspace defense optimization. There are other researchers who have analyzed and discussed the application of AI algorithms for cyberspace security issues, such as the Internet of Things (IoT) and cyber-physical systems. Research aimed at protecting cyber-physical systems was summarized by Wickramasinghe et al. (2018). Furthermore, Al-Garadi et al. (2020) reviewed ML and DL techniques to solve IoT issues.

This review is particularly comprehensive because it includes the comprehensive AI algorithms in cyberspace security issue applications not only for intrusion detection but also for assessment and defense decision-making. The expansion of AI algorithms in cyberspace security employment is reviewed. The highlights of this work are as follows:

1. The research material has the characteristics of timeliness, authority, and universality. More than 150 articles were collected, especially from 2011 to 2022, 52.68% of which came from Elsevier, Springer, and IEEE journals, and 25% came from international academic conferences in this field.

2. The research problems include difficult problems and active directions in the field of network security. This review studies and analyzes three specific technical problems that need to be solved urgently in the field of cyberspace security, including network attack detection, security situation assessment, and network security defense strategy optimization. In each direction, AI algorithms are classified as ML, DL, or optimization algorithms, and are sorted and analyzed from the aspects of algorithms, datasets, simulation, and comparative experiments, even advantages and disadvantages.

3. Through omni-directional, multi-dimensional, and detailed research, the potential advantages of AI algorithms in solving specific problems in the field of network security are statistically analyzed, showing the current development trends of technologies and applications, the hot issues, and the focuses of different countries in the field of network security.

Furthermore, this survey is unique in that it covers a wide range of AI algorithms in three kinds of cyberspace security applications, not only ML and DL, but also optimization algorithms. We attempt to create a reference point for technical experts who want to

recognize the potential of AI technology in cyberspace security, to guide researchers to further tap the potential advantages of AI algorithms to solve the current complex and changing cyberspace security issues, and to open up the three links and multi-dimensional cyberspace security problem-solving paths of "intrusion detection modeling, situation assessment and prediction modeling, and defense decision-making modeling," thus building a comprehensive solution framework with perspective, integration, and intelligence.

2 AI technologies

The concept of AI first arose at the Dartmouth Conference in 1956 and reflected the wish for machines that would think and react like the human brain. Because of its tremendous difficulty and attractiveness, AI has attracted scientists and enthusiasts to invest in research since its birth. According to the level of AI implementation, we can describe three kinds of AI:

Artificial narrow intelligence (ANI): This type of intelligence is good at a particular task. It is used in image and speech recognition systems to identify objects, people, and other elements in images and audio recordings. Other examples are the automatic classification of spam, self-driving vehicles, face recognition on mobile phones, and so on. Most of the current AI is ANI.

Artificial general intelligence (AGI): At the beginning of the proposal of the concept of AI, people expected to build complex computers to achieve the same complex intelligence as what human beings have. This type of intelligence requires machines to be proficient in listening, speaking, reading, and writing like people. At present, AGI has not been achieved.

Artificial super intelligence (ASI): ASI, an addition to AGI, is the intelligence that is smarter than the human brain in almost every field, including innovation, social interaction, and thinking. Aaron Saenz, an AI scientist, once had an interesting metaphor: ANI, like the early amino acids on the Earth, might suddenly produce life.

In the technological development of AI, the three cornerstones are algorithms, data, and computing power, among which algorithms are fundamental. ML is a significant subarea of AI, and it is currently one of the core research directions in AI and data analysis (Zhang R and Wang, 2016; Zhou ZH, 2016). DL is a pivotal research direction in ML. Optimization algorithms are a part of AI algorithms, and have been comprehensively used to solve issues in the field of AI.

This section expounds the basic ideas and applications of AI algorithms from three aspects: ML, DL, and optimization algorithms (see supplementary materials, Section 4).

2.1 Machine learning

Tom Mitchell defined ML as follows: "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*." ML algorithms use a large number of statistical rules to settle optimization problems. According to the current mainstream classification methods with different training samples and feedback methods, ML can be separated into four groups: supervised learning, unsupervised learning (Bitaab and Hashemi, 2017), semi-supervised learning (Zhou XY and Belkin, 2014; Kunal and Dua, 2019; Gupta ARB and Agrawal, 2020), and reinforcement learning (RL) (Buşoniu et al., 2010; Hessel et al., 2018; Gronauer and Diepold, 2022) (see supplementary materials, Sections 1.1, 1.2, 1.3, and 1.4, respectively).

ML is a powerful method to realize AI, and it is also the earliest AI algorithm developed. Unlike the traditional rule-based design algorithm, ML determines the law from massive data and automatically learns the parameters needed by the algorithm. The Bayesian theorem is a kind of ML algorithm based on the historical data of similar events to obtain the possibility of occurrence.

2.2 Deep learning

Tom Mitchell defined DL as follows: "Deep learning is a particular kind of machine learning that achieves great power and flexibility by learning to represent the world as nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones." DL is one of the technical branches and subset in ML, through mainly the construction of deep artificial neural networks to learn knowledge. The input data are usually complex, large scale, and high

dimensional. DL has been said to be one of the most significant breakthroughs since the advent of ML.

Because DL methods are a popular research topic in the domain of AI algorithms, they are listed separately here. The ultimate goal is to make the machine have the same analytical learning ability as human beings and be able to recognize text, images, sound, and other data. With the high-speed expansion of graphics processing units (GPUs), a lot of GPUs working together with CPU, the application of DL to various fields has been accelerated (Pouyanfar et al., 2019).

The core idea of DL is to use a multi-layer neural network frame to obtain information from datasets. This method is much better than traditional ML methods in terms of learning ability. It has been used in many areas, including computer vision (such as video recognition, medical image analysis, and image classification), autopilot, natural language processing, speech recognition, handwriting recognition, and human-computer interaction. It continues to be extended to other fields. The DL framework based on neural networks includes mainly deep neural networks (DNNs), convolutional neural networks (CNNs) (Waibel et al., 1990; Krizhevsky et al., 2012; Long et al., 2015), recurrent neural networks (RNNs) (Socher et al., 2011a; Graves et al., 2013; Sutskever et al., 2014), deep belief networks (DBNs), autoencoder (AE), generative adversarial networks (GANs) (Socher et al., 2011b; Goodfellow et al., 2014, 2016; Ledig et al., 2017), and so on (see supplementary materials, Section 2).

2.3 Optimization algorithms

The classical optimization algorithms include linear programming, dynamic programming, the mountain climbing method, the fastest drop method, the simulated annealing algorithm (SAA), the tabu search algorithm, and the genetic algorithm (GA). These algorithms have the advantages of simple parameter settings, strong adaptability, no special requirements for the analytical properties of objective functions and the selection of initial parameters of the algorithm, and fast convergence. These algorithms are commonly used to solve practical engineering problems, such as linear programming, quadratic programming, and convex function optimization. The traditional optimization algorithm requires that the objective function be convex, continuously differentiable, and so on, and thus has a poor ability to deal with nondeterministic information. It has great limitations when directly applied to complex large-scale engineering problems.

The swarm intelligence (SI) optimization algorithm and group search optimizer algorithm are new kinds of optimization algorithm, inspired by a social behavior mechanism from insect groups and animals (Mohiuddin et al., 2016; Elbes et al., 2019; Shafiqur et al., 2020). The SI algorithm simulates the group behavior of insects, fish, herds, and birds. Each group always alters the search direction by learning its own experience and others, also known as the metaheuristic method. The group search optimizer algorithm is to optimize the problem by simulating the foraging behavior of the animal population. Compared with other optimization algorithms, such as GAs and the particle swarm optimization (PSO) algorithm, the group search optimizer algorithm is superior in exploitation ability. It has been comprehensively researched and used to solve problems of load economic dispatching in power systems in recent years, including multi-objective decision-making optimization (Park et al., 2010) and other large-scale complex engineering optimization and decision-making problems. For details, see Section 3 in the supplementary materials.

3 Application of AI algorithms to cyberspace security

With the development of network structures to a larger scale, complexity, concentration, and heterogeneity, cyberspace has become a complex system. Network attacks are large-scale, distributed, and hidden. The attack target extends from the traditional network system to other systems, such as IoT, industrial equipment, smart homes, and driverless system networks, which brings huge security risks to cyberspace. The traditional security defense no longer satisfies the "supervision, prevention, control, and evaluation" systematic linkage defense requirements for cyberspace security defense in the future. AI technologies have accelerated the development of cyberspace security technology. The research fields of cyberspace intrusion detection, resource allocation optimization and defense evaluation, defense decision-making, and strategy generation based on AI algorithms have

shown unprecedented advantages (Sagar et al., 2019; Zhang HY et al., 2019; Zeadally et al., 2020).

This section describes the different AI algorithms used to solve cyberspace security issues, including intrusion detection, prediction and evaluation of the security situation, and generating and optimizing strategies. References to important methodology, mainly from 2017 to 2022, are provided for each technique.

3.1 AI algorithm modeling for detection of network attacks

Network attack detection, also known as intrusion detection, is one of the significant approaches in cyberspace security. Network attack detection is a means to prevent networks from attack, by finding suspicious attacks and taking corresponding measures to reduce economic losses effectively. Researchers continue to introduce more advanced AI algorithms, as new technologies arise for making the model more efficient and more stable.

The approach includes several main steps: data preprocessing, feature processing, dataset selection, and index evaluation. In general, most researchers use existing, recognized, and widely used public datasets to train models. Examples are the KDD CUP99 dataset, NSL-KDD dataset, UNSW-NB15 dataset, and CICIDS2017 dataset. For the performance evaluation index of the model, which is adopted to explain the performance of the system generally, the following indexes are mainly used: accuracy, precision, recall, F score, true positive rate (TPR), false positive rate (FPR), and area under curve (AUC).

3.1.1 Modeling methods based on machine learning

Network attacks are growing more sophisticated and challenging. The establishment of an intrusion detection model is the first step in the implementation of cyberspace security defense actions and behaviors to better classify the threats of network attacks. ML has become the mainstream classification problem solving scheme due to its relatively simple architecture and low computing overhead. Currently, the use of ML in the sphere of intrusion detection is growing significantly. ML algorithms are used to explore and study a series of network attack classification issues (Hühn and Hüllermeier, 2009; Faker and Dogdu, 2019; Gu et al., 2019). Traditional ML algorithms, namely, K-nearest neighbor (K-NN), support vector machine (SVM), naive Bayesian (NB), and decision tree (DT), play a significant role in modeling for intrusion detection (Buczak and Guven, 2016; Nisioti et al., 2018; Mishra et al., 2019). The detection accuracy of the intrusion detection system (IDS) is increased, and the false alarm rate is lowered (Narudin et al., 2016; Al-Yaseen et al., 2017). The researchers apply these methods to a variety of cyberspace security problems, such as network intrusion detection, IoT attacks, smart grid attacks, and the "network-physical" attack of industrial control systems (ICSs).

Researchers apply the K-NN algorithm and two or more algorithms to design intrusion detection models, improving the detection rate, accuracy, and efficiency against new attacks (Aung and Min, 2018; Jain and Kaur, 2019). However, the classification performance of K-NN decreases significantly with the improvement of the characteristic dimension of cyberspace data. To settle this problem, Chen F et al. (2018) proposed a combined model including the tree seed algorithm and K-NN. The tree seed algorithm is used to generate the original data, and then K-NN is used to classify the effective features. This model can effectively remove redundant features and help increase detection accuracy.

Aiming at security threat issues in the IoT, Al-Omari et al. (2021) proposed an intelligent intrusion detection model with DT that overcame the shortcomings of the previous DT model for identifying the impurity of the features with a Gini index method to rank the security characteristics. Compared with traditional ML techniques (K-NN, SVM, logistic regression (LR), and NB), this model can effectively detect and predict network attacks and even reduce the computing time. To settle the question that the system integrity protection (SIP) scheme in a smart grid can be easily attacked, Wang PY and Govindarasu (2020) proposed a supervised learning model based on an SVM embedded layered decision tree (SVMLDT), which converts anomaly detection into a multi-class classification problem. Decentralized SIP has greater flexibility and resilience in the face of malicious attacks compared with traditional centralized protection.

Bayesian networks are helpful in inferring valuable information from uncertain events. The traditional Bayesian model can detect normal network attacks and conserve sensitive data, but cannot detect a

new attack threat (Olowononi et al., 2021). Sapavath et al. (2021) proposed an intrusion detection method based on AI algorithms for the dynamic detection of network attacks. In this model, three adaptive agent models with system immunity are designed to observe the private agent of user activity in cyberspace, detect the suspicious user of public agents, and locate a malicious user of mobile agents, separately. This AI algorithm is trained using the attack data in the cloud environment, and the transmission power is used as the threshold for classifying suspicious behaviors. The numerical simulation outcomes show that the Bayesian model is superior to the stochastic forest model and the DNN model in terms of accuracy, precision, recall, and learning time. Specifically,the accuracy is 99.8% for the Bayesian network algorithm, 99.1% for DNN, and 89.2% for the random forest (RF).

Researchers improved the *K*-means algorithm to decrease the detection delay in IDS and improve the classification performance. To solve the problems of distributed denial of service (DDoS) detection, Gu et al. (2019) used a semi-supervised *K*-means algorithm using hybrid feature selection (SKM-HFS). This model is at a higher level with respect to the specification in detection display and sequential preference technology; it also has the smallest detection delay compared with the K-NN algorithm and NB algorithm. Gu et al. (2019) made a performance comparison of different feature selection methods using different datasets. The Caida "DDoS attack 2007" dataset, DARPA DDoS dataset, CICIDS "DDoS attack 2017" dataset, and a real-world dataset were applied to change the data test. The algorithms above perform only one-level classification for the original data, which is not ideal. Al-Yaseen et al. (2017) provided a multi-level intrusion detection model. The original training dataset was optimized to decrease the training time spent by improved *K*-means. SVM and a limit learning machine were used for multi-level classification, and the presentation of classifiers was significantly improved. The accuracy was 95.75% on the KDD CUP99 dataset.

Combining the game theory with the ML algorithm to design an intrusion detection model and then describing the behavior of attackers and defenders are a hot topic. Game research on network attacks and protection is in the forefront. Burke (1999), Lye and Wing (2002), and Liu P and Zang (2003) have used incomplete information games, stochastic games, and other models to assess the objective, purpose, and rules of intrusion. In 2017, Liu XX established the attack and defense game method using a hybrid strategy game theory. It was assumed that the attacker and the defender can obtain all information. However, the real situation is that the attack and defense parties can obtain only part of the other party's information using this model. In 2020, based on the previous model, the team built a hybrid attack defense game model based on the Bayesian theory with incomplete data to build a model of threat propagation between two nodes (Liu XX et al., 2021). The attacker and the defender also acquired parts of the data of each other and attempted to infer the behavior of the other party. By processing the refined Bayesian Nash equilibrium, a quantitative analysis method for the "network-physical" attack of the ICS was obtained. Time overhead was used to evaluate the model performance. The calculation results in ICSs showed that when the total number of nodes was smaller than 50, the time cost did not exceed 0.1 s. Even if the number of nodes increased to 120, the time was still less than 0.5 s. When the number of nodes increased from 10 to 200, the time increased slowly. The results showed that the model can achieve higher efficiency with fewer nodes.

To make reasonable use of unmarked security data (such as system logs), avoid the shortcomings in existing tag training data, and enhance the detection ability, Nishiyama et al. (2020) proposed a new semi-supervised learning algorithm involving large-scale unlabeled logs (SILU). SILU can be used alone or added to any kind of classifier of supervised learning methods (such as LR, SVM, and RF), and is a supplement to the existing supervised learning methods. It has lower overhead and better outcomes than traditional supervised learning. It can not only improve detection ability but also suppress the error marks of data.

3.1.2 Modeling methods based on deep learning

In recent years, DL models such as neural networks have become effective solutions for classification tasks because of their ability to generalize more complex task features. Researchers have made much anomaly analysis for intrusion detection with ML and DL models. They provided a large amount of data to compare

the classification performances of different algorithms. The performance of intrusion detection has thus been greatly improved. Some scholars used a CNN algorithm to design an intrusion detection model for supply chain network attacks in power systems, IoT, computer networks, unmanned aerial vehicles (UAVs), and web applications, and further improved the detection accuracy.

Aiming at supply chain network attacks in power systems, Khaw et al. (2021) proposed a transmission line relay protection IDS with a one-dimensional (1D) CNN based AE, and the accuracy reached 100%. In view of the network attacks of IoT, Ullah et al. (2019) designed a TensorFlow deep convolution neural network (DCNN) model for malware data analysis. The marked and weighted features were used to filter the noisy data, leading to higher identification performance in large-scale malware detection and less time cost. Ullah et al. (2019) made a comparison of the DCNN models based on different image pixels. The results showed that the higher the image resolution, the higher the accuracy and precision. In addition, they compared the performances of the DCNN model and hybrid ML algorithms. The accuracy of DCNN was 97.46%, and the *F* score of DCNN was 97.44%, which were much higher than those of hybrid ML algorithms, such as CLGM+SVM, LBP+SVM, and GIST+SVM. Unfortunately, DCNN took a little longer time.

Following the DCNN method, Khoa et al. (2020) applied a new model deployed on the IoT gateway based on collaborative learning. In the new intelligent "filter" based on DNN, every filter uses the data included in cyberspace to teach the DNN algorithm to keep network attacks away in real time. The most significant characteristic of this model is that the filter shares the trained detection model with others, rather than exchanging the actual data. It can maintain the data privacy of multiple subnets and greatly enhance the detection accuracy and learning speed while reducing network traffic. Another DL model was proposed by Roopak et al. (2019), using different datasets to analyze and compare multiple algorithms. It was concluded that all DL models tested were better than ML models in network attack detection, such as SVM, Bayesian, and RF, except multilayer perceptron (MLP) (Roopak et al., 2019). In addition, Khoa et al. (2020) made a performance comparison of *K*-means models over three traditional datasets (KDD, NSL-KDD, and UNSW-NB15), showing that the classification accuracy was increased by 14.76% and that the communication overhead was reduced by 98.5% compared with *K*-means in the KDD dataset.

Parameter tuning based on DL technology in different environments is time consuming. To mitigate this problem further, Chen Y et al. (2022) proposed a multi-objective evolutionary CNN (MECNN) method in fog computing, providing low delay and high accuracy for IoT. Compared with ML and some hybrid DL models, it was shown that the MECNN model had good robustness while improving the detection performance. However, this model had some problems, and the accuracy will decline slightly due to the imbalance of training data. Balamurugan et al. (2022) designed a IDSGT-DNN framework in cloud computing. Compared with some hybrid algorithms and optimization algorithms, the results showed its better performance in accuracy, detection rate, precision, *F* score, AUC, and FPR on the CICIDS2017 dataset. The accuracy was 98.65%, and the *F* score was 99%.

The application of computer networks focuses mainly on the detection, modeling, and calculation of network attack behavior and safety issues (Kim et al., 2019; Roopak et al., 2019; Yeom and Kim, 2019). Ho et al. (2021) proposed a model that separates all packet traffic from benign or malicious classes in CNNs, giving an innovative dataset preprocessing method to improve the multi-class classification performance of IDSs based on CNNs. The model consists of two convolution layers and two fully connected layers. The input of the model is a batch of matrices composed of 77 features. In the model, all super parameters are selected from a subset of manually specified values, and an iterative evaluation model is used to search for the best combination of parameter values. Ho et al. (2021) improved the multi-class classification performance and achieved an accuracy of 99.78% on the CICIDS2017 dataset, higher than those of nine other well-known classifier models.

Hossain et al. (2020) established an intrusion detection model based on long short-term memory (LSTM) for secure shell (SSH) and file transfer protocol (FTP) brute force attacks (BFAs), to solve dictionary-based BFA detection in network traffic detection. They made a comparison of the LSTM model and other

ML algorithms, such as J48, NB, K-NN, DT, and MLP. Although the classification accuracy of their model was as high as 99.88%, they conducted only characteristic studies and extractions for SSH and FTP attacks.

In the study of extracting elements of cyber threats from unstructured network text, Ma et al. (2021) proposed a novel cybersecurity entity identification model based on bidirectional long short-term memory with conditional random fields (Bi-LSTM with CRF) to extract security-related concepts and entities from unstructured text, named XBiLSTM-CRF. The method contains a Bi-LSTM layer, a word embedding layer, and a CRF layer, and concatenates X input with Bi-LSTM output. The Bi-LSTM model with CRF connects word embedding with Bi-LSTM, which greatly improves the detection ability without increasing the complexity when compared with the traditional Bi-LSTM model. XBiLSTM-CRF can identify six categories, including software, network, attack, file name, hardware, and modifiers. By extracting the network threat and constructing a network threat knowledge map, which helps cyberspace security technicians understand the nature of cyberspace security threats, the accuracy was up to 90%.

In the application of the algorithm combined with DL, researchers combined the traditional supervised learning algorithm and DNN, which enhances the ability of intrusion detection. To resolve the problem that the available intrusion systems cannot accurately detect the threat of unknown sources, anomaly analysis models were designed and implemented combining the K-NN algorithm with the DNN algorithm for system intrusion classification based on the CICIDS2017 dataset (Atefi et al., 2019). The results indicated that this model performs better in anomaly detection and classification. Issa and Albayrak (2021) established a new DL classification method based on CNN and LSTM, named CLSTMNET, for persistent network attacks. The detection accuracy was 99.28% in the NSL-KDD dataset, higher than those of the traditional methods, such as SVM, MLP, DT, RF, NB tree, NB, and J48.

To further detecte various network threats to industrial "network-physical" systems, Li BB et al. (2021) designed a new DL method with a CNN and gated recursive unit (GRU), which was followed by an MLP module, and then a softmax layer. The CNN module consisted mainly of three convolution blocks. The results of numerical analysis showed that this model is superior in its main parameters. When the communication number R increased from 1 to 10, this model's ability was greatly improved, and the gradient was stable when R was large enough. The model was effective in detecting different network threats in industrial "network-physics," including denial-of-service (DoS), reconnaissance, response injection, and command injection attacks.

The combination of DL and RL is widely used to settle the question of high-dimensional decision-making, and research on this kind of combination algorithm is still in the primary stage. Among such algorithms, deep Q-network (DQN) has been a major subject of investigation in recent years. To improve the reliability of IDS for UAVs, DQN has been used to build a multi-class classifier for training, and the custom reward function has been used to consider unbalanced datasets. A method of regular offline learning was proposed to ensure that the UAV can independently learn to adapt to the evolution of intrusion attacks (Bouhamed et al., 2021), so that the UAV can detect suspicious activities independently and take necessary actions to ensure safety.

To ensure the safety of unprotected web applications, Tekerek (2021) proposed an anomaly-based web attack detection model based on the DL method. This model is composed of data preprocessing and CNN steps, which can effectively settle the issue that the extracted features cannot be classified in traditional ML anomaly detection research. After certain training on CSIC2010v2 HTTP datasets, their method achieved better results than NB (Nguyen et al., 2011), SVM, and CNN (Zhang M et al., 2017)—while maintaining a lower FPR, the accuracy reached 97.07%.

In recent years, the problem of data imbalance has attracted research. Aiming at the imbalance problem of attack types in datasets, AE has been used for feature engineering and learning. Kunang et al. (2019) built an AE model for feature extraction, which had an overall accuracy reaching 86.96% and a precision of 88.65% on the NSL-KDD dataset. Kherlenchimeg and Nakaya (2018) proposed a sparse AE and RNN with accuracy reaching 80% on the NSL-KDD dataset. Mushtaq et al. (2022) proposed a hybrid framework AE-LSTM, which obtained features using AE and

LSTMs for classification. The results showed higher accuracy and smaller prediction error on the NSL-KDD dataset when compared with other ML or DL techniques. On this basis, Shaikh and Shashikala (2019) proposed a stacked AE with an LSTM network for only DoS attacks, in which the overall accuracy reached 94.3% and FPR reached 5.7% on the NSL-KDD dataset. Similarly, Qazi et al. (2022) proposed a stacked non-symmetric deep autoencoder (S-NDAE) model, which achieved an accuracy of 99.65% and a precision of 99.99% on the KDD CUP99 dataset. Hindy et al. (2020) proposed an AE method for zero-day attacks. The accuracy reached 90.01%, 98.43%, 98.47%, and 99.67% for DoS (GoldenEye), DoS (Hulk), port scanning, and DDoS attacks on CICIDS2017, respectively.

To further settle the problem of accuracy reduction caused by dataset imbalance, researchers began to apply GAN technology to dataset feature enhancement. Gupta et al. (2022) applied a CSE-IDS model based on cost-sensitive DL and ensemble algorithms, which was evaluated on three well-known datasets with accuracy, recall, precision, *F* score, receiver operating characteristic (ROC) curve, AUC value, and computational time. Ding et al. (2022) proposed a tabular auxiliary classifier GAN (TACGAN) model; they added two loss functions in the generator to compute the information loss. The accuracy of TACGAN was 95.86% for the CICIDS2017 dataset, 92.39% for the UNSW-NB15 dataset, and 93.53% for the KDD CUP99 dataset. The method achieved excellent results compared with nine other algorithms. Huo et al. (2022) proposed a model of combining learning with DL, light gradient boosting machine (LGBM), which improved the accuracy to 78.64%.

Aiming at the analysis of ML and DL algorithms, researchers have conducted many experiments and comparisons on several well-known public datasets. Atefi et al. (2019) and Yeom and Kim (2019) tested the performance of NB, SVM, and CNN based classifier on the CICIDS2017 dataset. After comprehensive evaluation, Atefi et al. (2019) and Yeom and Kim (2019) concluded that CNN and SVM generally have high detection rates. In addition, CNN was better than SVM with respect to the processing time. However, Atefi et al. (2019) and Yeom and Kim (2019) observed that CNN had medium performance on datasets with many labels. Atefi et al. (2019) verified through experiments that the DL algorithm performed better than the ML algorithm and that the performance of DNN was obviously better than that of K-NN, with higher accuracy and lower CPU time. Andresini et al. (2020) established an intrusion detection model named MINDFUL. This model combines the unsupervised phase of multi-channel feature learning with a supervised phase using cross-channel feature dependency. Through flexible and effective learning, the difference between normal flow and attack flow can be better reduced. Using three benchmark datasets to test the difference performance of multiple algorithms, including MINDFUL, nearest neighbor (NN), artificial neural network (ANN), CNN, and advanced CNN (ACNN), it was shown that MINDFUL usually had high accuracy and *F* score on three traditional datasets. The accuracy reached 97.9% on the CICIDS2017 dataset.

3.1.3 Modeling methods based on optimization algorithms

Because different feature selection methods will affect the network attack detection accuracy, the SI optimization algorithm began to be widely used to enhance the performance of feature selection. Seth and Chandra (2018) established a K-NN and a modified grey wolf optimization (MGWO) method for cloud, and the method is effective for feature selection; the accuracy reached 99.87% on Solaris datasets and 98.94% on Windows datasets.

Moizuddin and Jose (2022) recommended a generalized mean GWO (GMGWO) method, which can obtain an average accuracy of 99.07% with a lower training time on two publicly available datasets (NSL-KDD and BoT-IoT). Chohra et al. (2022) combined a Chameleon model based on an SI optimization algorithm with ensemble learning techniques. The ensemble learning classifiers were used as fitness and evaluation functions for each individual in the population, and the anomaly detection AE was used to obtain excellent models by an iterative update. The simulation results showed that the best *F* score reached 97.302% on the IoT-Zeek dataset.

The optimization algorithm can be combined with the DL algorithm to optimize the parameters to improve the network attack detection accuracy of the model. Kan et al. (2021) used a new adaptive particle swarm

optimization CNN (APSO-CNN) method to accurately detect diverse types of IoT network attacks. In PSO, small fitness values were searched by updating the speed and position of the particle swarm. The possibility of PSO falling into a local optimum was reduced by adaptively changing the inertia weight through the fitness value. Kan et al. (2021) compared the APSO-CNN algorithm with three commonly used algorithms (SVM, R-CNN, and FNN). The accuracy was 96%, much higher than those of the other methods. This model is reliable and has been successfully used for many types of intrusion detection tasks.

3.2 Modeling of AI algorithms for security situation assessment

An assessment of the security situation is performed to describe the security of the network operation. At present, with the existing research modes and algorithm frameworks in the field of network security assessment, there is no outstanding optimal solution. The network security situation assessment consists of several stages: situation acquisition, situation analysis, and situation assessment. Among them, situation acquisition and situation analysis refer mainly to obtaining the classification results of network attack threats and identifying the threats with AI algorithms, which is similar to network intrusion detection.

Researchers typically combine AI algorithms with different stages of situation assessment to improve model performance. Through a series of mathematical models and algorithms to extract information from massive network security data, the model can obtain the macro network security situation correctly, so that network managers can make decisions and take protective measures in advance and provide a basis for the next situation prediction. Different from intrusion detection, only when network attack data detection and classification are completed, can network security situation assessment be carried out.

In general, the results of the network situation assessment model are directly related to the attack probability, attack type, sample number, influence value of each attack type, and number of occurrences of each attack obtained from situation classification analysis. The attack probability and the impact of various attacks will be calculated according to the test classification results. The network security situation value will also be calculated and evaluated. A common vulnerability scoring system (CVSS) is usually given. In this way, the current network security status and level are judged, and the management personnel can fully grasp the current network security situation and take timely measures.

Researchers believe that it is necessary to build a reasonable and objective situation indicator system according to certain principles. This is of great significance for the assessment to reflect whether the results of network security situation assessment are scientific and reasonable. Some researchers established the index system from the perspectives of vulnerability, threat, disaster tolerance, and stability, which can fully reflect the security risks and operating state of the network system (Zhang R et al., 2022).

3.2.1 Modeling methods based on machine learning

The ML algorithms, such as Bayesian algorithms, are widely used for cyberspace security risk assessment, network vulnerability assessment, and network reliability assessment. Meanwhile, the hybrid methods, such as ML algorithms combined with SI optimization algorithms, are the main research directions in cyberspace security defense prediction and assessment.

To calculate the probability of the cyberspace security risk, researchers proposed a risk evaluation model combining an SI optimization algorithm with ML algorithms. Li DT et al. (2021) established a new model based on the fruit fly optimization algorithm and SVM (FOA-SVM). The model can evaluate the various factors affecting cyberspace security quantitatively. The calculation formula is $S=iT+jV+kR$, where T is the threat index, V is the vulnerability index, R is the asset state index, and S is the cyberspace security situation index, i, j, and k are the cyberspace security impact factors corresponding to each index, and their values are determined according to expert scoring method. The accuracy of the FOA-SVM model was 81.2%, the AUC value was 83%, and the F score was 83%. Kumar VS and Narasimhan (2021) established the economic risk evaluation model for virtual power plants by using the NB algorithm and CRQ-J48 algorithm. The security risk equation is $R=f(T, V, C)$, where T, V, and C represent the threat, vulnerability, and cost, respectively. The prediction accuracy of the NB algorithm and CRQ-J48 model was 82% in the simulation data.

For comprehensive vulnerability evaluation of cyberspace security, it is a common pratice to combine the Bayesian algorithm with other algorithms to evaluate the system security. By establishing a reasonable evaluation index frame and evaluation algorithm, the cyberspace security state can be quantitatively represented, which can provide technical guidance for network system designers. Chen SS et al. (2008) proposed a vulnerability state assessment method with a Bayesian network. By analyzing the relationship between the Bayesian network and the fragile state structure, four multi-angle evaluation indexes, namely network reliability, vulnerability point criticality, minimum order minimum path set, and minimum cut set of the vulnerability state map, are established. The network attack is simulated and analyzed by using the sensor protocol for information via negotiation (SPIN) model verification tool, which verifies that this algorithm and this evaluation index set can reflect the cyberspace security state correctly. Mehta et al. (2006) calculated the possibility of arrival of various security states in the attack graph using a Bayesian network, and calculated the threat degree of the corresponding attack scene using these possibilities.

A reliability evaluation method based on state-space classification (SSC) and sequential Monte Carlo simulation (MCS) was proposed to evaluate the reliability of distribution network systems (Li GF et al., 2020). This method was applied as a supervised learning algorithm to conduct the classification calculation, and the empirical model of reliability evaluation was established by replacing the traditional emergency analysis method based on continuity. The model can avoid the workload of topology analysis. The results showed that the framework ensured the evaluation accuracy. Pu (2020) proposed a complex attack assessment method based on a dynamic Bayesian network. It can comprehensively evaluate the attack capability of network nodes. This method combines the node information and observation data of multiple nodes related to the attack and obtains high attack accuracy and efficiency.

Luan and Tan (2021) proposed the EWM-IFAHP model, and the DT algorithm was used to extract elements. The authors built the fuzzy evaluation matrix, including the vulnerability and seriousness of attacks, the support degree, and the asset value. The EWM-IFAHP model had better recall rate and an accuracy of 90%. It can even have a higher generalization ability when compared with other methods on the KDD CUP99 dataset.

3.2.2 Modeling methods based on deep learning

DL algorithms with neural network algorithms as the core are widely used in the cyberspace security field to handle high-dimensional and massive threat data owing to their self-learning, adaptive, and nonlinear processing characteristics.

Yang HY and Zeng (2021) proposed a network security situation assessment model named deep autoencoder (DAE) CNN with the under-over sampling weighted (UOSW) algorithm (DAENDD(UOSW)) based on DL. To resolve the problem of extreme imbalance of classification results for different datasets, the UOSW algorithm was proposed to deal with datasets, and DAE was used to classify network attacks. After the classification of network attacks, the network security situation was quantitatively evaluated. In particular, DAENDD(UOSW) improved the recall rate and accuracy with a small number of data samples for training.

The uncertainty and nonlinearity of situation data pose a great challenge to the accuracy of the situation prediction model. Research in recent years has tried to solve this problem using optimization algorithms. To improve the performance of a situation prediction model, Hu CH et al. (2021) proposed a model combining the sparrow search algorithm (SSA) and the simplicial algorithm (SA) named SA-SSA, which had better performance than traditional situation prediction models; it improved the accuracy of the neural network model and minimized the training loss. The results showed that the method can accomplish the task of situation prediction with better capability on KDD CUP99.

The serious imbalance of situation data brings great challenges to the accuracy of situation prediction models. At present, researchers are beginning to solve this problem using optimization algorithms. Zhang R et al. (2021) proposed an SSA with back propagation (SSA-BP) model. The simulation results showed that the model can achieve higher accuracy on the KDD CUP99 dataset and less training loss than the traditional model. To improve further the accuracy and convergence speed of the model, an SAA-SSA-BPNN

model was proposed (Zhang R et al., 2022). This model converges fast and does not easily fall into local optima. This model can evaluate the threat degree of the network system. The comparison test showed that this evaluation model has higher accuracy and higher convergence speed than other improved BP models. To better master the dynamic changes of network security, Zhang ZQ (2021) proposed a combination of genetic simulated annealing and BP (GSA-BP) to enhance the BP neural network. The square error was 0.0146, and the model can prevent the BP neural network from falling into local minima.

Facing massive network attack data, traditional methods cannot meet the needs of huge data processing. The DL algorithm shows its advantages. Yang HY et al. (2021) proposed an NSSA model based on adversarial DL, combining DAE and DNN. The DAE network is used to extract features and classify network attacks. Experimental redults showed that the model can identify network attacks accurately and evaluate the network status on the NSL-KDD dataset more comprehensively and flexibly. Yang XJ and Jia (2021) used the improved PSO algorithm to establish a new network structure and select the optimal network parameters for the LSTM neural network to reduce human subjectivity. Diao (2021) proposed an improved NAWL-LSTM method. The improved LSTM neural network was used to analyze and process network security situation data, making effective use of the attack logic contained in sequence data. This prediction model can reduce the complexity of implementation and improve the stability. Su (2021) proposed a hybrid optimization deep reinforcement learning (DRL) method. A differential time-series prediction structure was created, and a normalized DRL prediction model was built. The experimental results showed that, compared with the time-domain analysis test group, DRL had a small prediction error, good prediction effect, and huge application potential.

In view of the existing network security situation assessment methods, there are some problems such as difficulty in extracting feature elements, low accuracy, and poor timeliness. Wang JH et al. (2021) established a new model with a GA probabilistic neural network (GA-PNN). The correction factor of PNN was optimized by GA-PNN to prevent the convergence speed from being too low, and then PNN training was carried out to obtain a stable model. In contrast to the traditional model, GA-PNN had a higher training speed and higher evaluation accuracy, with an average accuracy of more than 90% and a maximum accuracy of 98.46% on the KDD CUP99 dataset. Yang HY et al. (2022b) proposed a PFEN-ABiGRU method based on parallel feature extraction and improved BiGRU. The improved model was used to detect network threats, and the network security situation value was calculated. Zhang R et al. (2022) developed another deep weighted feature learning (PSAE-ATBiGRU) method to detect network threats. The model has three parts: data preprocessing, PSAE-ATBiGRU network threat detection, and network security situation assessment. Results showed that the proposed method achieved the highest accuracy of 82.13% on the test set, and that the recall rate and F score reached 83.36% and 82.74% respectively on the NSL-KDD dataset. It can efficiently and comprehensively evaluate the overall situation of network security compared with classical situation assessment methods such as SVM, LSTM, BiGRU, and AEDNN. Ye and Tan (2019) and Yang HY et al. (2021) used the DL algorithm for evaluation and obtained higher accuracy. However, the implementation efficiency of the algorithm was sacrificed, leading to a longer evaluation time. Wei MH (2021) proposed GRU-RNN by combining a particular cyclic neural network with a threshold regression element. It can describe the information security characteristics of time-series data much better than traditional models. This model synchronously abstracts the internal and external security situation features from the network's time-series data as security prediction features, inputs them into RNN for training, and trains RNN through a time-series error BP algorithm. Through iterative optimization, although requiring more training time, it can obtain higher evaluation accuracy and robustness for a complex network structure of nonstationary data.

3.3 AI algorithm modeling for defense strategy optimization

Although the optimization algorithm is not the main algorithm for intrusion detection or risk evaluation, it is always used to optimize some parameters and extract data features. The application of optimization algorithms in cyberspace security focuses mostly

on cyberspace security defense strategy optimization, including security defense decision-making optimization and security risk strategy allocation optimization.

One approach is to solve the network defense strategy optimization problem as a multi-objective optimization problem. Some researchers think that the ultimate goal of cyberspace security defense strategy optimization is to guarantee the confidentiality and integrity of the network system. Cyberspace security defense strategy optimization is a the multi-objective optimization problem. Hu BW et al. (2021) proposed a decentralized consensus decision-making (DCDM) method for cyberspace security protection in multimicrogrid (MMG) systems. It combines a fuzzy static Bayesian game model (FSB-GM) to acquire the optimal strategy with a hybrid consensus algorithm to obtain consensus. It ensures consistency and non-repeatability. Aiming at the cyberspace security defense strategy optimization problem of the IoT, Hamrioui and Bokhari (2021) proposed a model using the combinatorial optimization method and combining it with the actual situation of the backpack. This model is based on the optimization of two objective functions: one is to maximize the cyberspace security risk detection capability to resist all kinds of network attacks, and the other is to minimize the cost of construction and deployment. The results showed that this model can not only minimize the cost and budget, maximize the security level, and settle the cardinality constraint problem at the same time, but also trade-off among security, cost, some necessary security assets, and the budget. When compared with non-dominated sorting genetic algorithm-II (NSGA-II), the experimental results showed that this method produced higher security levels and lower costs in a reasonable time frame.

Researchers optimize the security strategy allocation by combining different optimization algorithms. In particular, there is a tendency to enhance the effectiveness of combinatorial optimization algorithms. Another popular solution, however, is to use game theory and incorporate cognitive models of human attackers to formulate the defense strategy by dynamic evolution of the network. Dealing with the uncertainty of network attack behaviors and the optimal allocation of cyberspace security resources, Hyder and Govindarasu (2020) proposed a game theory optimization framework in power grid deployment. By modeling various attacker configuration files and various practical characteristics of defenders, the optimal strategy of smart grid cyberspace security infrastructure investment was obtained. Game theory has been applied by some researchers to the generation of network defense strategies, and good results have been obtained. Bhuiyan et al. (2021) established a risk aversion "defender-attacker" stochastic Stackelberg game model with an attack graph and proposed an accurate algorithm, solving stochastic programming with three different acceleration algorithms PBA_NL, PBA_All_ACC, and PBA_TrSsNl. Numerical simulations showed that the average calculation time was reduced by 71%. Compared with deterministic and risk-neutral models, it can provide better network interception decisions.

Attack graph is a model-based network security vulnerability assessment technology (Touhiduzzaman et al., 2019). Using this technology, defenders can analyze the interdependence between host vulnerabilities, and thus obtain all possible attack paths of intruders and present them in the form of a tree graph. Stevens-Navarro et al. (2008) introduced game theory graph coloring to decide the premier allocation model with security path diversity. This model introduces a game theory strategy based on a graph coloring game, putting forward the best diversity by expanding the security index and increasing the attack acting on grid network assets. A limited number of software package security intensities (action: color) are assigned to make many security paths (player: node) so that the security of network assets (income: security index) of the whole power grid is increased. The introduction of the game theory strategy increases the heterogeneity of the safety mechanism, provides better security indicator allocation capability, reduces the attack propagation of the whole power network, and ensures the highest level of cyberspace security.

Defense strategy optimization in cyberspace security has increasingly relied upon stochastic game processes that combine game theory with a Markov decision process (MDP). On the basis of the game theory, some researchers have added DRL algorithms to the decision-making process and thus can use online learning and make adaptive adjustments to increase the learning rate as the defender collects additional historical data. Stochastic game process, which combines

game theory with MDP, is more and more widely used in defense strategy optimization. Researchers add a DRL algorithm to the decision-making process to provide the best decision support for network managers.

Liu XH et al. (2021) constructed a stochastic game model, which used RL to solve the game equilibrium in stochastic dynamic network systems by introducing a deep recurrent Q-learning (DRQN) algorithm and defense decision-making algorithm with online learning capability. Compared with the most advanced equilibrium solution method, the proposed DRQN algorithm can achieve the optimal defense strategy faster. In addition, some people believe that due to the current diversity of cyberspace threats, it is necessary to further understand the attackers' attack decisions. This kind of defense decision-making algorithm usually specifies the following assumption: attackers are rational decision makers, and they will take the best action every time they attack. To verify the rationality and effectiveness of the defense decision-making strategy proposed based on this assumption, Aggarwal et al. (2022) proposed a cognitive model based on instance-based learning theory to represent and predict the decision deployment of attackers in this task. The experimental results showed that the model can accurately capture the data of the attacker and make correct decisions.

4 Comparison and discussion

In the process of investigation and analysis, a total of 152 academic papers were collected, from which 85 papers were selected for analysis and comparison. Each paper is based on AI algorithms for solving specific problems in the field of network security. The collection and collation of paper materials involving network security and AI algorithm related fields come mainly from different kinds of journals and conferences, such as Elsevier, Springer, and IEEE journals, and international academic conferences. Most of them are indexed by Web of Science, and the authors are influential scholars in the fields of AI and cyberspace security around the world. Among them, 19.70% are from Elsevier journals, 6.58% are from Springer journals, 26.40% are from IEEE journals, and 25% are from conferences. Fig. 1 depicts the percentage of papers selected from different sources.

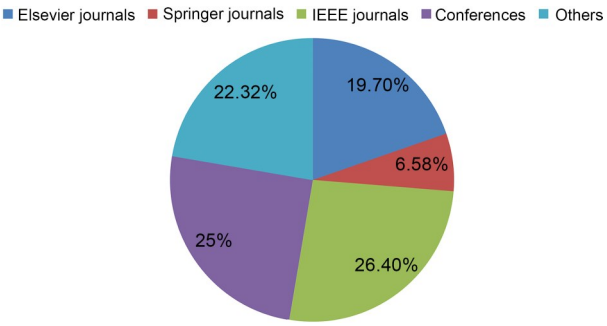

Fig. 1 The percentage of papers chosen from different sources

To summarize the effectiveness and advancement of leading technologies and their further development, we assessed the coverage of three aspects including intrusion detection, security situation assessment, and defense strategy optimization in specific issues of the literature mainly from 2017 to 2022.

Fig. 2 shows the relationship between the proportion of the number of papers and the country that the first affiliation belongs to from 2017 to 2022. This statistic further illustrates how much attention different countries have paid to security issues in cyberspace over a period of time. Aggregated data are from 17 countries: China, USA, Canada, India, Malaysia, Japan, Turkey, UK, Pakistan, Jordan, Vietnam, Italy, Nigeria, Saudi Arabia, Botswana, the Netherlands, and France. It can be seen that China pays the highest attention to cyberspace security. China's papers account for 47.62%, USA 14.29%, and Canada 8.58%.

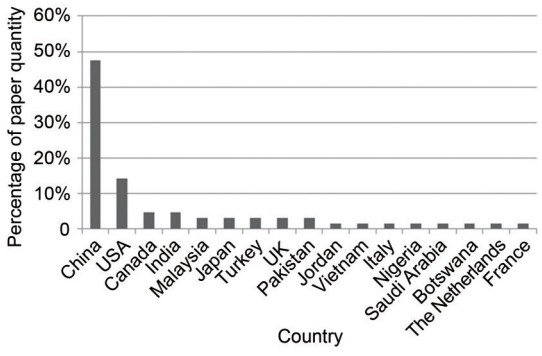

Fig. 2 Analysis of research situation in different countries

As can be seen in Fig. 3, there are six countries that pay high attention to security issues in the three types of cyberspace, namely China, USA, Canada, Malaysia, Japan, and India. Among them, China's

attention level is the highest, especially on the issue of security situation assessment, and the attention level reaches 90%. In contrast, USA pays more attention to defense strategy optimization.

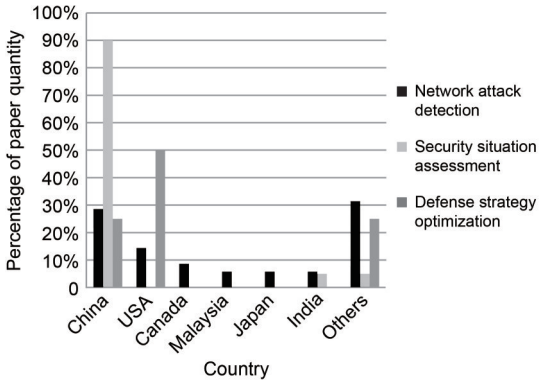

Fig. 3 Analysis of research directions in different countries

4.1 Performance analysis of AI algorithms in cyberspace security

4.1.1 Performance analysis of AI algorithms in network attack detection

The performance of the network attack detection method depends mainly on the dataset and several evaluation indexes. The accuracy, precision, recall, *F* score, and other indices are included in the evaluation index set. With the analysis of the data in Table 1, it can be concluded that the application of AI algorithms has many benefits to cyberspace security. Specifically, traditional ML and DL algorithms are widely used in intrusion detection.

The differences between ML algorithms and DL algorithms and between different datasets lead to different evaluation results. According to these results, the differences have mainly two reasons. We have analyzed these, and the conclusions can provide inspiration for researchers.

The first reason is the data processing types of the algorithm. The main processes of the traditional ML algorithm are dataset preparation, data preprocessing, data segmentation, definition of the neural network model, and network training. As described above, it takes much time to collect the data, screen the data, try different feature extraction algorithms, and combine different features to classify and regress the data. Further, researchers are committed to designing novel algorithms and methods for feature selection. Simulated annealing and GAs are classic feature selection algorithms. In addition, there are many methods based on evolutionary algorithms and random methods (for example, Monte Carlo). However, compared with the traditional ML algorithms, DL reduces the effort of designing feature extractors for each problem. Training often takes a long time, while testing takes a relatively short time. Therefore, DL relies more on high-performance machines with GPU (Xiao et al., 2021). With cyberspace attack behavior becoming more and more diversified, the explosion of traffic data inevitably leads to serious imbalance between normal data and data in the network attack. Traditional ML algorithms are over-reliant to artificial feature extraction methods in terms of performance, and are not suitable for mining the inherent laws of the dataset. In particular, traditional ML algorithms fail to consider the characteristics of time and space of network traffic and do not analyze the correlation of data in different dimensions, which makes them difficult to predict potential threats. In contrast, DL algorithms do not need to mention data features but carry out high-dimensional abstract learning of data automatically through neural networks, which reduces the synthesis of feature engineering and saves time. In fact, DL algorithms are suitable for dealing with big data. When the amount of data is relatively small, it may be more appropriate to use traditional ML methods. It can also be found that the network intrusion detection model based on the DL algorithm has a lot of barriers. The parameter adjustment processes, such as defining the neural network model structure, confirming the loss function, and determining the optimizer, are more burdensome and complex. DL models are inherently black boxes, so it is hard to ensure the main routes to improve the performance of the models. In addition, models based on DL are focused on a specific threat, which may greatly reduce their prediction accuracy, or even lead to missing attacks. So, some comprehensive solutions generalize or integrate different DL methods to deal with the large scope of the types of attack. Yet, there is a clear trend that as datasets become more complete, neural network algorithm models based on DL for network intrusion detection will achieve higher accuracy and efficiency (Bdrany and Sadkhan, 2020; Shende and Thorat, 2020).

Table 1 Performance comparison of multiple types of network attack detection models

Reference	Algorithm(s)	Dataset(s)	Accuracy	Precision	Recall	*F* score
Al-Omari et al., 2021	DT	UNSW-NB15	98.00%	97.00%	97.00%	97.00%
Sapavath et al., 2021	NB	The known attacks and novel attacks	99.80%	100.00%	96.00%	97.20%
Al-Yaseen et al., 2017	Improved *K*-means	KDD CUP99	95.75%	–	–	–
Khaw et al., 2021	1D-CNN+AE	OPAL-RT HYPERSIM	100.00%	–	–	–
Ullah et al., 2019	DCNN	Google Code Jam	97.46%	–	–	97.44%
Andresini et al., 2020	MINDFUL	CICIDS2017	97.90%	–	–	–
Ho et al., 2021	CNN	CICIDS2017	99.78%	–	–	–
Hossain et al., 2020	LSTM	CICIDS2017	99.88%	–	–	–
Ma et al., 2021	XBiLSTM-CRF	Open-source NER	–	90.54%	88.26%	89.38%
Atefi et al., 2019	K-NN and DNN	CICIDS2017	96.29%	96.43%	–	96.53%
Issa and Albayrak, 2021	CLSTMNet	NSL-KDD	99.28%	–	–	–
Li BB et al., 2021	CNN-GRU	Industrial CPS	99.20%	99.85%	97.36%	98.10%
Kan et al., 2021	APSO-CNN	Public IoT	96.00%	97.00%	–	–
Bouhamed et al., 2021	DQN	CICIDS2017	99.70%	95.00%	97.00%	96.00%
Tekerek, 2021	CNN	CSIC2010v2 HTTP	–	97.49%	–	97.51%
Roopak et al., 2019	CNN+LSTM	CICIDS2017	97.16%	–	–	–
Moizuddin and Jose, 2022	GMGWO	NSL-KDD	99.07%	–	–	–
Ding et al., 2022	TACGAN	CICIDS2017	95.86%	–	95.42%	95.83%
Chen Y et al., 2022	MECNN	AWID and CICIDS2017	99.84%	–	–	–
Kunang et al., 2019	AE	NSL-KDD	86.96%	88.65%	–	–
Kherlenchimeg and Nakaya, 2018	AE+RNN	NSL-KDD	80.00%	–	–	–
Shaikh and Shashikala, 2019	AE+LSTM	NSL-KDD	94.30%	–	–	–
Balamurugan et al., 2022	IDSGT-DNN	CICIDS2017	–	98.65%	–	99.00%
Mushtaq et al., 2022	AE-LSTM	NSL-KDD	89.00%	88.00%	94.00%	91.00%
Chohra et al., 2022	Chameleon	IoT-Zeek	–	–	–	97.30%
Seth and Chandra, 2018	MGWO	Solaris	99.87%	–	–	–
		Windows	98.94%	–	–	–

"–" indicates that this evaluation index is not used in the performance evaluation of this model

The second reason is the complexity of the algorithm design. The diversity of network attacks leads to the complexity of cyberspace. The types of attacks are different, and the algorithm design is also different. Therefore, it is necessary to use a variety of algorithms to design the algorithm model. As mentioned above, the researchers chose evolutionary algorithms and random methods to improve the prediction performance. For improving the performance on the evaluation index, different algorithms will show different advantages. It is difficult to improve the evaluation index in four dimensions at the same time. Therefore, the researcher needs to design different algorithms to improve the performance on the different indexes. Finally, the attacker's ability is a key element when designing some ML and DL methods. In particular, the sensitivity to data poisoning attack will be a valuable indicator for assessing the new algorithm. This is a new research field for study of the weakness and sensitivity of the ML model. The designers should consider how the intruder uses DL to bypass the DL based detection system. For example, Bahnsen et al. (2018) studied how attackers used DNN to improve the efficiency of phishing attacks,

ingeniously bypassing the ML based phishing detection system.

4.1.2 Performance analysis of AI algorithms in assessment of the security situation

Above all, AI technology research has had a remarkable effect on cyberspace security issues. ML algorithms such as Bayesian algorithms and hybrid algorithms (such as ML/DL algorithms combined with SI optimization algorithms) are used to settle three types of issues, including cyberspace security risk evaluation, network vulnerability evaluation, and network reliability evaluation (Table 2).

Researchers have explored the application of hybrid optimization algorithms that rely on ML/DL algorithms and SI optimization algorithms, such as cuckoo optimization algorithms or genetic optimization, to assess cyberspace security situations. The special advantage of applying DL algorithms is that they can process huge data. In addition, the DL algorithm can further improve the accuracy and real-time performance of security defense evaluation. However, with the analysis of the literature, we can see that there are still some problems in the algorithm-driven network security evaluation model. Researchers cannot obtain many quantitative results formed by the evaluation model because there is some difference between the calculation value and the expected value. Therefore, there is much more space for researchers to design algorithms to increase the reliability of the evaluation model. According to the research above, the DL algorithm will obtain higher accuracy and robustness than the ML algorithm, but the evaluation time will be longer.

4.1.3 Performance analysis of AI algorithms in defense strategy optimization

The application of a multi-objective optimization algorithm and game theory is the main research direction, and they are used to optimize and improve the efficiency of security defense decision-making and security risk strategy allocation (Table 3).

Research on AI algorithms for cyberspace security defense decision-making models is insufficient.

Table 2 Performance comparison of multiple types of assessment of security situation models

Reference	Algorithm(s)	Evaluation index(es)	Dataset
Li DT et al., 2021	FOA-SVM	Accuracy; *F* score; AUC	–
Kumar VS and Narasimhan, 2021	NB and CRQ-J48	Accuracy	1209 recordings, including webservers, servers, smart mobile phones, laptops, and few smart devices
Li GF et al., 2020	SSC and sequential MCS	Accuracy; reliability	The RBTS bus 2 system and an actual distribution network in the northeast USA
Yang HY and Zeng, 2021	DAENDD(UOSW)	Accuracy; recall	NSL-KDD
Wang JH et al., 2021	GA-PNN	Training speed	–
Wei MH, 2021	GRU-RNN	Accuracy; robustness; time	–
Luan and Tan, 2021	EWM-IFAHP	Accuracy; recall; vulnerability; asset	KDD CUP99
Diao, 2021	NAWL-ILSTM	Accuracy	KDD CUP99
Zhang R et al., 2022	SAA-SSA-BPNN	Network security situation value; convergence analysis; time complexity analysis	National Internet Emergency Response Center network security information and dynamic weekly report
Zhang R et al., 2021	SSA-BP		
Yang XJ and Jia, 2021	IPSO-LSTM		
Zhang ZQ, 2021	GSA-BP	Accuracy; error	IPS log information

"–" indicates that this evaluation index is not used in the performance evaluation of this model

Table 3 Performance comparison of multiple types of defense strategy optimization models

Reference	Algorithm	Evaluation index(es)	Advantages and disadvantages
Liu XH et al., 2021	DRQN-based defense decision-making	Reward	Improving the learning speed of collecting additional historical data and achieving the optimal defense strategy faster in defense game with incomplete information; solving the practical problem that MDP cannot use POMDP to analyze the network and capture the incomplete information
Aggarwal et al., 2022	Instance-based learning (IBL) model	Attacker's success rate; defender's losses	Capturing the risk of confrontation between people and predicting the attacker's decision-making and cognitive biases
Hu BW et al., 2021	DCDM	Strategic cost; system loss; state stability	Being decentralized; ensuring consistency and non-repudiation of results and improving the accuracy of distributed agent decision-making
Hamrioui and Bokhari, 2021	Iterative method	Trade-off cost; security assets	High-quality trade-off and safe assets; satisfying the budget, minimum cost, maximum security level, and cardinality constraints
Hyder and Govindarasu, 2020	Game-theory	Payoff matrix; mixed strategy Nash equilibrium	Obtaining the optimal policy; minimizing attack threat loss
Touhiduzzaman et al., 2019	Game-theoretic graph coloring technique	Steady state probabilities for security mechanism; security index	Better security; security mechanism diversity leading to increased costs

The related algorithms based on ML and DL have not yet been used in the subdivision field for solving cyberspace security defense decision-making issues. On one hand, cyberspace is a high-dimensional complex space, and the model needs to accurately detect a large number of network attacks as input, which will further affect the promotion and development of intelligent decision-making for cyberspace security defense. The accuracy of input data relies on the algorithm design in the intrusion detection model; the more accurate the input data, the more instructive the output results of decision-making. On the other hand, the related algorithms based on ML and DL will bring the serious problem that, if the design of the algorithm model is not perfect, the security strategy of the model output is unreasonable. The relatively small amount of research in this field shows that researchers are cautious about the application of algorithms in cyberspace security.

Today, it is unrealistic to rely solely on manpower to make strategy allocation due to the complexity of network attacks. To address the problem of limited defense allocation ability, the deep application of AI algorithms is an inevitable trend and can avoid risk to the maximum extent and prevent potential network attacks. As a matter of fact, based on ML algorithms (Arshad et al., 2012; Kumar N et al., 2015; Bahnsen et al., 2018) and DL algorithms (Stampa et al., 2017; Cao et al., 2018; Challita et al., 2018; Wei YF et al., 2019; Deng et al., 2020; He et al., 2020), scholars have researched and explored these issues. The algorithms improve the efficiency of decision-making strategy allocation, and realize the dynamic allocation of resources for optimizing the overall efficiency of network strategy allocation. It should be pointed out that cyberspace security defense decision-making is based on network attacks, and then a series of algorithms are designed to generate a system dynamic defense scheme. The generation of the defense decision-making model is consistent with the principle of effective utilization and allocation of network resources, so the design of the algorithm model for solving the network strategy allocation problem can serve as a reference. Supervised learning, RL, DL, and other algorithm models can be applied to explore how to design the cyberspace security defense decision-making algorithm.

4.2 Analysis of dataset application

We further analyzed the datasets used. In the collected literature, we summarized and analyzed the types of different datasets involved in solving network security problems. A total of 22 different datasets that occur most frequently were analyzed. Fig. 4 shows the percentage of papers in which the dataset is used.

It can be seen that the three types of datasets with the highest frequencies were the CICIDS2017 dataset, NSL-KDD dataset, and KDD CUP99 dataset. Among them, the CICIDS2017 dataset showed the highest percentage was 19.7%. The evaluation of a benchmark dataset for intrusion detection may assist a fairer evaluation of the different AI algorithms that have been proposed. In the cyberspace security field there are a large variety of data, but the performances of the AI algorithms are limited by the publicly available datasets. Benchmark datasets with large and regularly updated data are essential to enhance the study of security issues in cyberspace. With the development and application of DL based models in the cyberspace security industry, new datasets are emerging. Currently, the further application of DL algorithms relies on the quality and quantity of available data (Pouyanfar et al., 2019), which can directly influence the reliability of the results. Furthermore, other algorithms performed worse, and also depended on different categories of attacks and the features of the datasets. Although most work paid more attention to designing algorithms to improve the results, many studies have been dedicated to assessing the reliability of benchmark datasets. Some researchers have discussed the dataset including various categories of relevance to network attacks, and highlighted 11 factors, including diversity of traffic data, protocol diversity, amount of data collected, diversity of the attacks considered, inclusion of novel attack types, inclusion of full payloads without anonymity, presence or absence of informative features, updatability, consideration of realistic traffic, extent of labeling, and size of the feature set, which are the main elements for evaluating the reliability of benchmark datasets (Gharib et al., 2016). In addition, the adaptability of the dataset to changes over time should be considered as an indicator of the reliability of the dataset. Other researchers believed that reliable datasets should also provide anonymous means for payload information to ensure user privacy.

5 Conclusions and future directions

Modern information security architectures depend on the wide application of AI technologies. Cyberspace security is facing more severe risks and challenges, and AI technologies are also the weapon for defenders to maintain cyberspace security. Improving the active defense ability based on AI technologies and addressing the new risks and challenges of cyberspace security have become urgent needs. In this review, the latest technical exploration indicates that AI algorithms are playing an increasingly important role in guranteeing cyberspace security, not only developing security ability with good performance in intrusion detection and defense evaluation, but also providing new algorithm design ideas for researchers to explore cyberspace security defense decision-making methods. Network attack detection is an important area in the application of AI technologies. There are two technical hot spots for current research. One is to continue to improve the detection performance of the algorithm on the latest datasets and to analyze and compare the performance of ML and DL algorithms on different datasets. Second, from the research papers from 2021 to 2022, we can see that researchers regard the research of dataset imbalance as the focus, mainly on self-coding algorithms and generating antagonistic network algorithms, to establish a model framework based on a variety of hybrid algorithms for data feature extraction. In the past six years, Chinese researchers have conducted much research on the credibility assessment and prediction of the security situation, publishing 90% of the papers. The research on the optimization of the network security defense strategy was conducted mainly by researchers in USA, with the SI optimization algorithm, and the game theory and DRL algorithm for solving the optimization problem of the network security defense strategy.

Future work should consider human guidance and some new questions raised by AI algorithms. Regarding human guidance, some researchers are even concerned about the key question, which is not to improve endpoint detection but to make the analyst take

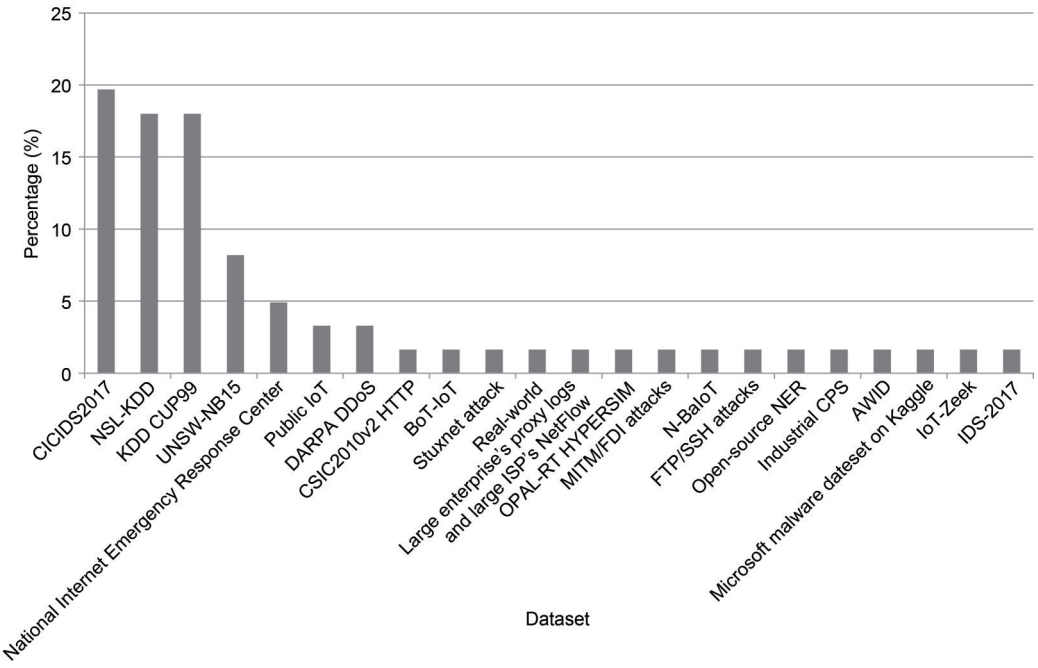

Fig. 4 The percentage of papers in which the dataset is used

action when the alerts are produced. In the future, one constraint is that AI technologies cannot be completely separated from human beings, because human guidance is very important. While formulating the development route of AI, we should focus on risk defense, strengthen the prediction of and research on potential risks, keep a watchful eye on the development of system security defense technologies, and clarify the defense development strategy. The other constraint is that, the combination of future networks and AI algorithms may be a double-edged sword. We can imagine that in future network vision, the application of AI algorithms in a variety of intelligent scenarios will bring rich heterogeneous connections and mass information storage and operation. However, it cannot be ignored that when AI algorithms meet the future network, they are also accompanied by many challenges about privacy protection. On one hand, a secure AI algorithm learning structure can defend the privacy of data. On the other hand, the AI algorithm is likely to be attacked by other networks or abused by other algorithms, which will lead to the invasion of user privacy. Therefore, when using AI algorithms to solve the cyberspace security problem, we must take into account the attack or abuse of those algorithms.

Contributors

Jie CHEN determined the whole research framework and drafted the paper. Dandan WU searched the literature and made algorithm data analysis. Ruiyun XIE reviewed the entire research framework and technology. Jie CHEN and Dandan WU revised and finalized the paper.

Compliance with ethics guidelines

Jie CHEN, Dandan WU, and Ruiyun XIE declare that they have no conflict of interest.

References

Aggarwal P, Thakoor O, Jabbari S, et al., 2022. Designing effective masking strategies for cyberdefense through human experimentation and cognitive models. *Comput Secur*, 117:102671. https://doi.org/10.1016/j.cose.2022.102671

Al-Garadi MA, Mohamed A, Al-Ali AK, et al., 2020. A survey of machine and deep learning methods for Internet of Things (IoT) security. *IEEE Commun Surv Tut*, 22 (3):1646-1685. https://doi.org/10.1109/COMST.2020.2988293

Al-Omari M, Rawashdeh M, Qutaishat F, et al., 2021. An intelligent tree-based intrusion detection model for cyber security. *J Netw Syst Manag*, 29(2):20. https://doi.org/10.1007/s10922-021-09591-y

Al-Yaseen WL, Othman ZA, Nazri MZA, 2017. Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. *Expert Syst Appl*, 67:296-303.

https://doi.org/10.1016/j.eswa.2016.09.041

Andresini G, Appice A, di Mauro N, et al., 2020. Multi-channel deep feature learning for intrusion detection. *IEEE Access*, 8:53346-53359. https://doi.org/10.1109/ACCESS.2020.2980937

Apruzzese G, Colajanni M, Ferretti L, et al., 2018. On the effectiveness of machine and deep learning for cyber security. Proc 10th Int Conf on Cyber Conflict, p.371-390. https://doi.org/10.23919/CYCON.2018.8405026

Arshad SA, Murtaza MA, Tahir M, 2012. Fair buffer allocation scheme for integrated wireless sensor and vehicular networks using Markov decision processes. IEEE Vehicular Technology Conf, p.1-5. https://doi.org/10.1109/VTCFall.2012.6399151

Atefi K, Hashim H, Kassim M, 2019. Anomaly analysis for the classification purpose of intrusion detection system with K-nearest neighbors and deep neural network. IEEE 7th Conf on Systems, Process and Control, p.269-274. https://doi.org/10.1109/ICSPC47137.2019.9068081

Aung YY, Min MM, 2018. Hybrid intrusion detection system using K-means and K-nearest neighbors algorithms. Proc IEEE/ACIS 17th Int Conf on Computer and Information Science, p.34-38. https://doi.org/10.1109/ICIS.2018.8466537

Bahnsen AC, Torroledo I, Camacho LD, et al., 2018. Simulating malicious AI. Proc Symp on Electronic Crime Research, p.15-17.

Balamurugan E, Mehbodniya A, Kariri E, et al., 2022. Network optimization using defender system in cloud computing security based intrusion detection system with game theory deep neural network (IDSGT-DNN). *Patt Recogn Lett*, 156:142-151. https://doi.org/10.1016/j.patrec.2022.02.013

Bdrany A, Sadkhan SB, 2020. Decision making approaches in cognitive radio—status, challenges and future trends. Int Conf on Advanced Science and Engineering, p.195-198. https://doi.org/10.1109/ICOASE51841.2020.9436597

Berman DS, Buczak NL, Chavis JS, et al., 2019. A survey of deep learning methods for cyber security. *Information*, 10(4):122. https://doi.org/10.3390/INFO10040122

Bhuiyan TH, Medal HR, Nandi AK, et al., 2021. Risk-averse bi-level stochastic network interdiction model for cybersecurity risk management. *Int J Crit Infrastr Prot*, 32: 100408. https://doi.org/10.1016/j.ijcip.2021.100408

Bitaab M, Hashemi S, 2017. Hybrid intrusion detection: combining decision tree and Gaussian mixture model. Proc 14th Int ISC (Iranian Society of Cryptology) Conf on Information Security and Cryptology, p.8-12. https://doi.org/10.1109/ISCISC.2017.8488375

Bouhamed O, Bouachir O, Aloqaily M, et al., 2021. Lightweight IDS for UAV networks: a periodic deep reinforcement learning-based approach. IFIP/IEEE Int Symp on Integrated Network Management, p.1032-1037.

Bresniker K, Gavrilovska A, Holt J, et al., 2019. Grand challenge: applying artificial intelligence and machine learning to cybersecurity. *Computer*, 52(12):45-52. https://doi.org/10.1109/MC.2019.2942584

Buczak AL, Guven E, 2016. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Commun Surv Tut*, 18(2):1153-1176. https://doi.org/10.1109/COMST.2015.2494502

Burke D, 1999. Toward a Game Theory Model of Information Warfare. Technical Report, AFIT/GSS/LAL/99D-1. Airforce Institute of Technology, USA.

Buşoniu L, Babuška R, de Schutter B, 2010. Multi-agent reinforcement learning: an overview. In: Srinivasan D, Jain LC (Eds.), Innovations in Multi-agent Systems and Applications. Springer, Heidelberg, p.183-221. https://doi.org/10.1007/978-3-642-14435-6_7

Cao G, Lu ZM, Wen XM, et al., 2018. AIF: an artificial intelligence framework for smart wireless network management. *IEEE Commun Lett*, 22(2):400-403. https://doi.org/10.1109/LCOMM.2017.2776917

Challita U, Dong L, Saad W, 2018. Proactive resource management for LTE in unlicensed spectrum: a deep learning perspective. *IEEE Trans Wirel Commun*, 17(7):4674-4689. https://doi.org/10.1109/TWC.2018.2829773

Chen F, Ye ZW, Wang CZ, et al., 2018. A feature selection approach for network intrusion detection based on tree-seed algorithm and K-nearest neighbor. IEEE 4th Int Symp on Wireless Systems within the Int Conf on Intelligent Data Acquisition and Advanced Computing Systems, p.68-72. https://doi.org/10.1109/IDAACS-SWS.2018.8525522

Chen SS, Lian YF, Jia W, 2008. A network vulnerability evaluation method based on Bayesian networks. *J Univ Chin Acad Sci*, 25(5):639-648 (in Chinese). https://doi.org/10.7523/j.issn.2095-6134.2008.5.011

Chen Y, Lin QZ, Wei WH, et al., 2022. Intrusion detection using multi-objective evolutionary convolutional neural network for Internet of Things in fog computing. *Knowl-Based Syst*, 244:108505. https://doi.org/10.1016/j.knosys.2022.108505

Chohra A, Shirani P, Karbab EB, et al., 2022. Chameleon: optimized feature selection using particle swarm optimization and ensemble methods for network anomaly detection. *Comput Secur*, 117:102684. https://doi.org/10.1016/j.cose.2022.102684

Choi YH, Liu P, Shang ZT, et al., 2020. Using deep learning to solve computer security challenges: a survey. *Cybersecurity*, 3(1):15. https://doi.org/10.1186/s42400-020-00055-5

Deng SG, Xiang ZZ, Zhao P, et al., 2020. Dynamical resource allocation in edge for trustable Internet-of-Things systems: a reinforcement learning method. *IEEE Trans Ind Inform*, 16(9):6103-6113. https://doi.org/10.1109/TII.2020.2974875

Diao WP, 2021. Network security situation forecast model based on neural network algorithm development and verification. IEEE 4th Int Conf on Automation, Electronics and Electrical Engineering, p.462-465. https://doi.org/10.1109/AUTEEE52864.2021.9668668

Ding HW, Chen LY, Dong L, et al., 2022. Imbalanced data classification: a KNN and generative adversarial networks-based hybrid approach for intrusion detection. *Fut Gener Comput Syst*, 131:240-254. https://doi.org/10.1016/j.future.2022.01.026

Elbes M, Alzubi S, Kanan T, et al., 2019. A survey on particle swarm optimization with emphasis on engineering and network applications. *Evol Intell*, 12(2):113-129. https://doi.org/10.1007/S12065-019-00210-Z

Faker O, Dogdu E, 2019. Intrusion detection using big data and deep learning techniques. Proc ACM Southeast Conf, p.86-93. https://doi.org/10.1145/3299815.3314439

Garcia AB, Babiceanu RF, Seker R, 2021. Artificial intelligence and machine learning approaches for aviation cybersecurity: an overview. Integrated Communications Navigation and Surveillance Conf, p.1-8. https://doi.org/10.1109/ICNS52807.2021.9441594

Gharib A, Sharafaldin I, Lashkari AH, et al., 2016. An evaluation framework for intrusion detection dataset. Proc Int Conf on Information Science and Security, p.1-6. https://doi.org/10.1109/ICISSEC.2016.7885840

Goodfellow IJ, Pouget-Abadie J, Mirza M, et al., 2014. Generative adversarial nets. Proc 27th Int Conf on Neural Information Processing Systems, p.2672-2680.

Goodfellow IJ, Bengio Y, Courville A, 2016. Deep Learning. MIT Press, Cambridge, USA.

Graves A, Mohamed AR, Hinton G, 2013. Speech recognition with deep recurrent neural networks. Proc IEEE Int Conf on Acoustics, Speech and Signal Processing, p.6645-6649. https://doi.org/10.1109/ICASSP.2013.6638947

Gronauer S, Diepold K, 2022. Multi-agent deep reinforcement learning: a survey. *Artif Intell Rev*, 55:895-943. https://doi.org/10.1007/s10462-021-09996-w

Gu YH, Li KY, Guo ZY, et al., 2019. Semi-supervised K-means DDoS detection method using hybrid feature selection algorithm. *IEEE Access*, 7:64351-64365. https://doi.org/10.1109/ACCESS.2019.2917532

Gupta ARB, Agrawal J, 2020. A comprehensive survey on various machine learning methods used for intrusion detection system. IEEE 9th Int Conf on Communication Systems and Network Technologies, p.282-289. https://doi.org/10.1109/CSNT48778.2020.9115764

Gupta N, Jindal V, Bedi P, 2022. CSE-IDS: using cost-sensitive deep learning and ensemble algorithms to handle class imbalance in network-based intrusion detection systems. *Comput Secur*, 112:102499. https://doi.org/10.1016/j.cose.2021.102499

Hamrioui S, Bokhari S, 2021. A new cybersecurity strategy for IoE by exploiting an optimization approach. 12th Int Conf on Information and Communication Systems, p.23-28. https://doi.org/10.1109/ICICS52457.2021.9464595

He XM, Wang K, Huang HW, et al., 2020. Green resource allocation based on deep reinforcement learning in content-centric IoT. *IEEE Trans Emerg Top Comput*, 8(3):781-796. https://doi.org/10.1109/TETC.2018.2805718

Hessel M, Modayil J, van Hasselt H, et al., 2018. Rainbow: combining improvements in deep reinforcement learning. Proc AAAI Conf on Artificial Intelligence, p.3215-3222. https://doi.org/10.1609/aaai.v32i1.11796

Hindy H, Atkinson R, Tachtatzis C, et al., 2020. Utilising deep learning techniques for effective zero-day attack detection. *Electronics*, 9(10):1684. https://doi.org/10.3390/electronics9101684

Ho S, Al Jufout S, Dajani K, et al., 2021. A novel intrusion detection model for detecting known and innovative cyberattacks using convolutional neural network. *IEEE Open J Comput Soc*, 2:14-25. https://doi.org/10.1109/OJCS.2021.3050917

Hossain D, Ochiai H, Doudou F, et al., 2020. SSH and FTP brute-force attacks detection in computer networks: LSTM and machine learning approaches. 5th Int Conf on Computer and Communication Systems, p.491-497. https://doi.org/10.1109/ICCCS49078.2020.9118459

Hu BW, Zhou CJ, Tian YC, et al., 2021. Decentralized consensus decision-making for cybersecurity protection in multi-microgrid systems. *IEEE Trans Syst Man Cybern Syst*, 51(4):2187-2198. https://doi.org/10.1109/TSMC.2020.3019272

Hu CH, Liu GK, Li M, 2021. A network security situation prediction method based on SA-SSA. 14th Int Symp on Computational Intelligence and Design, p.105-110. https://doi.org/10.1109/ISCID52796.2021.00033

Hühn J, Hüllermeier E, 2009. FURIA: an algorithm for unordered fuzzy rule induction. *Data Min Knowl Discov*, 19(3):293-319. https://doi.org/10.1007/s10618-009-0131-8

Huo D, Li XY, Li LH, et al., 2022. The application of 1D-CNN in microsoft malware detection. 7th Int Conf on Big Data Analytics, p.181-187. https://doi.org/10.1109/ICBDA55095.2022.9760349

Hyder B, Govindarasu M, 2020. Optimization of cybersecurity investment strategies in the smart grid using game-theory. IEEE Power & Energy Society Innovative Smart Grid Technologies Conf, p.1-5. https://doi.org/10.1109/ISGT45199.2020.9087634

Issa ASA, Albayrak Z, 2021. CLSTMNet: a deep learning model for intrusion detection. 3rd Int Scientific Conf of Engineering Sciences and Advances Technologies, Article 012244. https://doi.org/10.1088/1742-6596/1973/1/012244

Jain M, Kaur G, 2019. A novel distributed semi-supervised approach for detection of network based attacks. 9th Int

Conf on Cloud Computing, Data Science & Engineering, p.120-125. https://doi.org/10.1109/CONFLUENCE.2019.8776616

Kan X, Fan YX, Fang ZJ, et al., 2021. A novel IoT network intrusion detection approach based on adaptive particle swarm optimization convolutional neural network. *Inform Sci*, 568:147-162. https://doi.org/10.1016/J.INS.2021.03.060

Khaw YM, Jahromi AA, Arani MFM, et al., 2021. A deep learning-based cyberattack detection system for transmission protective relays. *IEEE Trans Smart Grid*, 12(3):2554-2565. https://doi.org/10.1109/TSG.2020.3040361

Kherlenchimeg Z, Nakaya N, 2018. Network intrusion classifier using autoencoder with recurrent neural network. Proc 4th Int Conf on Electronics and Software Science, p.94-100.

Khoa TV, Saputra YM, Hoang DT, et al., 2020. Collaborative learning model for cyberattack detection systems in IoT Industry 4.0. IEEE Wireless Communications and Networking Conf, p.1-6. https://doi.org/10.1109/WCNC45663.2020.9120761

Kim J, Shin Y, Choi E, 2019. An intrusion detection model based on a convolutional neural network. *J Multim Inform Syst*, 6(4):165-172. https://doi.org/10.33851/jmis.2019.6.4.165

Krizhevsky A, Sutskever I, Hinton GE, 2012. ImageNet classification with deep convolutional neural networks. Proc 25th Int Conf on Neural Information Processing Systems, p.1097-1105. http://doi.org/10.1145/3065386

Kumar N, Zeadally S, Chilamkurti N, et al., 2015. Performance analysis of Bayesian coalition game-based energy-aware virtual machine migration in vehicular mobile cloud. *IEEE Netw*, 29(2):62-69. https://doi.org/10.1109/MNET.2015.7064905

Kumar VS, Narasimhan VL, 2021. Using deep learning for assessing cybersecurity economic risks in virtual power plants. 7th Int Conf on Electrical Energy Systems, p.530-537. https://doi.org/10.1109/ICEES51510.2021.9383723

Kunal, Dua M, 2019. Machine learning approach to IDS: a comprehensive review. 3rd Int Conf on Electronics, Communication and Aerospace Technology, p.117-121. https://doi.org/10.1109/ICECA.2019.8822120

Kunang YN, Nurmaini S, Stiawan D, et al., 2019. Automatic features extraction using autoencoder in intrusion detection system. Proc Int Conf on Electrical Engineering and Computer Science, p.219-224. https://doi.org/10.1109/ICECOS.2018.8605181

Ledig C, Theis L, Huszár F, et al., 2017. Photo-realistic single image super-resolution using a generative adversarial network. IEEE Conf on Computer Vision and Pattern Recognition, p.105-114. https://doi.org/10.1109/CVPR.2017.19

Li BB, Wu YH, Song JR, et al., 2021. DeepFed: federated deep learning for intrusion detection in industrial cyber-physical systems. *IEEE Trans Ind Inform*, 17(8):5615-5624. https://doi.org/10.1109/TII.2020.3023430

Li DT, Feng HY, Gao YH, 2021. A network security evaluation method based on machine learning algorithm. *Electr Des Eng*, 29(12):138-142, 147 (in Chinese). https://doi.org/10.14022/j.issn1674-6236.2021.12.030

Li GF, Huang YX, Bie ZH, et al., 2020. Machine-learning-based reliability evaluation framework for power distribution networks. *IET Gener Trans Distrib*, 14(12):2282-2291. https://doi.org/10.1049/iet-gtd.2019.1520

Liu P, Zang WY, 2003. Incentive-based modeling and inference of attacker intent, objectives, and strategies. Proc 10th ACM Conf on Computer and Communications Security, p.179-189. https://doi.org/10.1145/948109.948135

Liu XH, Zhang HW, Dong SQ, et al., 2021. Network defense decision-making based on a stochastic game system and a deep recurrent Q-network. *Comput Secur*, 111:102480. https://doi.org/10.1016/j.cose.2021.102480

Liu XX, Zhang JX, Zhu PD, et al., 2021. Quantitative cyber-physical security analysis methodology for industrial control systems based on incomplete information Bayesian game. *Comput Secur*, 102:102138. https://doi.org/10.1016/j.cose.2020.102138

Long J, Shelhamer E, Darrell T, 2015. Fully convolutional networks for semantic segmentation. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.3431-3440. https://doi.org/10.1109/CVPR.2015.7298965

Luan D, Tan XB, 2021. EWM-IFAHP: an improved network security situation assessment model. 2nd Int Conf on Machine Learning and Computer Application, p.1-6.

Lye KW, Wing J, 2002. Game Strategies in Cyberspace Security. Technical Report, No. CMU-CS-02-136, School of Computer Science, Carnegie Mellon University, Pittsburgh, USA.

Ma PC, Jiang B, Lu ZG, et al., 2021. Cybersecurity named entity recognition using bidirectional long short-term memory with conditional random fields. *Tsinghua Sci Technol*, 26(3):259-265. https://doi.org/10.26599/TST.2019.9010033

Mehta V, Bartzis C, Zhu HF, et al., 2006. Ranking attack graphs. Proc 9th Int Workshop on Recent Advances in Intrusion Detection, p.127-144. https://doi.org/10.1007/11856214_7

Mishra P, Varadharajan V, Tupakula U, et al., 2019. A detailed investigation and analysis of using machine learning techniques for intrusion detection. *IEEE Commun Surv Tut*, 21(1):686-728. https://doi.org/10.1109/COMST.2018.2847722

Mohiuddin MA, Khan SA, Engelbrecht AP, 2016. Fuzzy particle swarm optimization algorithms for the open shortest path first weight setting problem. *Appl Intell*, 45(3):598-621. https://doi.org/10.1007/s10489-016-0776-0

Moizuddin MD, Jose MV, 2022. A bio-inspired hybrid deep learning model for network intrusion detection. *Knowl-Based Syst*, 238:107894.

https://doi.org/10.1016/j.knosys.2021.107894

Mushtaq E, Zameer A, Umer M, et al., 2022. A two-stage intrusion detection system with auto-encoder and LSTMs. *Appl Soft Comput*, 121:108768. https://doi.org/10.1016/j.asoc.2022.108768

Narudin FA, Feizollah A, Anuar NB, et al., 2016. Evaluation of machine learning classifiers for mobile malware detection. *Soft Comput*, 20(1):343-357. https://doi.org/10.1007/s00500-014-1511-6

Nguyen HT, Torrano-Gimenez C, Alvarez G, et al., 2011. Application of the generic feature selection measure in detection of web attacks. In: Herrero Á, Corchado E (Eds.), Computational Intelligence in Security for Information Systems. Springer, Berlin, p.25-32. https://doi.org/10.1007/978-3-642-21323-6_4

Nguyen TTT, Armitage G, 2008. A survey of techniques for Internet traffic classification using machine learning. *IEEE Commun Surv Tut*, 10(4):56-76. https://doi.org/10.1109/SURV.2008.080406

Nishiyama T, Kumagai A, Kamiya K, et al., 2020. SILU: strategy involving large-scale unlabeled logs for improving malware detector. IEEE Symp on Computers and Communications, p.1-7. https://doi.org/10.1109/ISCC50000.2020.9219571

Nisioti A, Mylonas A, Yoo PD, et al., 2018. From intrusion detection to attacker attribution: a comprehensive survey of unsupervised methods. *IEEE Commun Surv Tut*, 20(4):3369-3388. https://doi.org/10.1109/COMST.2018.2854724

Olowononi FO, Rawat DB, Liu CM, 2021. Resilient machine learning for networked cyber physical systems: a survey for machine learning security to securing machine learning for CPS. *IEEE Commun Surv Tut*, 23(1):524-552. https://doi.org/10.1109/COMST.2020.3036778

Park JB, Jeong YW, Shin JR, et al., 2010. Closure to discussion of "An improved particle swarm optimization for nonconvex economic dispatch problems." *IEEE Trans Power Syst*, 25(4):2010-2011. https://doi.org/10.1109/TPWRS.2010.2069890

Pouyanfar S, Sadiq S, Yan YL, et al., 2019. A survey on deep learning: algorithms, techniques, and applications. *ACM Comput Surv*, 51(5):92. https://doi.org/10.1145/3234150

Pu ZY, 2020. Network security situation analysis based on a dynamic Bayesian network and phase space reconstruction. *J Supercomput*, 76(2):1342-1357. https://doi.org/10.1007/s11227-018-2575-3

Qazi EUH, Imran M, Haider N, et al., 2022. An intelligent and efficient network intrusion detection system using deep learning. *Comput Electr Eng*, 99:107764. https://doi.org/10.1016/j.compeleceng.2022.107764

Roopak M, Tian GY, Chambers J, 2019. Deep learning models for cyber security in IoT networks. IEEE 9th Annual Computing and Communication Workshop and Conf, p.452-457. https://doi.org/10.1109/CCWC.2019.8666588

Sagar BS, Niranjan S, Kashyap N, et al., 2019. Providing cyber security using artificial intelligence—a survey. 3rd Int Conf on Computing Methodologies and Communication, p.717-720. https://doi.org/10.1109/ICCMC.2019.8819719

Salih A, Zeebaree ST, Ameen S, et al., 2021. A survey on the role of artificial intelligence, machine learning and deep learning for cybersecurity attack detection. 7th Int Engineering Conf "Research & Innovation amid Global Pandemic", p.61-66. https://doi.org/10.1109/IEC52205.2021.9476132

Sapavath NN, Muhati E, Rawat DB, 2021. Prediction and detection of cyberattacks using AI model in virtualized wireless networks. 8th IEEE Int Conf on Cyber Security and Cloud Computing (CSCloud)/7th IEEE Int Conf on Edge Computing and Scalable Cloud, p.97-102. https://doi.org/10.1109/CSCloud-EdgeCom52276.2021.00027

Seth JK, Chandra S, 2018. MIDS: metaheuristic based intrusion detection system for cloud using k-NN and MGWO. 2nd Int Conf on Advances in Computing and Data Sciences, p.411-420. https://doi.org/10.1007/978-981-13-1810-8_41

Shafiqur R, Salman K, Luai MA, 2020. The effect of acceleration coefficients in particle swarm optimization algorithm with application to wind farm layout design. *FME Trans*, 48(4):922-930. https://doi.org/10.5937/fme2004922r

Shaikh RA, Shashikala SV, 2019. An autoencoder and LSTM based intrusion detection approach against denial of service attacks. Proc 1st Int Conf on Advances in Information Technology, p.406-410. https://doi.org/10.1109/ICAIT47043.2019.8987336

Shende S, Thorat S, 2020. A review on deep learning method for intrusion detection in network security. 2nd Int Conf on Innovative Mechanisms for Industry Applications, p.173-177. https://doi.org/10.1109/ICIMIA48430.2020.9074975

Socher R, Huang EH, Pennington J, et al., 2011a. Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. Proc 24th Int Conf on Neural Information Processing Systems, p.801-809.

Socher R, Lin CCY, Ng AY, et al., 2011b. Parsing natural scenes and natural language with recursive neural networks. Proc 28th Int Conf on Machine Learning, p.129-136.

Stampa G, Arias M, Sanchez-Charles D, et al., 2017. A deep-reinforcement learning approach for software-defined networking routing optimization. https://arxiv.org/abs/1709.07080

Stevens-Navarro E, Lin YX, Wong VWS, 2008. An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks. *IEEE Trans Veh Technol*, 57(2):1243-1254. https://doi.org/10.1109/TVT.2007.907072

Su JY, 2021. Intelligent network security situation prediction method based on deep reinforcement learning. IEEE Int Conf on Industrial Application of Artificial Intelligence, p.343-348. https://doi.org/10.1109/IAAI54625.2021.9699894

Sun YY, Liu JJ, Wang JD, et al., 2020. When machine learning meets privacy in 6G: a survey. *IEEE Commun Surv Tut*, 22(4):2694-2724. https://doi.org/10.1109/COMST.2020.3011561

Sutskever I, Vinyals O, Le QV, 2014. Sequence to sequence learning with neural networks. Proc 27th Int Conf on Neural Information Processing Systems, p.3104-3112.

Tekerek T, 2021. A novel architecture for web-based attack detection using convolutional neural network. *Comput Secur*, 100:102096. https://doi.org/10.1016/j.cose.2020.102096

Torres JM, Comesaña CI, García-Nieto PJ, 2019. Review: machine learning techniques applied to cybersecurity. *Int J Mach Learn Cybern*, 10(10):2823-2836. https://doi.org/10.1007/S13042-018-00906-1

Touhiduzzaman M, Hahn A, Srivastava AK, 2019. A diversity-based substation cyber defense strategy utilizing coloring games. *IEEE Trans Smart Grid*, 10(5):5405-5415. https://doi.org/10.1109/TSG.2018.2881672

Ullah F, Naeem H, Jabbar S, et al., 2019. Cyber security threats detection in Internet of Things using deep learning approach. *IEEE Access*, 7:124379-124389. https://doi.org/10.1109/ACCESS.2019.2937347

Waibel A, Hanazawa T, Hinton G, et al., 1990. Phoneme recognition using time-delay neural networks. In: Waibe A, Lee KF (Eds.), Readings in Speech Recognition. Elsevier, Amsterdam, the Netherlands, p.393-404. https://doi.org/10.1016/B978-0-08-051584-7.50037-1

Wang JH, Shan ZL, Tan HS, et al., 2021. Network security situation assessment based on genetic optimized PNN neural network. *Comput Sci*, 48(6):338-342 (in Chinese).

Wang PY, Govindarasu M, 2020. Multi-agent based attack-resilient system integrity protection for smart grid. *IEEE Trans Smart Grid*, 11(4):3447-3456. https://doi.org/10.1109/TSG.2020.2970755

Wei MH, 2021. A new information security evaluation algorithm based on recurrent neural. *J Mianyang Teach Coll*, 40(2):75-80, 87 (in Chinese). https://doi.org/10.16276/j.cnki.cn51-1670/g.2021.02.015

Wei YF, Yu FR, Song M, et al., 2019. Joint optimization of caching, computing, and radio resources for fog-enabled IoT using natural actor-critic deep reinforcement learning. *IEEE Int Things J*, 6(2):2061-2073. https://doi.org/10.1109/JIOT.2018.2878435

Wickramasinghe CS, Marino DL, Amarasinghe K, et al., 2018. Generalization of deep learning for cyber-physical system security: a survey. Proc 44th Annual Conf of the IEEE Industrial Electronics Society, p.745-751. https://doi.org/10.1109/IECON.2018.8591773

Wu SX, Banzhaf W, 2010. The use of computational intelligence in intrusion detection systems: a review. *Appl Soft Comput*, 10(1):1-35. https://doi.org/10.1016/j.asoc.2009.06.019

Xiao JP, Long C, Zhao J, et al., 2021. Survey of network intrusion detection based on deep learning. *Front Data Comput*, 3(3):59-74 (in Chinese). https://doi.org/10.12379/j.issn.2096-1057.2022.12.03

Xin Y, Kong LS, Liu Z, et al., 2018. Machine learning and deep learning methods for cybersecurity. *IEEE Access*, 6:35365-35381. https://doi.org/10.1109/ACCESS.2018.2836950

Yang HY, Zeng RY, 2021. Method for assessment of network security situation with deep learning. *J Xidian Univ*, 48(1):183-190 (in Chinese). https://doi.org/10.19665/j.issn1001-2400.2021.01.021

Yang HY, Zeng RY, Xu GQ, et al., 2021. A network security situation assessment method based on adversarial deep learning. *Appl Soft Comput*, 102:107096. https://doi.org/10.1016/j.asoc.2021.107096

Yang HY, Zhang ZX, Zhang L, 2022a. Network security situation assessment based on deep weighted feature learning. *J Cyber Secur*, 7(4):32-43 (in Chinese). https://doi.org/10.19363/J.cnki.cn10-1380/tn.2022.07.03

Yang HY, Zhang ZX, Zhang L, 2022b. Network security situation assessments with parallel feature extraction and an improved BiGRU. *J Tsinghua Univ (Sci Technol)*, 62(5):842-848 (in Chinese). https://doi.org/10.16511/j.cnki.qhdxxb.2022.22.006

Yang XJ, Jia YM, 2021. IPSO-LSTM: a new Internet security situation prediction model. 2nd Int Conf on Machine Learning and Computer Application, p.1-5.

Ye L, Tan ZJ, 2019. A method of network security situation assessment based on deep learning. *Intell Comput Appl*, 9(6):73-75, 82 (in Chinese). https://doi.org/10.3969/j.issn.2095-2163.2019.06.015

Yeom S, Kim K, 2019. Detail analysis on machine learning based malicious network traffic classification. Proc 8th Int Conf on Smart Media & Applications, p.49-53.

Zeadally S, Adi E, Baig Z, et al., 2020. Harnessing artificial intelligence capabilities to improve cybersecurity. *IEEE Access*, 8:23817-23837. https://doi.org/10.1109/ACCESS.2020.2968045

Zhang HY, Lin KY, Chen WW, et al., 2019. Using machine learning techniques to improve intrusion detection accuracy. IEEE 2nd Int Conf on Knowledge Innovation and Invention, p.308-310. https://doi.org/10.1109/ICKII46306.2019.9042621

Zhang M, Xu BY, Bai S, et al., 2017. A deep learning method to detect web attacks using a specially designed CNN. Proc 24th Int Conf on Neural Information Processing, p.828-836. https://doi.org/10.1007/978-3-319-70139-4_84

Zhang R, Wang YB, 2016. Research on machine learning with algorithm and development. *J Commun Univ China (Sci Technol)*, 23(2):10-18, 24 (in Chinese). https://doi.org/10.16196/j.cnki.issn.1673-4793.2016.02.002

Zhang R, Pan ZH, Yin YF, 2021. Research on assessment algorithm for network security situation based on SSA-BP

neural network. 7th Int Symp on System and Software Reliability, p.140-145. https://doi.org/10.1109/ISSSR53171.2021.00024

Zhang R, Pan ZH, Yin YF, et al., 2022. Network security situation assessment model based on SAA-SSA-BPNN. *Comput Eng Appl*, 58(11):117-124 (in Chinese). https://doi.org/10.3778/j.issn.1002-8331.2110-0391

Zhang ZQ, 2021. Research on network security situation prediction based on improved and optimized BP neural network. 2nd Int Conf on Electronics, Communications and Information Technology, p.1014-1018. https://doi.org/10.1109/CECIT53797.2021.00180

Zhou XY, Belkin M, 2014. Semi-supervised learning. *Acad Press Libr Signal Process*, 1:1239-1269. https://doi.org/10.1016/B978-0-12-396502-8.00022-X

Zhou ZH, 2016. Machine Learning. Tsinghua University Press, Beijing, China, p.390-392 (in Chinese).

List of supplementary materials