P. Abhilash and S. Sunanda, International Journal of Research in Engineering, IT and Social Sciences, ISSN 2250-0588, Impact Factor: 6.565, Volume 14 Issue 06, June 2024, Page 1292-1300

Deep Learning Based Attack Detection for Cyber-Physical System Cybersecurity: A Survey

P.Abhilash, Department of CSE, Narayana Engineering College, Gudur.
S.Sunanda,Assistant Professor, Department of CSE, Narayana Engineering College, Gudur.

Abstract: *With the booming of cyber attacks and cyber criminals against cyber-physical systems (CPSs), detecting these attacks remains challenging. It might be the worst of times, but it might be the best of times because of opportunities brought by machine learning (ML), in particular deep learning (DL). In general, DL delivers superior performance to ML because of its layered setting and its effective algorithm for extract useful information from training data. DL models are adopted quickly to cyber attacks against CPS systems. In this survey, a holistic view of recently proposed DL solutions is provided to cyber attack detection in the CPS context. A six-step DL driven methodology is provided to summarize and analyze the surveyed literature for applying DL methods to detect cyber attacks against CPS systems. The methodology includes CPS scenario analysis, cyber attack identification, ML problem formulation, DL model customization, data acquisition for training, and performance evaluation. The reviewed works indicate great potential to detect cyber attacks against CPS through DL modules. Moreover, excellent performance is achieved partly because of several high quality datasets that are readily available for public use. Furthermore, challenges, opportunities, and research trends are pointed out for future research*.

Index Terms*—Cyber-physical system, cybersecurity, deep learning, intrusion detection, pattern classification*

I.INTRODUCTION

CYBER-physical systems (CPSs) suffer from cyber attacks when they are increasingly connected to the cyber space. According to published in 2017, more than 30 surveys were published to cover the cybersecurity issue in the CPSs. Cyber attacks have become increasingly sophisticated and prevalent as automated attacking tools, and professional hacking groups have started to get involved. A successful cyber attack against a CPS may be disastrous, catastrophic, or even fatal .However, it is a challenge to defend against cyber attacks on CPSs. Many CPS systems lack cybersecurity mechanisms like message authentication, resulting in challenges to detect false data injection attacks. A lack of universal encryption, especially on the systems employing dated technologies, makes it challenging to defend against eavesdropping attacks. System states need to be referred to detect replay attacks. In addition, the use of dated technology in operation limits the choices of defenses to network traffic in most cases . Deep learning (DL) delivers superior performance to traditional machine learning (ML) solutions. Whenever there is adequate data, DL models almost deliver excellent results. However, DL models have been slowly applied to solve the CPS cybersecurity issue compared with other fields such as NLP, image processing, software vulnerability and many more. It is also observed that many DL models have been proposed in recent publications to detect CPS cyber attacks. A widely accepted view to explain the difficulty of detecting cyber attacks on CPSs was accredited to the degree of complexity when superposing cybersecurity over CPSs . There exist a few short-length survey papers on CPS cybersecurity . Some papers investigated data-driven methods for detecting cyber attacks against CPS systems. However, there is no detailed discussion on applying DL methods to detect CPS cyber attacks. A short survey was provided in with a four-step framework to apply DL methods on CPS issues, including cybersecurity, adaptability, recoverability, and many more, without a specific focus on cybersecurity. Furthermore, most of the cited works in were published between 2012 and 2016, but this survey includes most papers between 2017 and 2021. A survey of surveys was presented in without relevance to DL models. A comprehensive survey on the cyber attacks against CPSs was presented in without

investigating the DL models. Various methods of detecting cyber attacks in the CPSs were summarized in without using DL methods. A comprehensive list of CPS attacks and challenges were provided in but overlooking ML, or DL approaches. A cybersecurity analysis framework was proposed in without utilizing the rich sources of available data. A recently published survey in presents cybersecurity control and state estimation from active and passive defence perspectives.

II.FUNCTIONAL OVERVIEW

Cyber-Physical Systems (CPS) integrate physical processes with computational and networking capabilities, necessitating robust cybersecurity measures. Deep learning (DL) has become a key approach for detecting cyber-attacks on CPS due to its ability to learn complex patterns from large datasets. CPS face a variety of threats including denial of service (DoS), data integrity attacks, replay attacks, and false data injection. DL techniques such as supervised learning (e.g., CNNs, RNNs, LSTM), unsupervised learning (e.g., autoencoders, GANs), and reinforcement learning (e.g., deep Q-networks) are employed to identify these threats. The process involves data collection from sensors and network traffic, normalization, feature extraction, and model training. Evaluating these models using metrics like accuracy, precision, recall, and AUC-ROC ensures their effectiveness. Deploying DL models enables real-time attack detection, adaptive learning to update models, and automated or manual response systems. Despite challenges in scalability, robustness against adversarial attacks, and the need for interpretable AI, DL offers significant promise in enhancing CPS cybersecurity by providing advanced, continuous monitoring and detection capabilities

III.RESEARCH METHODOLOGY

Our methodology represents a deep understanding of the surveyed papers. The process consists of six steps, including CPS scenario analysis, cyber attack identification, DL problem formulation, DL model construction, data acquisition, and performance evaluation. Fig. 1 shows a process of detecting cyber attacks in the context of a CPS by using DL models. For example, a smart grid may suffer from erroneous controls derived by electric load forecasts. Falsely injected messages containing maliciously crafted information need to be identified and eliminated before committing the prediction process. A stacked AutoEncoder (AE) proposed in may serve as a reliable regressor to predict the energy load on the system. The chosen AutoEncoder was subsequently trained with sufficient simulation data. At last, the DL model delivered excellent prediction results with the mean absolute percentage error of 3.51% on annual predictions.

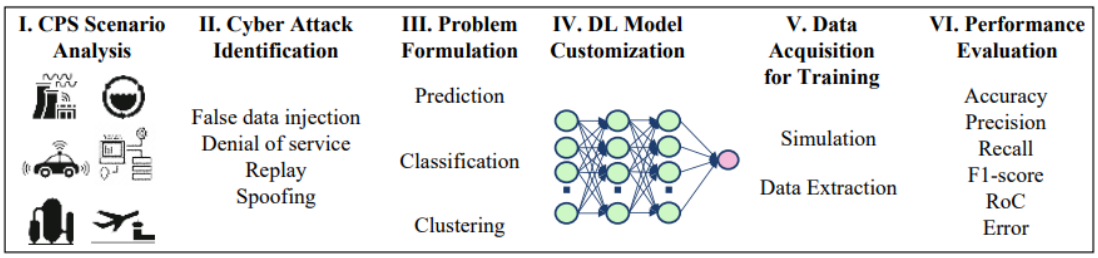

Fig. 1. The DL driven methodology for CPS cybersecurity considers the essential needs for training robust and usable DL models in the context of cyber attacks against the CPS systems

A. Step I: CPS Scenario Analysis

The normal operations of CPSs rely on several important factors, including dependability, real-time operation, fault tolerance, cybersecurity, and many more. We must consider these requirements holistically. Dependability consists of service availability and reliability to minimize the system downtime; real-time operation is a critical factor for maintaining the system operation when the inputs and environment rapidly change; fault tolerance requires that the critical components of the system have sufficient backups to prevent the system from shutting down; and cybersecurity requirements are becoming more and more prominent when many CPSs are connected to the cyber space to improve the quality of system control and the overall level of quality of

service. According to Mitchell et al., there are four primary categories of characteristics of CPS intrusion detection, including physical process monitoring, closed control loops, attack sophistication, and legacy technology.

B. Step II: Cyber Attack Identification

Upon completion of identifying the CPS scenario, we need to define a set of appropriate cyber attacks associated with CPS characteristics. For example, we will have more confidence to detect the falsely injected network packets if physical processes of the CPS components are properly monitored; cyber attacks like replay attacks may be detected on a CPS with a closed control loop; unknown attacks and sophisticated attacks like web attacks need to be considered if there is any concern of attack sophistication; denial of service (DoS) attacks and replay attacks are more prevalent in the I. CPS Scenario Analysis II. Cyber Attack Identification III. Problem Formulation IV. DL Model Customization V. Data Acquisition for Training VI. Performance Evaluation Accuracy Precision Recall F1-score RoC Error Simulation Data Extraction Clustering Prediction Classification False data injection Denial of service Replay Spoofing Fig. 1. The DL driven methodology for CPS cybersecurity considers the essential needs for training robust and usable DL models in the context of cyber attacks against the CPS systems. presence of legacy technology. Based on the surveyed articles, we identify many common cyber attacks. Some frequent cyber attacks against the industrial control network include false data injection attacks, DoS attacks, replay attacks, and alike; and some frequent cyber attacks against the software-based controllers with a centralized server include brute force attacks, botnets, web attacks, heartbleed attacks, infiltration attacks and many more. Effective and efficient detection of these cyber attacks can be leveraged by using DL models, so we will need to translate the cybersecurity problem to the ML domain.

C. Step III: ML Problem Formulation

After aligning the cyber attacks to the CPS characteristics, the research problem can be translated to the ML/DL domain. ML is defined in as “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.” DL is referred to in as solving a complex problem by using a hierarchy of more straightforward concepts without too much human intervention. The definition of ML is general, and we will implement an ML solution in multiple steps. In this step, we need to define the task T, including classification, clustering, regression, etc. A classification task requires that the trained model allocates its output to a pre-defined set of “classes” which could be the specific cyber attack categories; a clustering task often requires that the trained model allocates its output to a few “clusters” which could indicate normal traffic or attack traffic; a regression task is also known as a prediction task which requires the trained model to predict some numerical values. For example, a classification problem was found in to differentiate cyber attack types; a clustering problem was found in to separate covert messages from the normal messages; and a regression problem was set in to predict the electric load in a smart grid. The choice of the ML tasks will impact the construction of the DL models.

D. Step IV: DL Model Customization

The DL model is constructed by selecting an architecture suitable for the research problem and optimizing parameters. The choice of DL models should be made according to actual needs. For example, autoencoders are good at translating the input data so that they are suitable for learning the representations of the data often required in prediction or regression tasks convolution networks (CNNs) and other models are usually used in classification tasks . The configuration of the chosen DL model also depends on the available data. A DL model with a large number of neurons per layer will almost always require more data than a DL model with the same design but a few neurons per layer. Some trade-offs can also be made by stacking more hidden layers inside the DL model instead of expanding the layer size. The ways and insights of the customizing model can be explored based on a thorough understanding of DL algorithms and CPS cybersecurity data. Furthermore, we can achieve improvement at various levels by combining the choice of DL models with a specific research problem.

E. Step V: Data Acquisition for Training

Data acquisition is a critical step for training DL models. The quality and quantity of data determine the effectiveness of solving the research problem. Also, data can serve as the source for setting up ground truth and affect the prediction model's performance. One of the simplest methods to collect data is through simulation. This method is often used to generate datasets for power grids such as IEEE 9-bus, 14-bus, 30-bus, and 118-bus systems in Matlab. The other method relies on several existing datasets harvested by other researchers. These datasets include the SWaT dataset1 , the SCADA IDS dataset2 , the CICIDS2017 dataset3 , the UNSW NB15 dataset4 , and the KDD99 Cup dataset5.

F. Step VI: Performance Evaluation

The last step is used to determine whether the DL meets our expected objectives through performance evaluation. The performance is usually measured according to various metrics. We divide the performance metrics into two categories according to the tasks:

1) For prediction or regression tasks, a number of error metrics are used to measure the performance, including mean absolute error (MAE), mean relative error (MRE), root of mean squared error (RMSE), and mean absolute percentage error (MAPE).

2) For classification or clustering tasks, there are a few standard metrics, including accuracy, recall, precision, false positive rate (FPR), F1 score. And occasionally, graphical plots like receive operating characteristic (ROC) curves are used by plotting TPR as y-axis and FPR as x-axis to depict the trade-offs between benefits and costs. Finally, area under ROC curve (auROC) is used to indicate the cumulative strength of a particular ROC curve.

IV.CPS CYBERSECURITY WITH DEEP NEURAL MODELS

This Section surveys the relevant literature of detecting cyber attacks in the context of CPSs by following the research methodology described in Fig. 1. In particular, the body of the literature is divided into two parts according to the DL architectures, which will be elaborated below.

A.Representation Learning for Attack Detection

An AutoEncoder-based (AE) model was proposed in to preserve privacy information in the context of smart power networks. Data privacy violations are becoming more and more popular in smart power networks. It is challenging to defend against inference attacks, because the smart power networks represent the CPS characteristics of physical process monitoring, closed control loops, attack sophistication, and legacy technology. The research problem of defending against inference attacks was translated into a classification problem in the ML domain. A Variational AutoEncoder (VAE) was proposed to provide transformed features for the ultimate classification task and transform raw data into an encoded format for preventing inference attacks. A VAE is a feedforward model used for encoding an input into new data codes using a set of weighted parameters. The VAE consisted of one input layer, four hidden layers, and one output layer. The transformed data from the output layer were written to the database for publication. Two datasets were used to evaluate the VAE, i.e., the power system dataset and the UNSW NB15 dataset. The Power system dataset is a multi-class dataset involving 37 scenarios that include 8 natural events, 28 intrusive events, and 1 no event; and the UNSW-NB15 dataset includes a combination of current normal and attack records. 300,000 random samples of legitimate and attack observations were chosen from each dataset for assessing the performance of the proposed framework. Although the VAE was only employed as a part of the intrusion detection system, its strength was demonstrated while transforming complex data into a simple form. The VAE achieved 0.921 for accuracy and 0.005 for loss on the power system dataset, and 0.998 for accuracy and 0.0001 for loss on the UNSW-NB15 dataset.

B. Cyber Recognition with Deep Learning Methods

Cybersecurity Pattern Recognition with Deep Neural Networks (DNNs): A DNN-based model was proposed in to learn the communication patterns between electronics control units (ECUs) in the context of in-vehicular network security. The security of communication messages among ECUs is vital because a group of ECUs can control and monitor a vehicle's status during a maneuver. It is challenging to ensure cybersecurity because most communications between ECUs are through the controller area network protocol, which has no support for authentication or integrity check. Specifically, fake packets injected into the open communication channel through the controller area network protocol pose severe cybersecurity risks. Detecting the fabricated or modified packets in the vehicular setup needs to meet the requirements of physical process monitoring and

legacy technologies. This intrusion detection problem was translated into a binary classification problem in the ML domain. That is, statistical features were extracted from high dimensional CAN packet data through a dimension reduction process to represent the normal and attack packets. A 5- layered DNN model was constructed based on a standard DBN model by adding a binary classification layer as the final output layer. The DBN's coefficient weights were determined through an unsupervised pre-training process, but the final DNN model was trained with a bottom-up supervised manner. During each simulation round, a total of 200 000 packets were generated by the Open Car Test-bed and Network Experiments (OCTANE) generator. A 70:30 split was made to divide training and testing sets. Many experiments were conducted by varying the layers of the DNN model from 5 to 11 to investigate the trade-offs between performance and efficiency. The empirical results demonstrated the effectiveness of the proposed DNN model while comparing it with ANN and SVM. The best performance was achieved as 0.978 for accuracy, 0.016 for false positive rate, and 0.028 for false negative rate. Given the detection ratio of over 99%, the proposed DNN model showed good potentials to detect fake packets on vehicular networks despite that the DNN models' efficiency with more than five layers needed to be improved to meet the real-time requirements.

V. CHALLENGES AND FUTURE OPPORTUNITIES

Six potential areas are depicted in Fig. 2 where challenges and new research directions may arise. These six areas correspond to the six steps of our research methodology, as shown in Fig. 1. Our research methodology helps provide an overview of the research literature and extract important elements for comparative analysis. The analysis results underpin research challenges and opportunities in the near future.

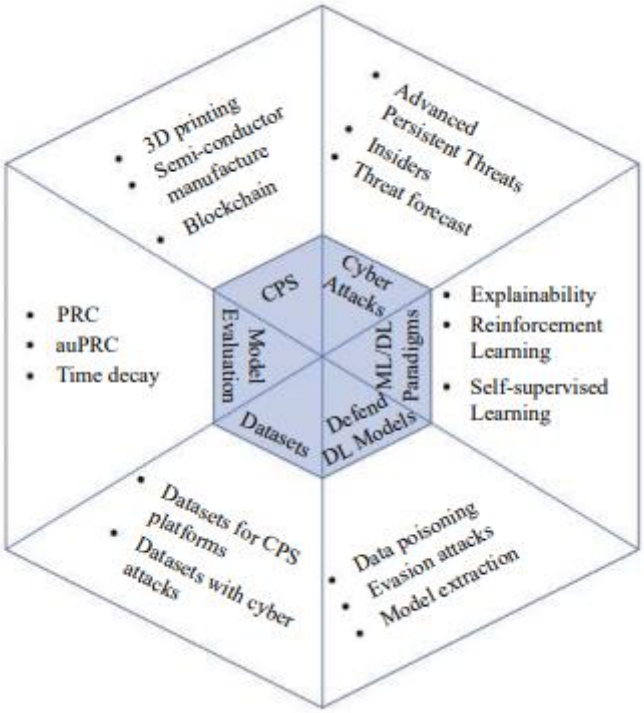

Fig. 2. Research directions for DL driven CPS cybersecurity

A.New CPS Cybersecurity Scenarios

The papers studied the communication networks in CPS scenarios. The majority of the surveyed publications investigated the CPS scenarios in water treatment plants or smart grids, which accounted for thirteen out of twenty surveyed papers. Among the remaining six papers, two studied vehicle networks, one on generic industrial control networks, one on chemical process plant, one on aviation communication networks, and one on gas pipeline controllers. The imbalanced topics suggest that CPS scenarios were underrepresented.

B. Identification of New Cyber Threats

Almost all the surveyed papers studied false data injection attacks. The detection of stealthy false data injection attacks is challenging because of the large amount of noise produced in the CPS and the lack of cybersecurity mechanisms to authenticate devices and messages transmitted across the network. There are a few types of false injection attacks, depending on the attacker's information and goals. For example, no advanced information is required for launching DoS attacks; only recorded packets are required for replay attacks; scanning tools are required for probing attacks; automated tools are required for fuzzy attacks. The effective and efficient analysis of the network traffic is crucial for defending against these attacks.

C. Adopting New ML/DL Paradigms

All the surveyed papers followed traditional ML paradigms, including supervised and unsupervised learning. Apart from four papers that examined regression problems and three papers on clustering problems, the

remaining papers studied classification problems. The dominating use of the supervised learning paradigm reflected the importance of the well-labeled data. In particular, network packets were labeled as normal or attack traffic, and the attack types often were differentiated. Such reliance on labeled data restricted the wide adoption of ML or DL methods. We advocate for researchers and practitioners to try new ML/DL paradigms. These new paradigms include reinforcement learning and self-supervised learning, together with improving the model's explainability. Instead of relying on learning the records, reinforcement learning focuses on experience . It is very suitable in isolated environments where many CPSs are currently deployed. For example, a reinforcement learning model was constructed in to predict the normal driving behaviors.

D. Defending the Trained DL Models

No surveyed works are considered to defend the trained DL models against various attacks. However, we would like to highlight the importance of defending the trained DL models because of the computational expenses to train the DL models, the important roles of the trained models, and potential dangers if the DL models are compromised. Due to the hunger for training samples, the DL models are sometimes trained with data from untrustworthy sources. Thus, adversarial attacks are prevalent because of linear behavior in high dimensional space . For example, Android HIV was proposed in to automatically generate adversarial Android malware that the existing detectors failed to detect.

E. Enriching CPS Cybersecurity Datasets

Among the surveyed papers, datasets collected in the field dominated the simulation with a 14:6 ratio. Simulated data were investigated in the two CPS scenarios — smart grids and vehicular networks. In smart grids, Matlab was the only choice for simulating electric load in five papers; and the OCTANE simulator was used in one paper on vehicular networks. However, there is a significant risk of solely relying on proprietary products like Matlab because the availability of such products may be discontinued unprecedently. On the other hand, field data is independent of the simulation platform and offers researchers good flexibility. Among the papers using field data, five papers chose the SWaT dataset, two papers the CICIDS2017 dataset, and the rest seven papers different datasets. The SWaT dataset dominated the field data category for a few reasons: 1) The network traffic data were continuously collected for 11 days from the control networks and from the sensors of a physical testbed, 2) the traffic with and without attack were chronologically separated for easy use, and 3) there were 36 attack scenarios against different components of the testbed. To our surprise, the NSL-KDD dataset [33], also known as KDD99-cup, was studied in one surveyed paper despite its age of 20+ years. Using such a

dated dataset may cause people to draw biased conclusions because many cyber attacks were not included in the NSL KDD dataset. Therefore, we recommend the researchers to use datasets like the UNSW-NB15 dataset [32] where recent cyber attacks were present.

F. Improving the Model Evaluation

Standard performance metrics were used in most of the surveyed papers. False positives were investigated, along with accuracy and error rates. It is proven in [34] that it is significantly more difficult to detect the rarely occurred attacks than the common ones derived by the Bayesian laws. In real-world CPSs, cyber attacks may rarely occur, so the DL models trained in the lab settings may be invalid. Since most papers did not investigate the impact of the imbalanced data between normal and attack traffic, the empirical results may be substantially biased or inflated. Cross comparisons in [75] showed that the precision-recall curve (PRC) and the area under precision-recall curve (auPRC) were more resilient to imbalanced data than ROC and auROC. Therefore, new studies should consider reporting PRC and auPRC.

Furthermore, time decay should be considered in future studies because each trained ML or DL model's performance will inevitably degrade over time. When the cyber attacks rapidly evolve, the models trained with old data will struggle with detecting new attacks. A time decay metric was proposed in [75] to evaluate a trained

model's performance loss. By studying the time decay, we will be able to decide when the model needs to be retrained. We strongly hope to see future work similar to [75] in the context of CPS and cyber attacks. Once the in-depth knowledge is developed and gained, we may expect to mitigate the risk of CPS cyber attacks.

VI.CONCLUSION

This survey provides a current view of detecting cyber attacks in the CPSs. A six-step DL driven methodology is proposed to summarize and analyze the twenty recently published papers in this survey. Specifically, a panoramic view is obtained through inspecting the CPS scenarios, identifying cybersecurity problems, translating the research problem to the ML/DL domain, constructing the DL model, preparing datasets, and finally evaluating the model. Cyber attacks persist as an ongoing and prominent threat to the security and safety of the CPSs. The reviewed works show great potential to exploit CPS cyber data through DL models because of their promising performances. The excellent performance is achieved partly because of several high-quality datasets that are readily available for public use. In addition to following the success of current research, we also identified promising research topics, including integration with blockchain, detection of advanced persistent threats, adopting new ML and DL paradigms, prevention of adversarial and model extraction attacks, enriching datasets, and applications of additional performance metrics. We are optimistic and confident that the research in this field will flourish.

.

VII. REFERENCES

[1] J. Giraldo, E. Sarkar, A. A. Cardenas, M. Maniatakos, and M. Kantarcioglu, "Security and privacy in cyberphysical systems: A survey of surveys," IEEE Design & Test, vol.34, no.4, pp.7–17, 2017. [1]

[2] R. Mitchell and I.-R. Chen, "A survey of intrusion detection techniques for cyber-physical systems," ACM Computing Surveys, vol.46, no.4, pp. 1–29, 2014. H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," IEEE Trans. Automatic control, vol.59, no.6, pp.1454–1467, 2014.

[3] X. Ge, Q.-L. Han, M. Zhong, and Z.-M. Zhang, "Distributed Krein space-based attack detection over sensor networks under deception attacks," Automatica, vol. 109, Article 108557, Nov. 2019.

[4] W. He, Z. Mo, Q.-L. Han, and F. Qian, "Secure impulsive synchronization in Lipschitz-type multi-agent systems subject to deception attacks," IEEE/CAA Journal of Automatica Sinica, vol.7, no. 5, pp.1326–1334, 2020.

[5] X. Yang, L. Shu, J. Chen, M. A. Ferrag, J. Wu, E. Nurellari, and K. Huang, "A survey on smart agriculture: Development modes, technologies, and security and privacy challenges," IEEE/CAA Journal of Automatica Sinica, vol.8, no.2, pp.273–302, 2020.

[6] X.-M. Zhang, Q.-L. Han, X. Ge, and L. Ding, "Resilient control design based on a sampled-data model for a class of networked control systems under denial-of-service attacks," IEEE Trans. Cybernetics, vol.50, no. 8, pp.3616–3626, 2020.

[7] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[8] D. Xiong, D. Zhang, X. Zhao, and Y. Zhao, "Deep learning for emg□based human-machine interaction: A review," IEEE/CAA Journal of Automatica Sinica, vol.8, no.3, pp.512–533, 2021.

[9] G. Lin, S. Wen, Q.-L. Han, J. Zhang, and Y. Xiang, "Software vulnerability detection using deep neural networks: A survey," Proc. the IEEE, vol.108, no.10, pp.1825–1848, 2020.

[10] S. Liu, G. Lin, Q.-L. Han, S. Wen, J. Zhang, and Y. Xiang, "Deepbalance: Deep-learning and fuzzy oversampling for vulnerability detection," IEEE Trans. Fuzzy Systems, vol.28, no.7, pp.1329–1343, 2020.