A preliminary SWOT evaluation for the applications of ML to Cyber Risk Analysis in the Construction Industry

Dongchi Yao1, Borja Garcia de Soto1

1 S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), Experimental Research Building, Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates
dy2037@nyu.edu, garcia.de.soto@nyu.edu

Abstract. Construction 4.0 is driving construction towards a data-centered industry. Construction firms manage significant amounts of valuable digital information, making them the target of cyberattacks, which not only compromise stored information but could cause severe harm to cyber-physical systems, personnel, and products. Therefore, it is critical to conduct cyber risk analyses to manage construction information assets to ensure their confidentiality, integrity, and availability. Traditional risk analysis methodologies like Fault Tree Analysis have limitations in dealing with the rapidly evolving cyber risks. As an alternative, Machine Learning (ML) methods are finding their way into the risk analysis field. ML models developed for cybersecurity purposes can learn from past results to make reliable predictions while removing the laboriousness of the traditional risk analysis. This article reviews ML techniques used for cyber risk analysis in different industries in recent years. Based on that, we investigate how ML techniques could be used for cyber risk analysis. Afterward, a SWOT analysis is conducted to identify the Strengths, Weaknesses, Opportunities, and Threats regarding the applications of ML in cyber risk analysis in the construction industry, and recommendations to address the weaknesses and threats are presented. Finally, future research areas using ML to prevent cyberattacks in the construction industry are proposed.

Keywords: Artificial Intelligence; Construction 4.0; Cybersecurity; Machine learning

1. Introduction
With the advent of Construction 4.0, big data, BIM, Machine Learning (ML), and cloud computing, to name a few, are finding their way into the different phases of the lifecycle of construction projects [1]. While digitalization undoubtedly facilitates the efficiency of project management, emerging challenges on data protection have to be addressed [2]. Apart from confidential data loss due to cyberattacks for profitability, harms to cyber-physical systems, personnel, and assets might also be incurred. A breach of blueprints in the planning phase [3], malicious attacks to BIM models during the design phase [4], or attacks on automatic control systems in the construction phase [5], data leakage due to information exchange among various parties [6] are all examples of cyber incidents. According to [7], more than 75% of respondents in the construction industry had experienced a cyber incident since 2019, and it is projected that cybercrime will cost businesses approximately $6 trillion per year on average through 2021[7]. However, the awareness and readiness on cybersecurity in the construction industry are still quite low, and in general, the construction sector treats cybersecurity as a lesser business priority compared to other industry sectors [8]. Therefore, considering cyber risk analysis is necessary to give stakeholders insights about preventing cyberattacks and minimizing potential losses. Traditional cyber risk analysis methods, such as Evidence Theory, and Analytic Hierarchy Process, rely excessively on...
experts’ subjective judgment, and the analysis processes have to be started over when placed in different risk scenarios, which is time-consuming [9]. Fortunately, with artificial intelligence (AI) advancements, ML techniques can address some of these problems and generate results more swiftly and objectively. The rest of this paper is organized as follows. Section 2 reviews recent publications on the application of ML in cyber risk analysis in other industries, and shows its applicability in the construction industry. Based on the findings from existing publications, Section 3 includes a summary of a SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis to get a preliminary evaluation of the applications of ML for cyber risk assessment in the construction industry; Section 4 lists corresponding recommendations regarding the weaknesses and threats identified; Section 5 concludes this study and suggests areas of future work.

2. Literature review
Cybersecurity risk analysis refers to the process of evaluating the security attributes of a cyber system, such as confidentiality, integrity, and availability as per the relevant information security technology and management standards. A typical framework of cyber risk analysis is shown in Figure 1. It can be seen that there are three main elements: (1) Risk identification: Consisting of asset, threat, and vulnerability identification, (2) Risk estimate: Including risk level determination and risk score calculation, and (3) Risk management: Consisting of several tasks. For purposes of this study, we focus on Threat and Vulnerability Identification, Risk Estimate, and Risk Monitoring (the shared boxes in Figure 1). The selection of the first three is evident. The last one was chosen because it can provide real-time feedback on a cyber system’s risk posturing, which should be prioritized [10].

ML learns from existing data to forecast future behaviors, outcomes, and trends rather than through programming. In the cyber risk analysis context, the past data are the results of previous threat and vulnerability identifications, cyber risk estimates, and risk management plans. In general, cybersecurity problems can be classified into three categories. (1) For classification category, supervised learning algorithms along with DL (Deep Learning, a branch of ML) algorithms are suitable, (2) for Clustering and (3) for Dimension Reduction categories, unsupervised learning algorithms are mainly adopted.

![Figure 1. Flow chart of cyber risk analysis (adapted from [11]).](image-url)
2.1. Machine learning in risk identification

Threats can exploit vulnerabilities existent in a cyber system to attack the corresponding asset. For example, in the construction industry, threats can be data exfiltration, contagious malware spread, denial of service attack, etc. In addition, vulnerabilities can come from collaborative working processes without strict rules, coding mistakes, software design defects, and unregulated data aggregation in BIM [12]. However, threats are becoming increasingly complex, and vulnerabilities are becoming more covert, making the manual identification of threats implausible. In this case, ML can fit in very well due to its automation ability. Examples of relevant recent (from 2016 to 2021) publications that investigate the use of ML for risk identification (in particular threat and vulnerability identification) in various fields include the following. Pang et al. [13] proposed a prediction technique targeting the early identification of potentially vulnerable software components by combining the Support Vector Machine (SVM) algorithm and ensemble learning strategy. Varma et al. [14] combined Bat Optimization Algorithm for Wrapper-based Feature Selection (BOAWFS) and the Random Forest (RF) algorithm to identify Android malware. Yuan et al. [15] presented a novel insider threat detection method with Deep Neural Network (DNN)-specifically, the Long Short-Term Memory-Convolutional Neural Network (LSTM-CNN) framework-to find employee’s anomalous behavior. Russell et al. [16] employed a Deep Representation Learning algorithm and leveraged the wealth of C and C++ open-source code available to develop a large-scale function-level vulnerability detection system. Ebrahimi et al. [17] proposed a Transductive Support Vector Machines (TSVM) algorithm and a deep bidirectional Long Short-Term Memory (LSTM) algorithm for semi-supervised threat identification and labeling. Medina et al. [18] employed Inception-ResNet-V2 architecture and VGG16 architecture on transfer learning and fine-tuning to detect vulnerabilities in industrial control systems. Of course, these are not an exhaustive list but give insights into how ML techniques could be used in the construction industry in risk identification, such as identifying an attack by detecting abnormal data flow in computers, identifying a phishing email received by contractors by analyzing the text, identifying an intentional data-stealing of building plans by detecting the excessive attempts of log-in, etc.

2.2. Machine learning in risk estimate

Following the identification processes for threats and vulnerabilities, an asset’s risk level or score should be determined (i.e., a risk estimate can be made). For example, in the construction industry, when a kind of threat to the BIM software appears and is detected, relevant personnel should calculate the score of risk this threat incurs or should classify the risk to a specific risk level, based on which a risk management measure is to be decided to deal with this threat. Since the risk level classification and risk scoring process align well with the functions of ML and with the advancement of Neural Network-based DL algorithms, new models based on the ML are being developed for achieving the goal of automatic risk estimate calculations. From a survey of existing literature from various fields focused on ML for risk estimates, we summarize some of the relevant work found. Matsika et al. [19] developed a risk estimate tool based on Multi-layer Artificial Neural Networks to calculate the risk of Metro and Light Rail Systems imposed by a potential terrorist attack. Bilge et al. [20] adopted fully and semi-supervised learning algorithms to create the RiskTeller system that can classify the risk level of a computer according to each computer’s usage patterns. Jiao [21] adopted the Genetic Algorithm based Back Propagation Neural Network (GABP-NN) model to calculate the cyber risk score of a computer where 12 risk factor indexes are established as the input and a risk level from 1 to 5 is set to be the output. Zhang et al. [22] proposed a DL-based model for industrial cyber risk level probabilistic classification (5 levels from “very low” to “very high”) using the real-world industrial dataset. Sun et al. [23] proposed Siamese Network Classification Framework (SNCF) algorithm that can classify the network machine to a specific risk level given the imbalanced original data. Kalinin et al. [24] proposed a 3-layer-perceptron NN algorithm to classify the risk level of a given infrastructure within the context of Smart City Construction, using the information of the network nodes as the input. Using ML methods, repetitive risk estimate processes are no longer needed, saving much time and labor. These frameworks can be applied to the construction field to manage cyber risk, such as predicting the risk level/score incurred by a malicious threat, not patching a vulnerability of a construction software, IT personnel’s maloperation, the incompleteness of certain contract text, etc.
2.3. Machine learning in risk management

Within the scope of risk management, risk monitoring is an important topic. Continuous cybersecurity monitoring provides real-time visibility into the network ecosystem, allowing incident response teams to stay ahead of impending cyber threats and mitigate information security risks. For example, in the construction industry, a complete risk monitoring system can, at any time, analyze the data flow from the computers in the site office, from the sensors in the field, from the equipment operating, etc. Any abnormal pattern of the data flow can be analyzed and judged whether it signifies an attack occurs, and then the monitoring system can automatically choose which action to take or remind personnel to make decisions to mitigate the risk. Recent publications of ML applications on risk monitoring include the following. Chung et al. [25] integrated a Naive Q-Learning algorithm into the cybersecurity monitoring system of a cyber infrastructure for modeling the decision-making process. He et al. [26] proposed a DL-based scheme by employing Conditional Deep Belief Network (CDBN) to monitor the high-dimensional temporal behavior features of the unobservable False Data Injection (FDI) attacks. Subrato et al. [27] compared various ML algorithms (Naive Bayes, K-Nearest Neighbors, SVM, Decision Tree, ANN) to monitor cyber risks of social media. Li et al. [28] combined distributed streaming processing mechanism and ensemble prediction algorithm to monitor the abnormal network traffic of a cyber-physical power system. Sakthivel et al. [29] employed a Recursive Neural Network (RNN) algorithm to monitor and prevent the cyber system from cyberattacks in a manufacturing company. Garrido et al. [30] adopted a graph learning algorithm to detect intrusion, score anomalous activities and monitor security in industrial automation systems. These applications give the construction industry useful insights into how to create a monitoring system to identify any threat or vulnerability in real-time and simultaneously calculate the risk, based on which an action can be automatically elicited. For example, a risk monitoring system can be installed in the computer controlling the 3D printing process. This system synchronously monitors the network data flow of the computer, parameters of the 3D model in STL form, and parameters of the physical printing components. Once these parameters are tampered with and recognized, the system itself is able to decide whether to stop printing based on the risk value it calculates. From this literature review, it can be seen that the automation of risk analysis and the use of ML techniques, especially DL, is gaining much attention. It is then reasonable to assume that AI applied to cyber risk analysis in construction has a bright future.

3. SWOT analysis

To get a better understanding of the potential of AI applied to cyber risk analysis in construction, we conducted a preliminary SWOT analysis to facilitate a realistic approach to evaluate the construction industry’s position regarding the applications of ML in cyber risk analysis by identifying strengths, weaknesses, opportunities (i.e., positive external factors) and threats (i.e., negative external factors) of such applications. In addition, we considered internal (financial and human resources, tangible and intangible assets, operational efficiencies, etc.) and external (policies and regulations, market and technology changes, suppliers, etc.) factors and current and future strengths and weaknesses of the construction industry. The results are summarized in Fig 2.

3.1. Strengths

Strengths here are regarded as the internal factors rendering ML techniques competitive in cyber risk analysis. Since the construction industry is marching towards digitalization and automation, we consider ML techniques as the backbone of cyber risk analysis in the future due to these strengths. The main benefits of ML are as follows: (1) ML can handle the volume. ML can analyze a very large volume of activities across a construction company’s network and the massive volume of digital information generated, such as emails, files, etc., quickly. (2) ML can learn over time. ML can identify malicious attacks based on the behaviors of applications and the network as a whole. Over time, ML cybersecurity solutions learn about a network’s regular traffic and behaviors, providing a reactive approach to attacks. (3) ML can identify unknown threats. By using clustering algorithms, ML can successfully identify previously unknown attacks and update itself with the new information of these attacks.
3.2. Weaknesses
Weaknesses, like strengths, are inherent features of ML techniques and indicate the shortages that hinder its applicability and advancement. Only by identifying the weaknesses and solving them could ML techniques develop up to a higher level. Identified weaknesses include: (1) The hardship of high-quality data acquisition. The quality of input data is the cornerstone of ML applications. However, obtaining high-quality data is difficult, and the process is complex [31]. (2) The cost of error. False positives require analysts to spend expensive time inspecting, and the damage caused by false negatives is also non-negligible [32]. (3) The difficulty of deploying and maintaining ML modules. Some tools used for implementing ML algorithms were newly launched (e.g., TensorFlow and PyTorch packages were launched in 2015 and 2017, respectively), which might not be compatible with the older versions of computer configuration; thus, the deployment of ML might necessitate the redesign of the relevant business processes, which is expensive for construction companies [33]. (4) The lack of skilled personnel in construction. Since ML is a relatively new technology, the construction industry might lack the skilled personnel familiar with the implementation and utilization of ML, which requires the organization to add new positions to existing staff or hire new staff, thus increasing the cost.

![SWOT analysis of ML application in cyber risk analysis in the construction industry.](image)

3.3. Opportunities
Opportunities here are positive external factors that help realize the set goals of ML techniques or advance them to achieve higher performance in cyber risk analysis in construction. Identifying these opportunities and seizing them at the right moment is critical for realizing the automation of cyber risk analysis. In this study, opportunities are identified as follows: (1) Better supporting infrastructures for ML to create intelligent systems. With the upgrading of software and other configuration that support ML implementation, such as better CPUs and GPUs, utilizing ML techniques to create a fully intelligent system to perform part or the whole cycle of cyber risk analysis is prospective. For example, ML-IDS (Intrusion Detection System) integration can be realized and applied to digital devices like sensors in the construction field and data-rich computers to detect intrusions [34]. (2) The availability of more data. ML is a data-driven technique, thus to achieve higher performance, more data is needed. Like in other industries, digitalization makes it possible for the construction industry to store data of virtually any software, equipment, and processes, such as CDE (Common Data Environment), BIM software, and SCADA systems, which could be sufficiently utilized to feed the ML algorithms. (3) The willingness of organizations to invest in ML. To facilitate the growth of organizations in the digitalized era, the decision levels are more willing (forced) to take action against cyberattacks swiftly. Knowing the prospect of ML, they tend to invest in better preparing existing staff, hiring new staff, purchasing advanced ML software, etc., thus iteratively propel the development of ML in the cyber risk analysis area [35].
3.4. Threats
Threats here refer to external factors that can negatively affect the implementation or the performance of ML techniques from the outside. It is vital to anticipate these threats and take action against them before ML becomes a victim of them and its growth stalls. The threats identified in this study include:
(1) Cyber threats are constantly evolving. As a significant number of new threats emerge, security solutions that use ML have to be regenerated to keep up, thus increasing labor and cost. (2) Cyber criminals use AI, too. Cybercriminals can use ML techniques for illegal information gathering of people and IT assets, malicious user impersonation through phishing emails, unauthorized access to social accounts or machines, etc. (3) ML system manipulation. A true ML system never ceases to learn for improvement by taking constant feedback from the environment. The attacker can misuse this property by steering the system in the wrong direction by providing falsified data as feedback to the system [36].

4. Recommendations
Focusing on the weaknesses and threats from the SWOT analysis (Section 3), we propose some recommendations for construction project managers that could be considered as a starting point to overcoming the weaknesses and threats when considering ML techniques for cyber risk analysis in their projects.

4.1. Recommendations to address weaknesses

4.1.1. Recommendations to address the hardship of high-quality data acquisition. All too often, the data collected for cybersecurity risk analysis is unstructured. To get high-quality data, data processing is necessary. The data processing phase consists of all the operations, from collecting the raw data to forming the final data set fed to the modeling tool [37]. These operation tasks may be performed multiple times in no fixed order, which includes: (1) Standardize the data collection process in the construction industry; (2) Clean and normalize the data to make the range of data value consistent; (3) Discard the redundant information by dimensionality reduction analysis; (4) Perform feature transformation to create structured data; (5) Conduct sample segmentation and sample balancing operations; (6) Augment the data to increase the diversity of data sources.

4.1.2. Recommendations to address the cost of error. To increase the precision of the ML model and prevent false positives and false negatives, several methods can be considered: (1) For a model whose accuracy rate is below 85% (or any threshold based on the organization’s risk tolerance), personnel should take part in the decision-making stage, combined with the decision AI suggests. (2) Consider results from different models and take the final decision with the most votes. (3) Construction companies can install an extra firewall to achieve the goal of multi-protection purposes. (4) Improve the performance of the model [38], which includes performing sufficient data processing as mentioned in section 4.1.1, adopting the K-fold cross-validation method to utilize the samples as many times as possible, designing different loss functions to find the best loss function, and adjusting the parameters consistently, including hyper-parameters, ways of optimization.

4.1.3. Recommendations to address the difficulty of deploying and maintaining ML modules. There are three recommendations regarding this weakness: (1) For middle-sized and larger construction companies, the board members should enhance their awareness of cybersecurity issues [39] and proactively take actions to improve their existing systems to accommodate ML modules; (2) For companies that are prone to cyberattacks, directives should hire information security professionals to manage cyber risks to prevent losses ahead of time. (3) When ML modules are being designed, the structure of the model should align with the goal of the task. ML modules should not be excessively complex if simpler structures meet the demands as well.

4.1.4. Recommendations to address the lack of skilled personnel. There are two ways to address this issue, and construction organization should base their choice on available budget and internal organizational structure. The first recommendation is that the training sessions can equip the existing
employees with ML knowledge on its implementation and maintenance procedures. The second is that the company can hire (or subcontract) new ML professionals to deal with relevant cyber risk analysis tasks.

4.2. Recommendations to address threats

4.2.1. Recommendations to address the constantly evolving threats. Facing the rapidly changing threats, besides improving the ML models’ performance, construction companies must strengthen IT security mechanisms [40]. These mechanisms are listed as follows: Make sure systems, applications, and users are patched; Make specific regulations regarding the information sharing among owners, contractors, suppliers, and other parties; Implement a consistent security model regardless of user location or device type; Practice the rule of least access; Embrace the advanced terminal approach; Require secure applications; Sufficiently use threat intelligence; Train the employees to fully understand threat environment; Proactively use new security technologies.

4.2.2. Recommendations to respond to cybercriminals’ usage of ML. Facing different usages of ML by cybercriminals, different preventive actions should be taken. (1) To prevent illegal information gathering of people and IT assets, personal information and properties of confidential IT assets must not be available in open sources. (2) To avoid being the victim of malicious impersonation of users, email/text receivers should double-check with the users who send suspicious messages through other channels. (3) To prevent unauthorized access, construction personnel can choose a captcha like MathCaptcha or its alternatives for websites and machines and simultaneously use complicated passwords and any form of multi-factor authentication (MFA) [41].

4.2.3. Recommendations to address ML system manipulation. There are several ideas to prevent ML system manipulation: (1) ML experts should prevent the damage by minimizing the amount of training data cybercriminals can control. (2) ML engineers can address this issue by maintaining a record of data ownership and streamlining and securing system operations. In addition, (3) ML engineers can design new functions to prevent data extraction and model function extraction by hackers to avoid system poisoning [42].

5. Conclusions and Outlook

This study reviews recent publications on the applications of ML techniques in risk identification, risk estimate, and risk monitoring and shows its potential in the construction industry for cyber risk analysis. Based on the literature review, a preliminary SWOT analysis was performed, addressing the strengths, weaknesses, opportunities, and threats that might be faced by the construction industry when considering ML techniques. Regarding the weaknesses and threats, we propose some recommendations for addressing them. From the literature review, it is apparent that NN-based DL techniques are obtaining an increasing amount of attention, mainly due to their capability of dealing with Natural Language Processing (NLP) and high accuracy for task classification. To expand this preliminary study, the authors’ work focuses more on image classification for threat identification and text analysis for vulnerability identification by integrating current DL algorithms like RNN-based LSTM algorithm, Probabilistic Latent Semantic Analysis (PLSA), and simultaneously developing new architectures to create an intelligent integrated system combining all processes of risk identification, estimate, and management in the long run.

Acknowledgments

The authors want to thank the support provided by the Center for Cyber Security at New York University Abu Dhabi (CCS-AD).
References

[1] Kling R and Turk Z 2021 Construction 4.0 – digital transformation of one of the oldest industries Econ Bus Rev 21(3) pp 393–496

[2] Mantha BRK and de Soto BG 2019 Cyber security challenges and vulnerability assessment in the construction industry Proc. of Creative Construction Conf. 2019 pp 29–37

[3] Mott G. 2018 A critical reflection on the construction of the cyberterrorist threat in the United Kingdom of Great Britain and Northern Ireland Nottingham Trent University

[4] Zheng R, Jiang J, Hao X, Ren W, Xiong F and Zhu T 2019 CaACBIM: A context-aware access control model for BIM J. Inf. 10(2) pp 1–15

[5] Lin CT, Wu SL and Lee ML 2017 Cyber attack and defense on industry control systems 2017 IEEE Conf. on Dependable and Secure Computing pp 524–6

[6] Stapleton KAJ, Gledson BJ and Alwan Z 2014 Understanding technological interoperability through observations of data leakage in Building Information Modelling (BIM) based transactions Proc. of the 32nd eCAADe Conf. pp 515–24

[7] The rise of cyber risk in the construction industry 2020 Gallagher [cited 2020 Mar 15] Available from: https://www.aig.com/us/news-and-insights/2020/jun/cyber-risk-within-construction-industry/

[8] Cyber security breaches survey 2021 GOV.UK [cited 2021 Jun 2] Available from: https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2021#chapter-3-awareness-and-attitudes

[9] Cherdantseva Y, Burnap P, Blyth A, Eden P, Jones K, Soulsby H, et al 2016 A review of cyber security risk assessment methods for SCADA systems J. Comput. Secur. 56 pp 1–27

[10] Williams C. Risk monitoring: 6 considerations for understanding this make or break moment for ERM [Internet]. ERM Insights. 2019 [cited 2021 Jun 2]. Available from: https://www.erminsightsbycarol.com/risk-monitoring/

[11] Zhao G 2020 Information Security Management and Risk Assessment 2nd ed. ed. Liu X (Beijing: Tsinghua University Press) p 48

[12] Kurtz S 2019 Cybersecurity vulnerabilities in the construction industry [Internet]. Total IT Information Technology [cited 2021 Mar 19] Available from: https://totalit.com/cybersecurity-vulnerabilities-in-the-construction-industry/

[13] Pang Y, Xue X and Namin AS 2016 Early identification of vulnerable software components via ensemble learning 2016 15th IEEE Int. Conf. on Machine Learning and Applications (ICMLA) pp 476–81

[14] P RKV, S KRM, K JS and D PL 2017 Bat optimization algorithm for wrapper-based feature selection and performance improvement of android malware detection Proc. of the 33rd Annual Computer Security Applications Conf. pp 362–72

[15] Yuan F, Cao Y, Shang Y, Liu Y, Tan J, Fang B 2018 Insider threat detection with deep neural network International Conference on Computational Science pp 43–54

[16] Russell R, Kim L, Hamilton L, Lazovich T, Harer J, Ozdemir O, et al 2019 Automated vulnerability detection in source code using deep representation learning 17th IEEE Int. Conf. on Machine Learning and Applications (ICMLA) pp 757–62

[17] Ebrahimi M, Nunamaker JF and Chen H 2020 Semi-supervised cyber threat identification in dark net markets: a transductive and deep learning approach J. Manag. Inf. Syst. 37(3) pp 694–722

[18] Blanco-Medina P, Fidalgo E, Alegre E, Vasco-Carofilis RA, Jañez-Martino F and
Villar VF 2021 Detecting vulnerabilities in critical infrastructures by classifying exposed industrial control systems using deep learning. *J. Appl. Sci.* 11(1) pp 1–14

[19] Matsika E, Robinson M and Neill CO 2016 Risk assessment tool for analysing terrorist attack. *European Transport Conf. 2016 Association for European Transport (AET)* pp 1–22

[20] Bilge L, Han Y and Dell’Amico M 2017 RiskTeller: Predicting the risk of cyber incidents. *Proc. of the 2017 ACM SIGSAC Conf. on Computer and Communications Security* pp 1299–311

[21] Jiao J 2018 Discussion on the neural network model of comprehensive evaluation of computer network security. *J. Inf. Commun.* 10 pp 14–15

[22] Zhang X, Lashkari AH, Maleki N, Mudge J and Ghorbani AA 2019 Probabilistic classification for industrial cyber risk using deep neural network. *the 2019 Int.Conf. on Security and Management* (Las Vegas) pp 208–15

[23] Sun D, Wu Z, Wang Y, Lv Q and Hu B 2019 Risk prediction for imbalanced data in cyber security: a Siamese network-based deep learning classification framework. *Int. Joint Conf. on Neural Networks (IJCNN)* pp 1–8

[24] Kalinin M, Krundyshev V and Zegzhda P 2021 Cybersecurity risk assessment in smart city infrastructures. *J.MDPI* 9(4) p 78

[25] Chung K, Kamhoua CA, Kwiat KA, Kalbarczyk ZT and Iyer RK 2016 Game theory with learning for cyber security monitoring. *2016 IEEE 17th Int. Symp. on High Assurance Systems Engineering (HASE)* pp 1–8

[26] He Y, Mendis GJ and Wei J 2017 Real-time detection of false data injection attacks in smart grid: a deep learning-based intelligent mechanism. *J. IEEE Trans. Smart Grid* 8(5) pp 2505–16

[27] Subroto A and Apriyana A 2019 Cyber risk prediction through social media big data analytics and statistical machine learning. *J. Big Data* 6(1) pp 1–19

[28] Li Q, Meng S, Zhang S, Wu M, Zhang J, Taleby Ahvanooey M, et al 2019 Safety risk monitoring of cyber-physical power systems based on ensemble learning algorithm. *IEEE Access* 7 pp 24788–805

[29] Sakthivel RK, Nagasubramanian G, Al-Turjman F and Sankayya M 2020 Core-level cybersecurity assurance using cloud-based adaptive machine learning techniques for manufacturing industry. *J. Trans. Emerg. Telecommun. Technol.* p e3947.

[30] Garrido JS, Dold D and Frank J 2021 Machine learning on knowledge graphs for context-aware security monitoring. *arXiv Prepr arXiv2015.08741 Available from: http://arxiv.org/abs/2105.08741*

[31] Goh GD, Sing SL and Yeong WV 2021 A review on machine learning in 3D printing: applications, potential, and challenges. *J. Artif. Intell. Rev.* 54(1) pp 63–94

[32] Yellenki RB 2020 Top 8 challenges for machine learning practitioners. *Towards Data Science* [cited 2021 Apr 1] Available from: https://towardsdatascience.com/top-8-challenges-for-machine-learning-practitioners-c4c0130701a1

[33] Paleyes A, Urma RG and Lawrence ND 2020 Challenges in deploying machine learning: a survey of case studies. *arXiv Prepr arXiv* pp 1–21. Available from: https://arxiv.org/abs/2011.09926

[34] Gassais R, Ezzati-Jivan N, Fernandez JM, Aloise D and Dagenais MR 2020 Multi-level host-based intrusion detection system for Internet of things. *J. Cloud Comput.* 9(1) pp 1–16

[35] Das S, Dey A, Pal A and Roy N 2015 Applications of artificial intelligence in machine learning: review and prospect. *Int. J. Comput. Appl.* 115(9) pp 31–41.
[36] Tatwadarshipn 2021 Security threats to machine learning systems *Analytics Vidhya* 2021 [cited 2021 Jun 4] Available from: https://www.analyticsvidhya.com/blog/2021/01/security-threats-to-machine-learning-systems/

[37] Rassam MA, Maarof MA and Zainal A 2017 Big data analytics adoption for cybersecurity: a review of current solutions, requirements, challenges and trends *J. Inf. Assur. Secur.* **11** pp 124–45

[38] Schelter S, Biessmann F, Januschowski T, Salinas D, Seufert S and Szarvas G 2018 On challenges in machine learning model management *Bull. IEEE Comput. Soc. Tech. Comm. Data Eng.* pp 5–13

[39] Armenia S, Angelini M, Nonino F, Palombi G and Schlitzer MF 2020 A dynamic simulation approach to support the evaluation of cyber risks and security investments in SMEs. *J. Decis. Support. Syst.* **June 2020** p 113580

[40] Goldstein M 2020 4 Ways small businesses can protect themselves from cyber attacks [Internet]. *The Intersection of Business and Technology* [cited 2021 May 27]. Available from: https://www.coxblue.com/4-ways-small-businesses-can-protect-themselves-from-cyber-attacks/

[41] Polyakov A 2021 Machine learning for cybercriminals 101 *Towards Data Science* [cited 2021 Jun 4] Available from: https://towardsdatascience.com/machine-learning-for-cybercriminals-a46798a8c268

[42] Jagielski M, Oprea A, Biggio B, Liu C, Nita-rotaru C and Li B 2018 Manipulating machine learning: poisoning attacks and countermeasures for regression learning *2018 IEEE Symp. on Security and Privacy (SP)* pp 19–35