International Transactions on Education Technology (ITEE)
Vol. 1 No. 1 November 2022

p-ISSN: 2963-6078
e-ISSN: 2963-1947

Author Notification
22 November 2022
Final Revised
27 November 2022
Published
30 November 2022

Analysis of Machine Learning Systems for Cyber Physical Systems

Anggi Rachmawati[1], Yossaepurrohman [2]

PGSD (Pendidikan Guru Sekolah Dasar), UPBJJ-UT Bogor, Jl. Sholeh Iskandar No.234, RT.02/RW.11, Kedungbadak, Kec. Tanah Sereal, Kota Bogor, Jawa Barat 16164

e-mail:
anggirachmawati82@gmail.com
yossaepurrohman@gmail.com

To cite this document:
Rachmawati, A. & Yossaepurrohman. (2022). Analysis of Machine Learning Systems for Cyber Physical Systems . International Transactions on Education Technology (ITEE), 1(1),1-9.

Abstract

This study summarizes major literature reviews on machine learning systems for network analysis and intrusion detection. Furthermore, it provides a brief lesson description of each machine learning approach. Because data is so important in machine learning methods, this study The primary tools for assessing network traffic and spotting anomalies are machine learning approaches, and the study focuses on the datasets utilized in these techniques. This research examine the multiple advantages (reasonable use) that machine learning has made possible, particularly for security and cyber-physical systems, including enhanced intrusion detection techniques and judgment accuracy. Additionally, this study discusses the difficulties of utilizing machine learning for cybersecurity and offers suggestions for further study..

Keywords: Machine learning systems, Cyber-physical Systems, Security systems

1. Introduction

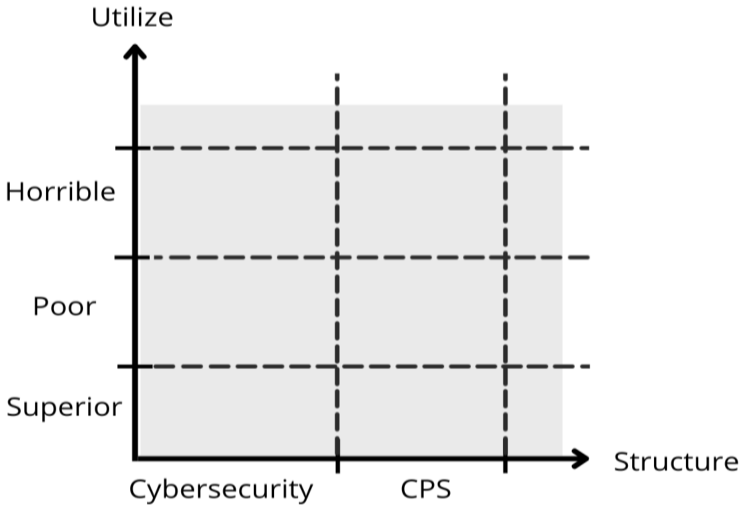
This study examines Applications for cyber security and reporting findings from a literature analysis of data mining and computer learning techniques. Furthermore, it describes Application of machine learning or data mining techniques to problems with cyber intrusion detection [1]. The study provides a set of comparative standards for data mining or machine learning techniques in addition to evaluating the difficulty of numerous algorithms for data mining or machine learning.Following that, best practices and a set of suggestions are provided based on the characteristics of the cyber issue[2]. The term "cyber security" refers to a group of techniques and tools used to stop unauthorized users from accessing, attacking, changing or eradicating computers, networks, data, and apps. The majority of cyber security systems are composed of network security systems and computer (host) security systems. Each of these entities must have a firewall, intrusion detection system, and at least one antivirus program (IDS). IDSs detect and assess unlawful use, change, duplication, and damage to information systems in addition to identifying them[3]Security breaches include both internal incursions (intra-organizational assaults) and external invasions (inter-organizational assaults).Because of the growth in the ratio of cyber-attacks, Security threat detection methods based on artificial intelligence as well as machine learning, a crucial part of our daily lives. [4]. Security standards are vital in providing the finest security applications while keeping implementation and security requirements in mind. In this regard, intrusion detection systems are critical for cyber security research[5]. Despite the fact that some research publications can validate specific datasets, machine learning techniques, and deep learning, there is little information in the literature about machine learning or data mining systems for cyber security and security-related datasets.[6].

This survey study emphasizes the methods (ML, DL, and DM) and their descriptions, with an emphasis on machine learning or data mining systems. [7]. Aside from various reviews, several articles including these approaches have been published. In comparison to previous assessments, the articles that meet specified criteria will be the primary focus of our paper. Google Scholar inquiries were performed using methods like as "machine learning," "cybersecurity," and "data mining". Among the key concerns were highly cited publications featuring well-known methodologie[8]s. Nonetheless, there was a recognition that unique and inventive approaches would be overlooked, thus a couple of these publications were also chosen. In general, articles were chosen with the goal of including at least one, ideally a few, representative studies from each of the machine learning or data mining areas[9].

2. Research Method

Machine learning applications give a distinctive technique to tackle fundamental scientific and technical challenges using computer software[10]. Machine learning has made enormous strides over the past two decades, and newbies can now easily use these technologies. Essentially, a "black-box" laboratory environment was used. Considering that machine learning has become a useful application, commercial businesses are gradually implementing it widelySoftware advancements in machine learning include robot control, speech recognition, natural language processing, computer vision, and other novel applications. Machine learning is used by large organizations like To improve customer experience, advertise special deals, and suggest products to buy, Amazon and Google. For machine learning developers, training a system is easier than creating the conventional input-process-output architecture. They do this by constructing simulations of the anticipated outcomes[11]. Machine learning is influencing a number of industries in the form of data-intensive problems like cybersecurity. Machine learning can also help other fields, such as biology, astronomy, and social science, to produce astounding outcomes[12]. Data from experiments can be processed and evaluated in unique ways using machine learning. Theoretically, working on machine learning algorithms can aid in our understanding of "big data" concepts. Additionally, these technologies may use deep learning and machine learning. Machine learning algorithms may be very different for various functions. But machine learning offers fresh ways to examine enormous amounts of data and create evolutionary systems. Moreover, it is possible to get higher optimization by multiple algorithm generations[13]. Processing massive amounts of data is a wonderful application for machine learning and cybersecurity. Networks and platforms areassaultable; susceptible. The number of tools used to scan and evaluate targets determines how successful these assaults are. Adversaries employ machine learning to make their attacks more successful. Peer-reviewed journals have lately released a number of evaluations of the security of machine learning and artificial intelligence.In addition, attention was made on current research on Machine learning techniques applied in the Internet of Things and intrusion detection for computer network security [4]. The key machine learning literature reviews and an explanation of each machine learning approach were Intrusion detection network analysis is described.

A complete a literature review of defensive strategies and security risks was offered during machine learning training and testing or inference from a data-driven perspective. Furthermore, informed on artificial intelligence security problems, particularly the supervised learning and reinforcement approaches. On the other hand, updates on security flaws and countermeasures were investigated throughout a machine learning-based system's life cycle, from training through inference[15].



Picture 1. The interconnections between Machine Cybersecurity education and CPS

To be more precise, the "superior" The phrase "use of machine learning" refers to the ethical or advantageous use of machine learning to enhance task-specific performance or the creation of advantageous applications. Examples include of enhancing medical diagnostics, enhancing speech and computer vision services, and among many others. In a similar vein, "bad" machine learning is characterized as possible dangers throughout the Data gathering, preprocessing, training, deployment, and decision-making processes, as well as training procedures (i.e. training, testing, validation, and implementation) These are open to exploitation and subversion, leading to decision-making and machine learning models that are incorrect or completely untrustworthy[16]. Last but not least, we describe "awful"Machine learning as detrimental apps or software that, among other things, help in the subversion of automated computer systems and security detection measures, increase the efficacy of malware, or otherwise apply machine learning in a harmful or undesired way.

3. Findings

3.1 Cyber Security : Machine Learning

Even while training and evaluating machine learning models takes a lot of time and work, they can nevertheless be used to make simple inferences. Additionallybigger, more representative datasets during training significantly enhances models for machine learning. To be able to keep systems secure and to notify system administrators and users of illegal access so that additional measures (attribution, mitigation, and removal) may be taken performed, intrusion detection is a vital area for active investigation and discovery. The next section goes into further depth on how to use various machine learning techniques to identify intrusion threats.

An established technology called the Intrusion Detection System (IDS) keeps an eye on secured systems and networks for illicit activity. It is a crucial technique for safeguarding system security and digital infrastructures. This research examines how various machine learning algorithms may be used to detect both abuse and abnormalities. The addition The use of machine learning has enhanced IDS performance and made it possible to detect unexpected intrusions..

a. Misuse Detection

Misuse detection in traditional intrusion detection systems analyzes current actions to massive databases of threat signatures. Misuse detection, in essence, detects irregularities in existing attack records. A variety of machine learning algorithms have been developed for this procedure. For example, Integrated Artificial Neural the performance of artificial neural networks (ANNs) for abuse detection, in particular by employing supervised and unsupervised processes intrusion detection systems (IDSIn The learning model is trained using a labeled training dataset in the supervised learning approach. The knowledge model is trained using an unlabeled dataset in the unsupervised learning method, and the Self-Organizing Map (SOM) is then used by the model to automatically finish the results without the need for human intervention. Also exhibited was an A three-layer, two-learning ANN model methods. Their assessment findings revealed that their IDS performance achieved 99.4% with the ANN model (compared to 96.3% without ANN).

Additionally, a lot of research has concentrated on lowering the learning curve to boost abuse detection's effectiveness. Subba et al., for example, suggested an ANN-based abuse detection method that makes use of a system-based intelligent agent. When learning new and previously undisclosed audit records, the intelligent agent may quickly spot both aberrant and regular data exhibit underlying patterns. The proposed approach could outperform current intrusion detection models using the C4.5 algorithm, Naive Bayes, and SVM. By keeping track of crucial traits in a number of files and training on a dataset of anomalous properties, the recommended method for identifying abuse locates abnormal system behaviors. Additionally, the technique speeds up detection and the characteristics that were extracted from the trace files are auditable data.

b. CPS Machine Learning

In addition to cybersecurity, control, networking, and computing issues in CPS may be addressed using machine learning. This section will review current studies using machine learning for CPS, followed by obstacles and possible routes for further research.

1) Fixing CPS Security Concerns

CPS systems and applications prioritize security and privacy. Due to the essential infrastructures that Industry 4.0, smart homes, smart grids, and many other applications of the "smart world" monitor and govern, effective and long-lasting security measures are necessary. Cyberattacks against CPS also commonly share characteristics, such as fake data injection attacks, also known as data integrity attacks, which try to deceive the targeted CPS, such as smart grid and smart transportation, among others, by changing or adding false data. For instance, it has been demonstrated that false data injection attacks have a considerable impact on important power grid functional aspects such as state estimate, optimal power flow, and energy price, among others. As the scope of data gathering and distribution inside CPS increases, malicious data samples will unavoidably be obtained. Machine learning technologies can analyze massive input datasets and enhance CPS security because to their unmatched analytical capabilities. We may train machine learning algorithms to spot cyberattacks by providing them with damaging data or distinctive characteristics as inputs. For instance, attacks on data integrity are important vulnerabilities in power grid systems, which make up a shared energy infrastructure. CPS An et al. provided a method for detecting data integrity breaches in the AC power grid based on deep reinforcement learning.Their studies concentrated on making use of a target network and the main network both using deep-Q-network detection (DQND)to enhance the defensive strategy.The effectiveness of the learning process was increased by quantifying the observation space sliding window. The recommended strategy performed better than several current detection algorithms in various IEEE bus systems, according to their testing findings.. Aside from that, privacy issues specific to machine learning technology exist, especially when it comes to distributed machine learning systems. Significant privacy issues exist when transferring between dispersed computer units and a centralized server for data and parameters Aside from that, privacy issues specific to machine learning technology exist, especially when it comes to distributed machine learning systems. Significant privacy issues exist when transferring information and settings between dispersed computer units and a centralized server[17].

2) The Difficulties of Using Machine Learning for CPS Safety

In general, numerous conditions must be met before constructing an efficient machine learning model. It is vital to strike a compromise between strong detecting capabilities and resource usage. Furthermore, in order to provide correct results, machine learning models must be fed relevant and timely high-quality data. Furthermore, data streams provide constant input for real-time detection but are challenging to verify in real-time. It is increasingly difficult to identify and detect malware and harmful code due to its diversity and rapid expansion. Malware may spread to 3 million new samples in an hour, and some developing assaults can launch at different speeds while evading end-point detection. The machine learning training procedure is also frequently time-consuming, and although it is possible to find new infections in the field, doing so requires a trained model. Attackers regularly use malware detection technology to test their programs against, and such detection is obviously unexpected. Simply said, malware is always evolving, making it challenging to create and maintain well-trained machine learning models that would enable effective detection. Some of the aforementioned issues may be resolved by using high-dimensional data and incremental learning for non-stationary data. Malware fingerprinting can benefit from The potential for machine learning can be used in many systems for various applications, includes. However, analyzing high-dimensional data requires expensive computer resources. Online learning platforms have been created to manage incremental learning and speed up instruction. "Liang et al."l, for instance, looked at an online learning method that updated machine learning models based on slices of continuous data in order to dynamically fit new datasets. Additionally, distributed learning systems can improve computing capabilities, offering another solution[18].

3) CPS Performance Improvement

Remember that machine learning has shown a remarkable capacity to analyze data and provide unmatched insights and expertise. It is evident that The potential for machine learning can be used in many systems for various applications, includes wireless networks, smart transportation, smart grid, image and video recognition, and others. Machine learning may also be used to boost system automation in network and control systems. The use of artificial intelligence in CPS has been attempted in several recent research. The optimization of network scheduling has received particular attention. Jiang et al. researched available machine learning methods and employed a range of learning techniques to enhance IoT system network performance. They improved scheduling, specifically for the IoT network, using machine learning. Co-design or integration of networking and control elements in industrial IoT, energy production and monitoring, and traffic congestion avoidance, and a few more applications are only a few of the additional machine learning applications that help CPS operations.[19].

4) The Difficulties of CPS Machine Learning Utilization

There are issues with CPS applications, much like when machine learning is used for cybersecurity.First off, one or more scenarios may arise when a single machine learning model is unable to do all tasks. Most of the time, a machine learning model is trained for a specific issue and can only be retrained for a job that is comparable.. Furthermore, it is challenging to generalize a single machine learning model to account for every conceivable case since CPS differ. For system-wide solutions, a variety of models and data are therefore required.. There are issues with CPS applications, just like when machine learning is used for cybersecurity. Second, one or more scenarios may exist when a single machine learning model is unable to do all tasks. Typically, a machine learning model can only be retrained for a job that is comparable to the issue for which it was first trained. Furthermore, CPS varies, making it challenging to generalize a single machine learning model to account for all circumstances. For system-wide solutions, a variety of models and data are therefore required. Third, edge computing technologies might provide the CPS with crucial real-time decision-making and actuation, as well as flexibility and redundancy. Edge computing nodes' low computational capacity in comparison to cloud infrastructures, however, may prevent them from supporting the deep learning training process. As a result, significant problems being researched at the moment are how to improve deep learning models and systems to decrease the amount of computing needed and boost the effectiveness of distributed learning.. Furthermore, the development of error recovery algorithms for distributed machine learning is necessary since the It is crucial for deep learning models to be stable in edge computing infrastructures[20].

Last but not least, there are particular limitations on the types, amounts, and formats of input data for machine learning models[21]. Even if CPS gather enormous volumes of data, the data's quality cannot be guaranteed, particularly because it is uncertain how long newly designed IoT equipment will last. Machine learning systems either need to must naturally be able to deal with such raw data and noise, or they must be able to convert input data from raw data into a preset data format, an expensive operation.As was previously said, a variety of heterogeneous sensors continuously gather data in CPS systems, and handling the raw data is quite challenging.

3.2 Cyber Physical Sets of Data

a. ISOT (Information Security and Object Technology) Set of data

A total of 1,675,424 traffic flows made up of publicly accessible botnets and regular datasets make up the ISOT dataset. The Storm and Waledac botnets were among the honeynet project's French chapter's malicious traffic that was used in ISOT.[22].

b. CTU-13 (Czech Technical University) Set of Data

Czech Technical University's CTU-13 dataset is a collection of 13 distinct malware seizures in a nonfictional network environment. The dataset's goal is to collect real mixed botnet traffic.. Keeping the aforementioned in mind, ordinary hosts that have been confirmed produce normal traffic, whilst infected ones produce botnet traffic. One of the advantages of utilizing this dataset is that it is a thoroughly annotated dataset that captures processes performed in a controlled setting[23].

c. UNSW-NB15 Set of Data
It appears that the IXIA Perfect Storm program was used to create the UNSW-NB15 the Australian Centre for Cyber Security's Cyber Range Lab data set (ACCS). This dataset includes about an hour's worth of anonymous traffic records from a 2007 DDoS attack. Among the principal attacks included in this collection are fuzzers, backdoors, analysis, exploits, denial of service, reconnaissance, generic worms, and shellcode. [24].

4. Conclusion

The protection preventing assaults on computer systems is a critical national and international security problem. This article presents a survey on security problems using a variety of machine learning approaches. A variety of datasets have been used in various studies. Furthermore, Artificial intelligence and machine learning are important elements of computer system security. This article describes the literature review Among the data mining and artificial intelligence methods employed by Cybe. The sample articles that explain different machine learning and data mining Techniques in the cyberspace were picked with consideration, and extensive groups of various datasets were created, together with their benefits and drawbacks. Future free access to a new dataset will be made possible by this work..

References

[1] Q. Aini, N. Lutfiani, H. Kusumah, and M. S. Zahran, “Deteksi dan Pengenalan Objek Dengan Model Machine Learning: Model Yolo,” *CESS (Journal of Computer Engineering, System and Science)*, vol. 6, no. 2, p. 192, 2021, doi: 10.24114/cess.v6i2.25840.

[2] U. Rahardja, Q. Aini, A. Khoirunisa, N. Lutfiani, and D. Martika Sari, “A Blockchain Based Online Business Intelligence Learning System,” *Blockchain Frontier Technology*, vol. 1, no. 01, pp. 87–103, 2021.

[3] M. Chammas, A. Makhoul, and J. Demerjian, “Writer identification for historical handwritten documents using a single feature extraction method,” *Proceedings - 19th IEEE International Conference on Machine Learning and Applications, ICMLA 2020*, pp. 61–64, 2020, doi: 10.1109/ICMLA51294.2020.00010.

[4] V. P. Srinivasan, R. Kalidhasan, U. S. Jiviithesh, P. M. Logesh, and R. Kiran Prasad, “Design and Fabrication of Waste Sorting Robot,” vol. XIV, no. 11, pp. 62–66, 2021, [Online]. Available: http://www.paideumajournal.com

[5] R. Gupta, S. Tanwar, F. Al-Turjman, P. Italiya, A. Nauman, and S. W. Kim, “Smart Contract Privacy Protection Using AI in Cyber-Physical Systems: Tools, Techniques and Challenges,” *IEEE Access*, vol. 8, pp. 24746–24772, 2020, doi: 10.1109/ACCESS.2020.2970576.

[6] T. Hariguna, Y. Durachman, M. Yusup, and S. Millah, “Blockchain Technology Transformation in Advancing Future Change,” *Blockchain Frontier Technology*, vol. 1, no. 01, pp. 13–20, 2021, doi: 10.34306/bfront.v1i01.4.

[7] D. Ding, Q. L. Han, Z. Wang, and X. Ge, “A Survey on Model-Based Distributed Control and Filtering for Industrial Cyber-Physical Systems,” *IEEE Trans Industr Inform*, vol. 15, no. 5, pp. 2483–2499, 2019, doi: 10.1109/TII.2019.2905295.

[8] C. Greer, M. Burns, D. Wollman, and E. Griffor, “Cyber-Physical Systems and Internet of Things NIST Special Publication 1900-202 Cyber-Physical Systems and Internet of Things,” *NIST Special Publication 1900-202*, 2019, [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1900-202.pdf

[9] T. Fitz, M. Theiler, and K. Smarsly, “A metamodel for cyber-physical systems,” pp. 1–33.

[10] Sedjelmaci H., Fateh G., Sidi-Mohammed S., Hassnaa M., Jiajia L., and Shuai H., “Cyber SecurityBasedonArtificial Intelligence for Cyber-Physical Systems,” *IEEE Netw*, vol. 34, no. 3, pp. 6–7, 2020.

[11] D. Ding, Q. L. Han, X. Ge, and J. Wang, “Secure State Estimation and Control of Cyber-Physical Systems: A Survey,” *IEEE Trans Syst Man Cybern Syst*, vol. 51, no. 1, pp. 176–190, 2021, doi: 10.1109/TSMC.2020.3041121.

[12] T. Ayuninggati, E. Purnama Harahap, Mulyati, and R. Junior, “Supply Chain Management, Certificate Management at the Transportation Layer Security in Charge of Security,” *Blockchain Frontier Technology*, vol. 1, no. 01, pp. 1–12, 2021, doi: 10.34306/bfront.v1i01.3.

[13] S. Maesaroh, H. J. Permana, P. Dirgayusa Febrianaga, Noviyanti, and R. A. Pardosi, “Blockchain Technology in the Future of Enterprise Security System from Cybercrime,” *Blockchain Frontier Technology*, vol. 2, no. 1, pp. 1–8, 2022, doi: 10.34306/bfront.v2i1.88.

[14] I. Wijaya, E. Haryatmi, and A. B. Kurniawan, “Implementasi Teknologi Blockchain pada Sistem Presensi Staff VM LePKom Berbasis Web,” *Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 5, no. 1, pp. 162–169, 2020.

[15] U. Rahardja, Q. Aini, M. Yusup, and A. Edliyanti, “Penerapan Teknologi Blockchain Sebagai Media Pengamanan Proses Transaksi E-Commerce,” *CESS (Journal of Computer Engineering, System and Science)*, vol. 5, no. 1, p. 28, 2020, doi: 10.24114/cess.v5i1.14893.

[16] A. Adiyanto and R. Febrianto, “Authentication Of Transaction Process In E-marketplace Based On Blockchain technology,” *Aptisi Transactions On Technopreneurship (ATT)*, vol. 2, no. 1, pp. 68–74, 2020, doi: 10.34306/att.v2i1.71.

[17] U. Sharma, P. Astya, A. Baliyan, S. Krit, V. Jain, and M. Z. Khan, *Advancing Computational Intelligence Techniques for Security Systems Design*. 2022. doi: 10.1201/9781003229704.

[18] X. Gu and A. Easwaran, “Towards safe machine learning for CPS infer uncertainty from training data,” *ICCPS 2019 - Proceedings of the 2019 ACM/IEEE International Conference on Cyber-Physical Systems*, pp. 249–258, 2019, doi: 10.1145/3302509.3311038.

[19] B. K. Park, B. Jeon, and R. Y. C. Kim, “Improvement practices in the performance of a CPS multiple-joint robotics simulator,” *Applied Sciences (Switzerland)*, vol. 10, no. 1, 2020, doi: 10.3390/app10010185.

[20] T. Kotsiopoulos, P. Sarigiannidis, D. Ioannidis, and D. Tzovaras, “Machine Learning and Deep Learning in smart manufacturing: The Smart Grid paradigm,” *Comput Sci Rev*, vol. 40, no. March 2021, 2021, doi: 10.1016/j.cosrev.2020.100341.

[21] G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa, and A. Fazzio, “From DFT to machine learning: Recent approaches to materials science - A review,” *JPhys Materials*, vol. 2, no. 3, 2019, doi: 10.1088/2515-7639/ab084b.

[22] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, *Machine Learning and Deep Learning Approaches for CyberSecurity: A Review*, vol. 10, no. March 2021. Springer International Publishing, 2022. doi: 10.1109/ACCESS.2022.3151248.

[23] B. M. Rahal, A. Santos, and M. Nogueira, “A Distributed Architecture for DDoS Prediction and Bot Detection,” *IEEE Access*, vol. 8, pp. 159756–159772, 2020, doi: 10.1109/ACCESS.2020.3020507.

[24] S. Choudhary and N. Kesswani, “Analysis of KDD-Cup’99, NSL-KDD and UNSW-NB15 Datasets using Deep Learning in IoT,” *Procedia Comput Sci*, vol. 167, no. 2019, pp. 1561–1573, 2020, doi: 10.1016/j.procs.2020.03.367.