Available online at www.alphanumericjournal.com

alphanumeric journal
The Journal of Operations Research, Statistics, Econometrics and Management Information Systems

Volume 11, Issue 2, 2023

Received: December 13, 2023
Accepted: December 31, 2023
Published Online: December 31, 2023

AJ ID: 2023.11.02.MIS.03
DOI: 10.17093/alphanumeric.1404181
Research Article

Using Artificial Intelligence in the Security of Cyber Physical Systems

Zeynep Gürkaş Aydın, Ph.D.*

Assist. Prof., Department of Computer Engineering, Faculty of Engineering, Istanbul University Cerrahpasa, Istanbul, Turkiye, zeynepg@iuc.edu.tr

Murat Kazanç

M.Sc., Istanbul University Cerrahpasa, Istanbul, Turkiye, murat.kazanc@ogr.iuc.edu.tr

* İstanbul Üniversitesi-Cerrahpaşa Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü Üniversite Mahallesi Bağlariçi Caddesi No:7, 34320 Avcılar/İstanbul, Türkiye

ABSTRACT

The prominence of cyber security continues to increase on a daily basis. Following the cyber-attacks in recent years, governments have implemented a range of regulations. The advancement of technology and digitalization has led to the creation of new vulnerabilities that cyber attackers can exploit. The digitalization of facilities such as energy distribution networks and water infrastructures has enhanced their efficiency, thereby benefiting states and society. The modern sensors, controllers, and networks of these new generation facilities have made them susceptible to cyber attackers. While all forms of cyber-attacks are detrimental, targeting critical cyber-physical systems presents a heightened level of peril. These assaults have the potential to disrupt the social structure and pose a threat to human lives. Various techniques are employed to guarantee the security of these facilities, which is of utmost importance. This study examined the applications of machine learning and deep learning methods, which are sub-branches of artificial intelligence that have recently undergone a period of significant advancement. Intrusion detection systems are being created for the networks that facilitate communication among the hardware components of the cyber-physical system. Another potential application area involves the development of models capable of detecting anomalies and attacks in the data generated by sensors and controllers. Cyber physical systems exhibit a wide range of diversity. Due to the wide range of variations, it is necessary to utilize specific datasets for training the model. Generating a dataset through attacks on a functional cyber-physical system is unattainable. The study also analyzed the solutions to this problem. Based on the analyzed studies, it has been observed that the utilization of artificial intelligence enhances the security of cyber physical systems.

Keywords: Cyber Physical System, Deep Learning, Machine Learning, Cyber Security, Critical Infrastructures

.

1. Introduction

Cyber Physical Systems (CPS) refer to systems where computer-based algorithms control input and output processes, which are continuously monitored in real time (Wang et al., 2022). The demand for CPSs has been steadily rising in recent years. Cyber-physical systems encompass various examples such as smart factories, transportation systems, critical infrastructures, robotics, and internet of things-based systems. Cyber-physical systems are formed by the integration of sensors, controllers, network, and communication devices. Special emphasis should be given to Critical Cyber Physical Systems (CCPS), which are a subset of CPSs. The National Cyber Security Strategy and 2013-2014 Action Plan in our country have identified certain structures, including transportation, energy, communication, finance, health, water management, and critical public services, as critical infrastructures (Information Technologies and Communications Authority, 2013). In addition, he has conducted comprehensive research and formulated precise definitions for critical infrastructure in various countries across the globe. For example, critical infrastructure in Europe, according to the 114/2008/EC regulation, every system that is essential to maintain, the infrastructure that provides the vital functions of the society and protects the health, security, and economic and social well-being, is considered critical infrastructure (EU monitor., 2008). Given the diverse and complex nature of cyber-physical systems, ensuring cyber security poses a challenging problem. Attackers have access to a wide range of attack surfaces.

Upon examination of the studies, it is feasible to categorize attacks against cyber-physical systems into two primary classifications. The first component refers to the medium employed by the devices comprising the cyber-physical system to establish communication among themselves. This setting can be accurately referred to as a network. The network can be categorized as wired, wireless, or hybrid, a blend of wired and wireless technologies. Common network attacks include denial of service (DoS), distributed denial of service (DDoS), interception (Man in The Middle - MiTM), and reconnaissance attacks. These attacks can facilitate the execution of other nefarious actions by limiting system access, manipulating system operations, and collecting information (Geiger et al., 2020). In 2014, a malicious software known as BlackEnergy propagated by exploiting macros in Microsoft's Office software and executed Distributed Denial of Service (DDoS) attacks on energy infrastructure in different countries, resulting in the disruption of services. Furthermore, these attacks specifically target the physical components of the cyber-physical system. These attacks include malware, erroneous data injections, and device replication. Malicious software that infiltrates controllers can result in system malfunctions or complete destruction. As an illustration, in 2010, the Stuxnet worm infected the Iranian State-owned Uranium Enrichment facility and triggered explosions within the facility (Chen & Abu-Nimeh, 2011). A crucial aspect is that the facility is entirely isolated from the global network. The malware initially infected programmable logic controllers (PLCs) that oversee the operation of centrifuges. This was accomplished by exploiting vulnerabilities in the computers that manage the SCADA (Centralized Control and Data Acquisition) system, using flash memory as the entry point.

Various conventional techniques are employed to identify cyber intrusions. Intrusion detection systems (IDS) do not actively prevent an attack in progress, but they do

furnish valuable information regarding the occurrence of such attacks (Wang et al., 2022). There exist two primary categories: Network-based intrusion detection systems (NIDS) operate on network devices exclusively, without requiring access to the end devices. In contrast, host-based intrusion detection systems (HIDS) are deployed on servers and workstations. In the second approach, if the data being transmitted is encrypted, it is possible to detect the data arriving at the server because the password is decrypted at the application layer of the OSI reference model. Firewall devices are strategically placed between the external and internal networks to protect network traffic by implementing rule-based protection against potential attacks.

Statistical methods are employed to detect anomalies in data security derived from sensors and controllers in cyber-physical systems. Signature-based methods can be employed to detect malware capable of infecting devices. In the present era, attackers seek novel vulnerabilities and formulate innovative techniques. This process can be considered a race. Novel approaches must be developed to enhance cyber security, continuously focusing on advancing system security. Deep learning and machine learning applications have become more prevalent in various industries and aspects of daily life. Machine learning and deep learning are specific branches of Artificial Intelligence that prioritize the development of systems capable of acquiring knowledge from existing data, identifying patterns, and making logical decisions autonomously or with minimal human involvement (Wazid et al., 2022). Artificial intelligence has the capability to enhance the effectiveness of cyber security techniques and facilitate the identification of zero-day attacks with reduced reliance on human intervention.

The use of artificial intelligence in the domain of cyber security yields significant advantages. However, to train machine learning and deep learning models effectively, it is necessary to have datasets that contain large amounts of high-quality data. The primary issue at discussion is the need for confidentiality in the data produced by cyber physical systems, particularly in critical cyber physical systems. Another concern arises from the fact that enabling attackers to target an operational system to generate the dataset is difficult. Scientists have created simulations, emulations, and test environments to solve this issue.

An examination of the existing body of literature indicates that a considerable number of studies have been undertaken utilizing obsolete and insufficient data sets. To the greatest extent possible, this study incorporates research that utilizes current data sets. This research includes investigations into various categories of cyber-physical systems. Insufficient research has been conducted in our country regarding the security of cyber-physical systems. In the realm of digitalization research, the development of a dataset and the training of artificial intelligence models pertaining to the modernized systems are essential.

The rest of the paper is organized as follows: In the "Related Work" section, we review existing literature and highlight the contributions of previous studies. Subsequently, the "Test Environments" section elucidates the experimental setups and conditions employed. The "Datasets" section delves into the datasets' intricacies, underscoring their relevance and significance. The "Machine Learning and Deep Learning Methods" section details the methodologies and algorithms harnessed in our research. Lastly,

the "Conclusion" section synthesizes our findings, discusses their implications, and outlines potential avenues for future research.

2. Related Work

Within the realm of machine learning and deep learning algorithms, researchers have approached the problem in two fundamental manners. These tasks encompass the categorization of attacks and the detection of anomalies in the flow of data.

Choosing an appropriate defense strategy for protecting Electrical CPSs and Smart Grids from attacks is a challenging task. It is essential to prioritize the examination of the measurement system across multiple subsystems to protect against diverse cyber attacks. To create an innovative approach, the suggestion is to utilize Lambda calculus to design the security strategy, considering both classical and quantum perspectives (D. et al., 2023). The study presents a Quantum Machine Learning technique to estimate vulnerability, exploit and execute strategies, and evaluate attack probability. Quantum threats, such as local and non-local interactions, have been employed to trigger attack and mutual attack events through quantum causal connections. These threats include methods like False Data Injection, which exploits quantum entanglement. The Quirk simulator was utilized to evaluate the efficacy of the developed model in addressing cyber attacks on power system networks.

A water distribution system test environment, which accurately represents Cyber-Physical Systems (CPS), was developed by (Perrone et al.,2021). Data was collected from this test environment under 15 different scenarios. The training process involved utilizing the generated dataset to train five distinct machine learning models across three distinct system scenarios. In the first scenario, two classifications were implemented to identify an abnormal situation. In the second scenario, various categorizations were employed to determine the nature of the cyber attack. Multiple classifications were employed to detect event-oriented destructive attack types, such as cyber sabotage or accidents, in the third scenario. Following the evaluations, Random Forest appeared as the most successful algorithm, achieving an accuracy rate of nearly 100%.

False data injection attacks are one of the worst attacks in smart grid systems and have been widely implemented recently. Attackers inject false data into the measurement, control, and calculation station in the smart grid network system. An attacker can easily change/manipulate the measured data in the smart grid system and transmit it to the control center via field devices. Decisions made using this data are dangerous enough to render the cyber-physical system inoperable. (Habib et al., 2023) conducted their study with statistical methods and supervised and unsupervised machine learning methods, Linear Regression was the most successful method with 100% accuracy.

Internet of Things (IoT) devices employed in the Cyber-Physical Systems (CPS) framework have limited resources such as processing power and energy (Wang et al., 2022). Proposed is a knowledge distillation and triplet convolution neural network approach called KD-TCNN, which is both lightweight and effective. This approach aims to enhance the performance of the model, accelerate anomaly detection for CPS, and reduce model complexity. The model was evaluated using the CIC IDS2017

dataset. The computational cost and size of this model are reduced by approximately 86% compared to the original benchmarking model, while experiencing only a 0.4 percent drop in accuracy.

The proposed approach by (Alguliyev et al., 2022) utilizes a combination of deep bidirectional gated recurrent unit and variational autoencoder model to effectively identify anomalies in a cyber-physical system. Due to the infrequency of anomalies in CPSs, the datasets generated by natural processes often exhibit an imbalance. Cyber Physical Systems (CPSs) store huge amounts of different types of data, which is frequently unlabeled, resulting in the limited effectiveness of machine learning techniques. The training performed on the SWAT dataset, consisting of 36 attack scenarios, resulted in an accuracy rate of nearly 100% and an F1 score of 87%.

A medical cyber-physical system, also known as MCPS, integrates medical sensor devices with cyber components to create a highly responsive system that plays an important part in maintaining security (Alrowais et al., 2023). The role of MCPS in hospitals is crucial as it involves identifying and protecting against attacks, ensuring the security of patients' medical information. The existing algorithms exhibit inefficiency and significant error rates. In this study, it is suggested to utilize artificial bee colony optimization and fuzzy C-Means algorithm to detect the attack and overcome the associated difficulties. The proposed model demonstrated a success rate of over 90% in countering remote server poisoning, physical brute force, and network penetration attacks conducted on the Cleveland dataset from the UCI repository. Furthermore, it demonstrated greater accuracy compared to conventional models.

Cyber Physical Power System (CPPS) has emerged as an advancement of the conventional power system through the implementation of control systems, computing units, and communication and information networks. The integration of them into CPPS has presented novel difficulties in ensuring security. From a machine learning viewpoint, this issue can be regarded as a multi-class classification problem (Lu & Wu, 2022). This study presents a model for detecting cyber attacks in CPPS (Cyber-Physical Production Systems) using collective learning to predict possible scenarios. In the proposed model, the data was divided into sections based on its sub-attributes and sent to the classifier for the ensemble. The preliminary classification process in this model is carried out by random vector functional link networks (RVFLNs), while decision trees are employed as the base classifier. A final classification process was conducted to determine the outcome by comparing various sub-feature spaces. Following the training, the proposed model achieved an accuracy of over 95%, surpassing the accuracy of conventional machine learning and deep learning approaches.

Digital Twins (DTs) are digital duplicates of physical objects that provide valuable information about physical processes and act as sources for collecting and distributing data in Cyber-Physical Systems (CPS). Furthermore, DTs offer an assessment platform to evaluate the functional performance and safety of the CPS without impeding ongoing operations. Their platform offers a controlled and supportive virtual training environment for security analysts, generating possible strategies to identify and address vulnerabilities in digital twins. In order to deal with the security issues associated with digital twins (DTs), the study conducted by (Suhail

et al., 2023) introduces a gamification approach known as "Securing Digital Twins through the Gamification Approach." In this game strategy, humans assume the role of the attacking team while artificial intelligence agents adopt the role of the defending team. The approach was tested testing and validation on two datasets, specifically SAD and ORNL, which consist of attacks on in-vehicle data as a representation of Cyber-Physical Systems (CPS). Azure IoT Hub was utilized to generate and gather sensor data (such as indicators, speedometers, and coolant levels) from DT environment devices for the purpose of evaluating the approach.

3. Test Environments

Testing environments are crucial for Cyber-Physical Systems (CPS). Executing real attacks on operational systems is extremely challenging. Attacks that halt the system, induce malfunctions, or result in physical damage are treated as unacceptable. Therefore, researchers have developed solutions such as producing synthetic data, conducting simulations, emulating scenarios, and developing authentic test environments. Another issue arises from the fact that most researchers do not readily share or only partially share the dataset they generate. This poses a challenge in terms of comparing studies. Researchers have employed one of the aforementioned methodologies, and occasionally a combination of them, to generate datasets for various CPSs.

Industry 4.0 relies on Internet of Things (IoT) technologies to connect devices and systems, with CPSs serving as distributed and decentralized backbone infrastructure. A test environment was established in (Funchal et al., 2020) to implement security mechanisms in a self-organizing cyber-physical conveyor system based on multi-agent systems (MAS) and using modular and smart conveyor modules. A machine learning-supported Intrusion Detection System (IDS) has been developed to analyze agent communication, monitor system events, extract indicators of intrusion, and help in decreasing cyber attacks (Shi et al., 2022). The study utilized a dataset obtained from a modern and genuine Industry 4.0 production system. A miniature manufacturing facility was established as part of this research study. The dataset encompasses seven distinct scenarios, encompassing normal operation, a sequence of cyber attacks, anomalies resulting from dissatisfied employees, and errors in production operations. Two datasets were generated by gathering both physical and network data from the test environment.

The authors (Marino et al., 2021) introduce a testbed that employs network flow and authentic industrial communication protocols to simulate ICT interactions in a wind-powered system. The testing environment is fully virtualized, employing emulations rather than physical devices. The benefits of the provided virtualized test environment are as follows: We offer the integration of authentic industrial protocols, network analysis tools, data engineering and machine learning tools. A comprehensive analysis of Cyber-Physical Systems (CPSs) is offered by gathering and examining both cyber and physical data simultaneously. A cost-efficient solution for prototyping and testing, capable of running on a single laptop, is available.

4. Datasets

Utilizing a significant amount of well-balanced, high-quality, and labeled dataset during training will greatly improve the possibility of success for the implemented models. Nevertheless, as previously stated, there are issues associated with the sharing of data sets. Upon examination of the research, it was observed that studies were still being conducted using low-quality data sets that failed to meet modern requirements, continuing for ages. Researchers have aggregated multiple published datasets into a unified dataset. Researchers have enhanced the dataset by generating synthetic data derived from actual data.

The thesis study conducted in (Turnipseed, 2015) involves the capture of network traffic from a SCADA system using the ModBus communication protocol. Subsequently, the Modbus ICS dataset was generated by extracting distinctive characteristics from the ModBus protocol data. The dataset comprises incidents encompassing denial of access, reconnaissance attacks, and incorrect operating parameters. The Tshark network analysis tool was employed to capture the network. The Matlab software processed 68 features from the captured traffic to generate a dataset (Frazão et al., 2019).

In the study carried out by Teixeira et al. (2018), researchers gathered network traffic data from the control system of a water storage tank, which is a component of the water purification and distribution process. Network traffic is analyzed to extract features, which are then used to create a dataset for training and testing various machine learning algorithms. The dataset contains examples of denial of access and reconnaissance attacks.

The WDT (Water Distribution Testbed) dataset comprises both physical and network data. This data is collected from a loop Water Distribution Testbed that simulates the flow of water between eight tanks using solenoid valves, pumps, pressure sensors, and flow sensors (Faramondi et al., 2021). The dataset is limited in size, yet it contains a wider range of cyber and physical attacks. Researchers have developed a balanced and complex dataset that is capable of presenting realistic situations. Thus, they have presented a smaller dataset that achieves an ideal balance between complexity and ease of use.

The testbed outlined in (Ferrag et al., 2022) consists of seven layers: Cloud Computing Layer, Network Functions Virtualization Layer, Blockchain Network Layer, Fog Computing Layer, Software-Defined Networking Layer, Edge Computing Layer, and IoT and IIoT Perception Layer. A dataset has been constructed heterogeneously to fulfill practical requirements. The test environment consists of a range of devices, including sensors, controllers, network devices, and computers, which are distributed across different layers. In addition, the Edge-IIoTset dataset included over 1000 extracted features, obtained from various protocols operating at different layers. Additionally, the dataset comprises various datasets and all the data is properly labeled.

Malicious activities and intrusion attacks targeting both local and satellite networks have been identified as significant security risks. Attacks targeting satellites can result in significant financial expenses. The STIN (Satellite Networks and Terrestrial Networks) dataset was generated to develop a detection methodology for countering

denial-of-service attacks in modern network settings. This was accomplished by using data collected from both satellite and terrestrial networks (Ashraf et al., 2022).

For this study, recent data sets and studies were prioritized. Table 1 lists the datasets utilized in this investigation as well as the artificial intelligence techniques that were implemented on those datasets.

Paper	Methods	Year	Data set
(D. et al., 2023)	Quantum ML	2023	Not Public
(Perrone et al.,2021)	DT, SVM, BYS, KNN and RF	2017	WADI (Ahmed et al.,2017)
(Wang et al., 2022)	DNN, CNN, RNN and Triplet CNN	2009 2018	NSL-KDD (Tavallaee et al.,2009) CIC IDS 2017 (*Sharafaldin et al.,2018*)
(Alguliyev et al., 2022)	LSTM, Bidirectional GRU and USAD	2022	SWaT (*Singapore University of Technology and Design,2022*)
(Alrowais et al., 2023)	Artificial Bee Colony, SVM, LSTM and Fuzzy C-Means	1989	UCI Cleveland (Detrano et al.,1989)
(Lu & Wu, 2022)	ANN, DT, RVFL, SVM and EnDTRVFL	2022	Not Public
(Suhail et al., 2023)	DNN	2020 2018	SAD (Verma et al.,2020) ORNL (Han et al.,2018)
(Teixeira et al., 2018)	KNN, DT, RF and SVM	2021	WUSTL-IIOT-2018 (Teixeira et al.,2020)
(Frazão et al., 2019)	DT, RF, NB, LR and KNN	2019	Modbus ICS (Frazão et al.,2019)
(Faramondi et al., 2021)	KNN, NB, SVM and RF	2021	Pysical WDT Dataset Network WDT Dataset (Guarino et al.,2021)
(Ferrag et al., 2022)	DT, RF SVM, KNN and DNN	2022	Edge-IIoTset (Ferrag et al., 2022)
(Ashraf et al., 2022)	RF, SVM, LR, MLP and RFMLP	2020	STIN Satallite STIN Terrestrial (Li et al.,2020)

Table 1. Datasets and Applied Methods

5. Machine Learning and Deep Learning Methods

Machine learning and deep learning algorithms, which fall within the umbrella of artificial intelligence, offer particular advantages. Deep learning algorithms succeed in acquiring knowledge of complex problems. Examples such as image classification or machine translation can be provided. Machine learning algorithms offer distinct benefits, including rapid processing speed and minimal system requirements for resources.

5.1. Machine Learning

Machine learning involves using mathematical and statistical algorithms to perform inference and classification tasks on new data, based on existing data and algorithms.

Typically, the steps involved in the machine learning process are as follows:

- Generating and preprocessing the dataset.
- Identifying and implementing machine learning models that are appropriate for the intended task.

- Model evaluation.
- Conducting re-model training with parameter adjustments, if needed, until the outcomes achieve the desired success metrics.

Machine learning encompasses three fundamental categories: supervised learning, unsupervised learning, and semi-supervised learning. Supervised learning requires a dataset that has been appropriately labeled. The primary objective is for the model to acquire the ability to determine the resulting output in response to the input provided during model training. The objective is for the model to correctly predict the appropriate output when presented with novel inputs that it has not previously encountered (Liu & Wu, 2012). A regression model is referred to as such in supervised learning when its output consists of continuous values. If the output consists of a limited and distinct set of values, it can be regarded as a classification model (Kazanç, 2022). A wide range of research exists that examines the security of various system types by implementing feature extraction and classification techniques such as (Ozogur et al., 2023). Unsupervised learning algorithms train the model solely based on the input data. The notion of output data is absent. The objective is to partition complex information into clusters or detect abnormal situations. A scenario might arise where we lack precise knowledge about the specific attributes we seek in these models.

5.2. Deep Learning

Deep Learning refers to algorithms that employ multiple layers to progressively extract more complex features from raw data, without the need for prior data preprocessing. Each layer contains a varying number of computational units known as neurons. Currently, deep learning is widely applied across various domains, incorporating diverse enhancements like memory and coding techniques. For instance, it finds applications in various domains, including computer vision, voice processing, translation, and future prediction. Deep Learning is a highly adaptable field that offers numerous possibilities for advancement. Researchers devise various solutions and methodologies through the development of novel algorithms. These solutions are also utilized in the field of cyber security.

Self-Organizing Maps (SOM) are used to represent the topological relationship of data accurately in order to identify any suspicious attacks. SOMs can be analyzed from the perspective of both static layered architecture and dynamic layered architecture (Qu et al., 2019). The utilization of a static layered architecture can significantly decrease the computational costs and proficiently depict the hierarchy of data. The dynamic layered architecture is well-suited for online infiltration detection due to its low computational latency, dynamic self-adaptation, and self-learning capabilities.

Quantum computers have the potential to significantly enhance general deep learning tasks by offering exponential acceleration for certain problems. Additionally, the use of a quantum neural network in classical deep learning can overcome the problem of memory bottlenecks. Hence, it is believed that it will fulfill the expected requirements of outstanding achievement and capacity for learning in the field of vulnerability detection (Zhou et al., 2022).

Multiple deep learning algorithms are employed together in specific studies. A novel approach for classifying network traffic data is introduced, which combines hierarchical LSTM (long short-term memory) and attention mechanisms for attack detection (Hou et al., 2022). HLSTM has been employed for extracting sequential features across hierarchical structures in multidimensional network data. Subsequently, the attention layer was employed to effectively capture the relationships among features and subsequently adjust the weights assigned to each feature.

6. Evaluation Metrics

During the model development process, there are three distinct phases: training, validation, and testing. The model is trained by utilizing a particular portion of the dataset during the training phase. The errors that occur during the training phase serve as a measure of how accurately the model aligns with the training data. Due to the adjustment of the model based on the training data, the success ratio may exceed that of the validation evaluation. During the verification phase, once the model has been trained, an assessment is conducted using data that the model has not been exposed to in the dataset. The testing phase uses external data that is not part of the dataset. The evaluation of this stage is naturally ambiguous due to the unknown nature of the data's class labels or outputs. The primary approach employed for assessing a classifier is the confusion matrix (Hou et al., 2022). For instance, a 2x2 confusion matrix illustrating four potential outcomes for a binary classifier is displayed in Table 2. When the data in the dataset is accurate and correctly predicts a positive outcome, it is referred to as a true positive (TP). A misclassified prediction is referred to as a false negative (FN), which is categorized as a type II error. Within the dataset, when the output is false and the prediction is positive, it is denoted as a true negative (TN). A positive prediction is considered a false positive (FP) and is known as a type I error (Tharwat, 2020).

		Correct Classes	
		Positive (P)	Negative (N)
Predicted Classes	True (T)	True Positive (TP)	True Negative (TN)
	False (F)	False Positive (FP)	False Negative (FN)

Table 2. Confusion Matrix

Accuracy (Acc) is a commonly used metric for evaluating classification performance. It quantifies the proportion of correctly classified outputs in relation to the total number of outputs in the dataset as shown in (1). However, if the distribution of the number of samples for each class in the data set is balanced, accuracy can be considered a suitable metric.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Sensitivity, also known as recall, measures the total number of positive outputs classified as true positives as shown in (2). When classes in the data set are not uniformly distributed, it is advisable to prioritize it. It provides insight into the

predictive accuracy of a class represented by a small number of samples in the dataset.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

Precision is the ratio of correctly identified positive outputs (true positives) to the total number of positive outputs, including both true positives and false positives and its formula is shown in (3). When classes are distributed unevenly within the data set, it is advisable to take advantage of it. This metric illustrates the precision with which the positive predictions of a category comprised of a limited number of samples are predicted.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

The F1 score is calculated as the harmonic mean of precision and sensitivity as indicated in (4). The F1 score ranges from 0 to 1, with higher values indicating better classification performance. When evaluating the success of a study employing classification algorithms, focusing solely on accuracy is among the most obvious errors. One primary reasoning for substituting the Accuracy metric with the F1 Score is to prevent erroneous model selections in datasets characterized by an uneven distribution of sample sizes across classes. Furthermore, the F1 Score may be chosen over other metrics due to its comprehensive inclusion of all error costs, not limited to False Positives and False Negatives.

$$F1\ score = 2\ x\ \frac{Precision\ x\ Sensitivity}{Precision + Sensitivity} \quad (4)$$

When determining which metric value should serve as the success criterion, the distribution of the number of samples for each class in the data set is the most important consideration. Accuracy is the more suitable metric for datasets in which the number of samples for each class is approximately equal. However, when dealing with datasets that contain an imbalanced number of samples, relying solely on the accuracy value would be highly erroneous. Recall, precision, and F1 score values should be evaluated for each class. Identical attempts are made to identify anomalies when IoT data sets are analyzed. Among the hundreds of thousands of network packets that are transmitted in an IoT network, for instance, only a few malicious packets are attempted to intercept. This circumstance results in imbalanced data sets.

7. Conclusion

The significance of protecting critical infrastructures against cyber threats has grown considerably during recent times. These infrastructures are crucial in providing essential services like energy, water, transportation, and communications. If they are targeted in an attack, it can lead to severe consequences. Consequently, ensuring the digital security of critical infrastructure has emerged as a paramount concern for countries.

In recent years, there has been a rise in the development of cyber security applications that utilize machine learning and deep learning techniques. These techniques are employed to detect and prevent attacks by analyzing extensive datasets. Deep learning and machine learning techniques are widely employed across various domains within the field of cyber security. Examples of usage areas include attack detection, authentication, security vulnerability detection, and data protection.

Employing machine learning and deep learning techniques in securing critical infrastructures enhances the speed and effectiveness of attack detection and prevention. These methods are more efficient in terms of time and resources for detecting attacks when compared to traditional methods. Nevertheless, there are certain challenges associated with employing these techniques in the field of cyber security. Data privacy is a matter of concern. Additionally, there is a requirement for datasets that comprise substantial quantities of high-quality data. Frequently, acquiring the dataset from actual systems is not possible. The capacity of data generated in test environments to accurately represent real systems is frequently inadequate.

As future work, we aim to create an openly accessible data set by establishing a test environment within the laboratory that simulates a dataset originating from our nation, which is currently insufficient. Contributing to scientific research with this dataset and the model training studies that will be conducted on it is the objective. Consequently, the issue of securing critical infrastructures from cyber threats has become a growing concern in contemporary times. Machine learning and deep learning are believed to have a significant impact on protecting these infrastructures. It is essential to thoroughly assess the proper application of these techniques and entrust the management of the procedure to professionals in the field of cyber security.

References

Ahmed, C. M., Palleti, V. R., & Mathur, A. P. (2017, April 21). WADI: a water distribution testbed for research in the design of secure cyber physical systems. Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks. https://doi.org/10.1145/3055366.3055375

Alguliyev, R., Sukhostat, L., & Mammadov, A. (2022, October 12). Anomaly Detection in Cyber-Physical Systems based on BiGRU-VAE. 2022 IEEE 16th International Conference on Application of Information and Communication Technologies (AICT). https://doi.org/10.1109/aict55583.2022.10013581

Alrowais, F., Mohamed, H. G., Al-Wesabi, F. N., Al Duhayyim, M., Hilal, A. M., & Motwakel, A. (2023, May). Cyber attack detection in healthcare data using cyber-physical system with optimized algorithm. Computers and Electrical Engineering, 108, 108636. https://doi.org/10.1016/j.compeleceng.2023.108636

Ashraf, I., Narra, M., Umer, M., Majeed, R., Sadiq, S., Javaid, F., & Rasool, N. (2022, February 21). A Deep Learning-Based Smart Framework for Cyber-Physical and Satellite System Security Threats Detection. Electronics, 11(4), 667. https://doi.org/10.3390/electronics11040667

Chen, T. M., & Abu-Nimeh, S. (2011, April). Lessons from Stuxnet. Computer, 44(4), 91–93. https://doi.org/10.1109/mc.2011.115

D., L., Nagpal, N., Chandrasekaran, S., & D., J. H. (2023, March). A quantum-based approach for offensive security against cyber attacks in electrical infrastructure. Applied Soft Computing, 136, 110071. https://doi.org/10.1016/j.asoc.2023.110071

Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J. J., Sandhu, S., Guppy, K. H., Lee, S., & Froelicher, V. (1989, August). International application of a new probability algorithm for the diagnosis of coronary artery disease. The American Journal of Cardiology, 64(5), 304–310. https://doi.org/10.1016/0002-9149(89)90524-9

EU monitor. (2008, December). Directive 2008/114 - Identification and designation of European critical infrastructures and the assessment of the need to improve their protection. Retrieved December 12, 2023, from https://www.eumonitor.eu/9353000/1/j9vvik7m1c3gyxp/vitgbgipfoqy

Faramondi, L., Flammini, F., Guarino, S., & Setola, R. (2021). A Hardware-in-the-Loop Water Distribution Testbed Dataset for Cyber-Physical Security Testing. IEEE Access, 9, 122385–122396. https://doi.org/10.1109/access.2021.3109465

Ferrag, M. A., Friha, O., Hamouda, D., Maglaras, L., & Janicke, H. (2022). Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning. IEEE Access, 10, 40281–40306. https://doi.org/10.1109/access.2022.3165809

Frazão, I., Abreu, P. H., Cruz, T., Araújo, H., & Simões, P. (2019). Denial of service attacks: Detecting the frailties of machine learning algorithms in the classification process. In Lecture Notes in Computer Science. Critical Information Infrastructures Security (pp. 230–235). https://doi.org/10.1007/978-3-030-05849-4_19

Funchal, G., Pedrosa, T., Vallim, M., & Leitao, P. (2020, July 20). Security for a Multi-Agent Cyber-Physical Conveyor System using Machine Learning. 2020 IEEE 18th International Conference on Industrial Informatics (INDIN). https://doi.org/10.1109/indin45582.2020.9478915

Geiger, M., Bauer, J., Masuch, M., & Franke, J. (2020, September). An Analysis of Black Energy 3, Crashoverride, and Trisis, Three Malware Approaches Targeting Operational Technology Systems. 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). https://doi.org/10.1109/etfa46521.2020.9212128

Guarino, S., Faramondi, L., Setola, R. & Flammini, F. (2021). A hardware-in-the-loop water distribution testbed (WDT) dataset for cyber-physical security testing. IEEE Dataport. https://dx.doi.org/10.21227/rbvf-2h90

Habib, A. A., Hasan, M. K., Alkhayyat, A., Islam, S., Sharma, R., & Alkwai, L. M. (2023, April). False data injection attack in smart grid cyber physical system: Issues, challenges, and future direction. Computers and Electrical Engineering, 107, 108638. https://doi.org/10.1016/j.compeleceng.2023.108638

Han, M.L., Kwak, B.I., & Kim, H.K. (2018). Anomaly intrusion detection method for vehicular networks based on survival analysis. Vehicular Communications, Volume 14, 2018, Pages 52-63. https://doi.org/10.1016/j.vehcom.2018.09.004

Hou, H., Di, Z., Zhang, M., & Yuan, D. (2022, May). An Intrusion Detection Method for Cyber Monintoring Using Attention based Hierarchical LSTM. 2022 IEEE 8th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS). https://doi.org/10.1109/bigdatasecurityhpscids54978.2022.00032

Information Technologies and Communications Authori. (2013, January). National Cyber Security Strategy and 2013-2014 Action Plan. Retrieved December 12, 2023, from https://www.btk.gov.tr/uploads/pages/2-1-strateji-eylem-plani-2013-2014-5a3412cf8f45a.pdf

Kazanç, M. (2022). Resim formatindaki dijital dokümanlarin bilgisayar görüsü ve makine öğrenmesi yöntemleri kullanilarak LaTex formatina çevrilmesi [MSc Thesis, İstanbul University-Cerrahpaşa].

Li, K., Zhou, H., Tu, Z., Wang, W., Zhang, H. (2020). Distributed network intrusion detection system in satellite-terrestrial integrated networks using federated learning. IEEE Access, vol. 8, pp. 214852-214865. https://doi.org/10.1109/ACCESS.2020.3041641

Liu, Q., & Wu, Y. (2012). Supervised Learning. Encyclopedia of the Sciences of Learning, 3243–3245. https://doi.org/10.1007/978-1-4419-1428-6_451

Lu, K. D., & Wu, Z. G. (2022, July 9). An Ensemble Learning-Based Cyber-Attacks Detection Method of Cyber-Physical Power Systems. 2022 International Conference on Advanced Robotics and Mechatronics (ICARM). https://doi.org/10.1109/icarm54641.2022.9959185

Marino, D. L., Wickramasinghe, C. S., Singh, V. K., Gentle, J., Rieger, C., & Manic, M. (2021). The Virtualized Cyber-Physical Testbed for Machine Learning Anomaly Detection: A Wind Powered Grid Case Study. IEEE Access, 9, 159475–159494. https://doi.org/10.1109/access.2021.3127169

Mitarai, K., Negoro, M., Kitagawa, M., & Fujii, K. (2018, September 10). Quantum circuit learning. Physical Review A, 98(3). https://doi.org/10.1103/physreva.98.032309

Ozogur, G., Erturk, M. A., Gurkas Aydin, Z., & Aydin, M. A. (2023, January 22). Android Malware Detection in Bytecode Level Using TF-IDF and XGBoost. The Computer Journal, 66(9), 2317–2328. https://doi.org/10.1093/comjnl/bxac198

Perrone, P., Flammini, F., & Setola, R. (2021, July 26). Machine Learning for Threat Recognition in Critical Cyber-Physical Systems. 2021 IEEE International Conference on Cyber Security and Resilience (CSR). https://doi.org/10.1109/csr51186.2021.9527979

Qu, X., Yang, L., Guo, K., Ma, L., Sun, M., Ke, M., & Li, M. (2019, October 2). A Survey on the Development of Self-Organizing Maps for Unsupervised Intrusion Detection. Mobile Networks and Applications, 26(2), 808–829. https://doi.org/10.1007/s11036-019-01353-0

Sharafaldin, I., Habibi Lashkari, A., & Ghorbani, A. A. (2018). Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. Proceedings of the 4th International Conference on Information Systems Security and Privacy. https://doi.org/10.5220/0006639801080116

Shi, L., Krishnan, S., Wen, S., & Xiang, Y. (2022). Supporting Cyber-Attacks and System Anomaly Detection Research with an Industry 4.0 Dataset. Network and System Security, 335–353. https://doi.org/10.1007/978-3-031-23020-2_19

Singapore University of Technology and Design (2022, June). Secure Water Treatment (SWaT). Retrieved December 20, 2023, from https://itrust.sutd.edu.sg/itrust-labs-home/itrust-labs_swat/

Suhail, S., Iqbal, M., Hussain, R., & Jurdak, R. (2023, October). ENIGMA: An explainable digital twin security solution for cyber–physical systems. Computers in Industry, 151, 103961. https://doi.org/10.1016/j.compind.2023.103961

Tavallaee, M., Bagheri, E., Lu, W. & Ghorbani, A. A., A detailed analysis of the KDD CUP 99 data set, 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, ON, Canada, 2009, pp. 1-6, https://doi.org/10.1109/CISDA.2009.5356528

Teixeira, M., Salman, T., Zolanvari, M., Jain, R., Meskin, N., & Samaka, M. (2018, August 9). SCADA System Testbed for Cybersecurity Research Using Machine Learning Approach. Future Internet, 10(8), 76. https://doi.org/10.3390/fi10080076

Tharwat, A. (2020, July 30). Classification assessment methods. Applied Computing and Informatics, 17(1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003

Turnipseed, I. (2015). A new scada dataset for intrusion detection research [Master of Science Thesis, Mississippi State University]. https://scholarsjunction.msstate.edu/td/209/

Verma, M.E., Iannacone, M.D., Bridges, R.A., Hollifield, S.C., Kay, B., & Combs, F.L. (2020). ROAD: The Real ORNL Automotive Dynamometer Controller Area Network Intrusion Detection Dataset (with a comprehensive CAN IDS dataset survey & guide). ArXiv, abs/2012.14600

Wang, Z., Li, Z., He, D., & Chan, S. (2022, November). A lightweight approach for network intrusion detection in industrial cyber-physical systems based on knowledge distillation and deep metric learning. Expert Systems with Applications, 206, 117671. https://doi.org/10.1016/j.eswa.2022.117671

Wazid, M., Das, A. K., Chamola, V., & Park, Y. (2022, September). Uniting cyber security and machine learning: Advantages, challenges and future research. ICT Express, 8(3), 313–321. https://doi.org/10.1016/j.icte.2022.04.007

Zhou, X., Pang, J., Yue, F., Liu, F., Guo, J., Liu, W., Song, Z., Shu, G., Xia, B., & Shan, Z. (2022, May 16). A new method of software vulnerability detection based on a quantum neural network. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-11227-3