Risk Accessment of Machine Learning Algorithms on Manipulated Dataset in Power Systems

Risk Assessment of Machine Learning Algorithms on Manipulated Dataset in Power Systems

1st Sayawu Yakubu Diaba
School of Technology and Innovations
University of Vaasa
Vaasa, Finland
sdiaba@uwasa.fi

2nd Miadreza Shafie-khah
School of Technology and Innovations
University of Vaasa
Vaasa, Finland
miadreza.shafiekhah@uwasa.fi

3rd Mike Mekkanen
School of Technology and Innovations
University of Vaasa
Vaasa, Finland
mike.mekkanen@uwasa.fi

4th Tero Vartiainen
School of Technology and Innovations
University of Vaasa
Vaasa, Finland
Tero.Vartiainen@uwasa.fi

5th Mohammed Elmusrati
School of Technology and Innovations
University of Vaasa
Vaasa, Finland
moel@uwasa.fi

***Abstract*—The emergence of the communication infrastructure in power systems has increased the variety and sophistication of network assaults. Intrusion Detection Systems' (IDS) importance has increased in relation to network security. IDS, however, is no longer secure when confronted with adversarial examples, and attackers can boost assault success rates by tricking the IDS. As a result, resilience must be increased. This paper assesses the Decision Tree, Logistic regression, Support Vector Machines (SVM), Naïve Bayes, K-Nearest Neighbours (KNN), and Ensemble's effectiveness. Using the WUSTL-IIoT-2021 dataset and CIC-IDS2017 dataset, we train the algorithms on the unmanipulated dataset and then train the algorithms on the manipulated dataset. Per the simulation results, the accuracy and prediction speed drop on the manipulated dataset while the training time rises.**

***Index Terms*—Communication infrastructure, Intrusion detection systems, Network security, Power systems**

I. Introduction

The growing interconnectedness of smart systems has led to increased concerns regarding their security vulnerabilities [1]. These systems are extensively utilized in several areas, including intelligent industries, smart cities, healthcare, and, many more, making security a critical issue [2]. Security measures such as authentication, authorization, encryption, Intrusion Detection Systems (IDS), and Intrusion Prevention Systems (IPS), have been employed to mitigate these vulnerabilities. Despite these efforts, these systems remain vulnerable to cyber-attacks [3].

The energy sector is not an exception and thus, the sector faces a wide range of cyber-attacks, including physical, internal, external, and cyber threats. Cyber threats, in particular, can emerge from any location and pose significant challenges for the industry. Although it is impossible to eliminate these threats [4], mitigation strategies can help reduce their impact. However, implementing these strategies often requires significant financial resources and effort. In addition, cyber-attack mitigations can affect negatively and create psychological consequences that can be detrimental to the industry, leading to a decline in performance and potentially affecting national economies. Therefore, it is crucial to prioritize and implement effective cyber-attack mitigation strategies while also considering their potential economic and psychological impacts [5].

Utilizing the power of information technology, smart grid technology uses a two-way communication system to intelligently distribute energy [6]. This enables the integration of green technologies to satisfy environmental requirements. However, the use of communication technology also leaves the system open to many security risks. Even though many survey studies have addressed these problems and suggested solutions, most of them categorize attacks according to how they affect confidentiality, integrity, and availability [7]. As the prevalence of cyber-attacks continues to rise in the smart grid [8], the need for reliable and accurate IDS becomes increasingly critical [9].

The authors of [10] classified attack scenarios into control-based and measurement-based attacks. The control-based attack includes altering or fabricating control signals sent to the targeted power system assets. It can directly result in frequency and transient voltage instability, line overloading, load reduction, and cascading failures. The measurement-based attack seeks to compromise measurements in order to conceal or falsely represent the system's current state, impair observability, and ultimately deceive operators or control systems.

It's critical to comprehend the possible repercussions [11] of data manipulation on the accuracy of IDS algorithms. Any system that relies on these algorithms to detect malicious traffic may be at risk if the data has been manipulated. As such, it is crucial to identify and address the challenges associated with data manipulation in IDS algorithms to improve their reliability and effectiveness. In this paper, we aim to explore the impact of data manipulation on the performance of machine learning algorithms. We will evaluate the strengths

and weaknesses of various approaches used to mitigate these effects. By examining the limitations and vulnerabilities of these algorithms in the face of manipulated data, we can develop more robust and effective systems to safeguard against cyber-attacks [11].

The remaining sections of the paper are structured as follows. The system model is presented in Section II, and the data description is summarized in Section III. The experimental analysis is presented in Section IV of the paper, and the conclusions are presented in Section V.

II. System Model

The modeling approach used in this paper is borrowed from Linnartz et. al. as shown in Fig.1. The figure depicts a simplified schematic summary of the crucial components of the Cyber-Physical Systems (CPS), including the manipulation strategy. Information and Communication Technology (ICT) is used to link the power system's assets (physical system) to the central control system (cyber system).

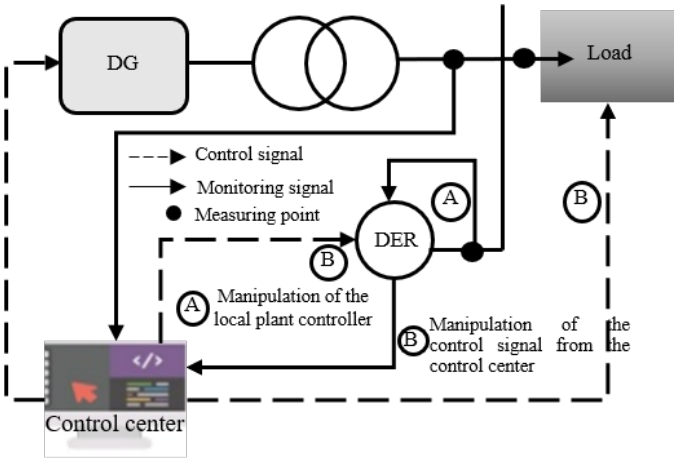

Fig. 1. Overview of the cyber-physical system. .

Supervisory control and data acquisition (SCADA) systems are used by Transmission System Operators (TSOs) and Distribution System Operators (DSOs) to continuously monitor and control the transmission and distribution assets of the power system (closed-loop control). Standardized protocols like IEC 61850, IEC 60870-5-104 or DNP3 are used for communication between the central management system and the assets. Typically, the information is transmitted through private channels, but neither of these protocols supports security features like authentication or integrity protection [12]. Thus, the communication channels are open to online threats.

For this experiment, the following assumptions are made.

- The attacker gains access and can interact with the assets according to the rules of the communication protocols. As a result, the attacker can alter control signals, create new ones, or stop all contact with the central control center.
- The attacker gained entry to the DSO's communication network. Due to earlier extensive reconnaissance operations, the attacker also has adequate knowledge of the control system, the power system, and its resources. He can then use the communication system to implement manipulation strategies by using denial of information attacks to transmit control signals to every connected DER.
- The attacker can attack and make data manipulations without the DSO noticing.

III. Data Description

Our proposed model's simulation evaluation stage utilizes the WUSTL-IIoT-2021 dataset [13] and the WUSTL-IIoT-2018 dataset. The WUSTL-IIoT-2021 dataset contains network data from an Industrial Internet of Things (IIoT) system. The dataset was developed using an IIoT testbed to support cybersecurity research, with the aim of emulating real-world industrial systems as accurately as possible. This testbed also allows for the simulation of actual cyber-attacks, providing a realistic environment for evaluating the effectiveness of the models. The normal and abnormal attacks are included in the WUSTL-IIOT-2018 dataset for ICS (SCADA) Cybersecurity research which closely matches true real-world data. It was constructed using a SCADA system testbed, which enabled the execution of an authentic cyber-attack [14].

A. Data pre-processing

The data preprocessing along with the data-cleaning process is depicted in Fig. 2. In the data pre-processing stage, columns with headings 'StartTime', 'LastTime', 'SrcAddr', 'DstAddr', 'sIpId', 'dIpId,' are removed from the data because they are specific to certain types of attacks. Including them in the training data would result in the algorithms being too specialized and not being able to accurately generalize to new, unseen data.

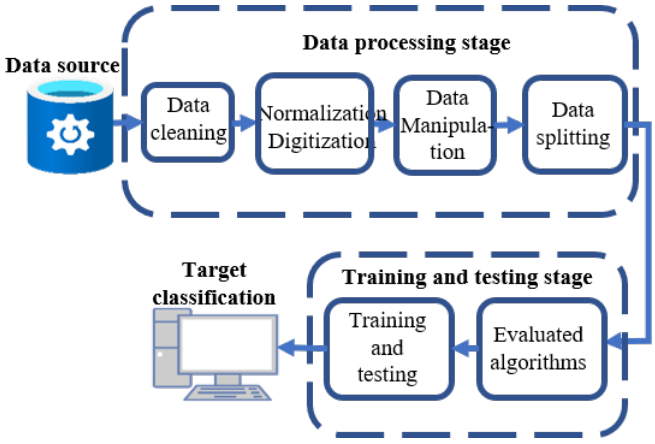

Fig. 2. The experimental model design.

Therefore, they were excluded to ensure the algorithms' ability to make accurate predictions on a wider range of input data. In the normalization and digitization stage, with respect to the target column, we represented normal with 0 and attack with 1. This column is manipulated by editing some 0's to 1

and some 1's to 0. The dataset is then split into a 70:30 ratio and used for the experiment.

However, Pearson's Correlation Coefficient (PCC) analysis is carried out on the dataset to perceive if it includes correlated features. This method was required because it guarantees the elimination of over-fitting. The PCC is mathematically given as

$$PCC = \frac{\sum (x_i + x_j)(y_i + y_j)}{\sum \sqrt{(x_i + x_j)^2 (y_i + y_j)^2}} \quad (1)$$

The symbol x_i represents the content of the variable in the dataset while x_j is referring to the average value of that variable. Similarly, y_i represents the values of the sample y_j represents the average value of that variable in the sample.

Correlation coefficients are statistical measures that describe the strength and direction of the relationship between two variables. A correlation matrix is typically square, with each variable listed on the table's rows and columns. The diagonal elements of the matrix are always 1 since each variable is perfectly correlated with itself, and the off-diagonal elements represent the correlation coefficients between the corresponding pairs of variables. A high positive correlation coefficient between two variables indicates that they tend to increase or decrease together, while a high negative correlation coefficient indicates that they tend to move in opposite directions. Correlation matrices score between -1 and +1 are often used to analyze the relationships between variables and identify data patterns and trends [15].

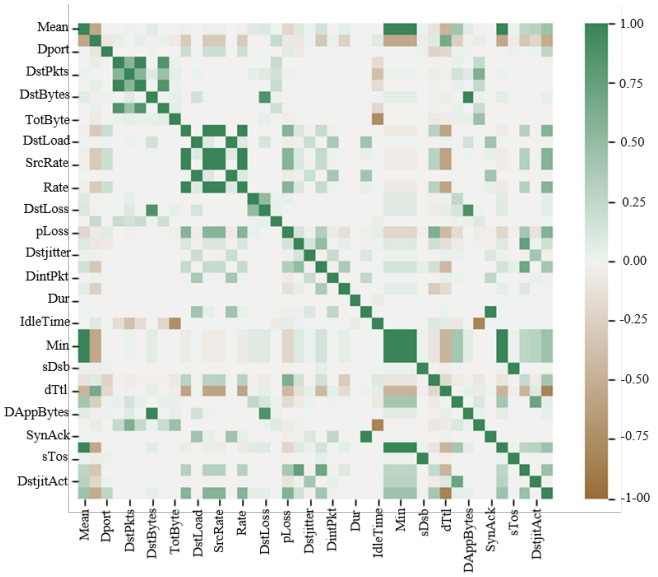

Fig. 3. The correlation matrix for the wustle-2021 dataset.

The effectiveness of the algorithm's categorization is assessed using a confusion matrix [16]. A confusion matrix is a method to evaluate how well a classifier algorithm is performing. It is composed of four fundamental parameters, including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [17].

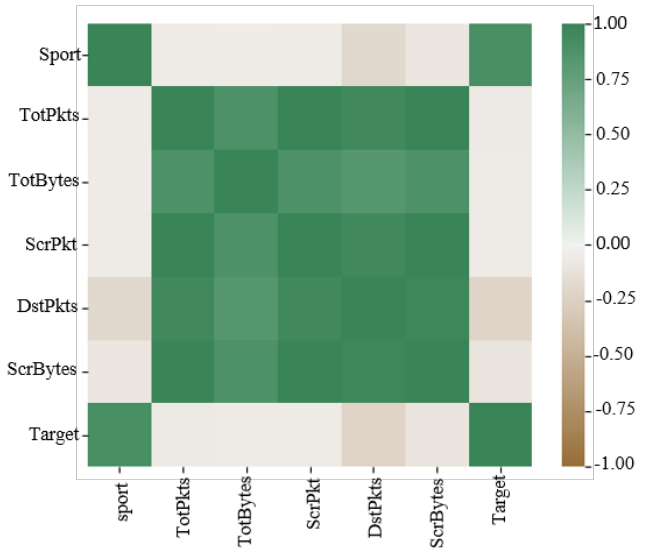

Fig. 4. The correlation matrix for the wustle-2018 dataset.

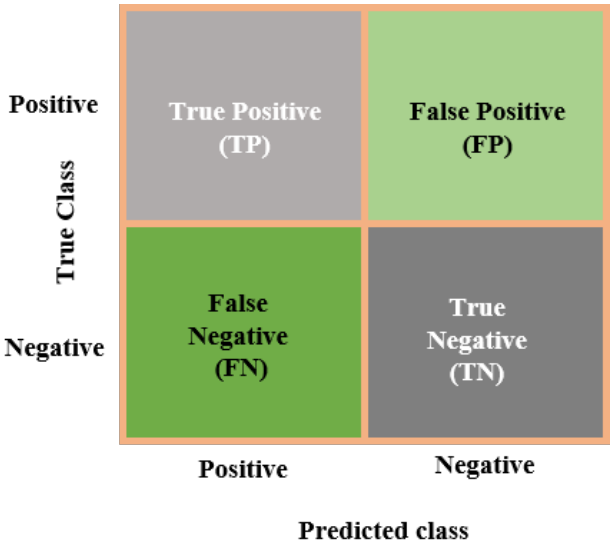

Fig. 5. The confusion matrix.

TP refers to when the algorithm correctly predicts a positive class, while TN indicates when the algorithm correctly predicts a negative class. FP happens when the algorithm predicts a positive class, but the actual class is negative, while FN refers to when the algorithm predicts a negative class, but the actual class is positive. Several performance metrics can be calculated based on these parameters, including accuracy, F1-score, precision, and recall. Accuracy measures the overall correctness of the algorithm's predictions. The algorithm's capability to correctly predict positive classes is measured by precision. The algorithm's capacity to recognize positive classes is measured by the recall, also known as sensitivity. Finally, the F1-score is a harmonic mean of precision and recall, providing a combined metric that balances both measures. The formulas are given as

follows [18], [19]

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F - score = \frac{2(Presicion * Recall)}{Precision + Recall} \tag{5}$$

IV. Experimental Analysis

In this section, we will present the experimental analysis of the impact of data manipulation on the performance of IDS algorithms. We conducted a series of experiments to evaluate the accuracy, prediction speed, and training time of these algorithms when the data they analyze has been manipulated. We will present our findings and analyze the strengths and weaknesses of these algorithms in the context of manipulated data. Evaluating the performances of the algorithms, we selected the best-performing algorithm in each category. Performance is evaluated in terms of the likelihood of a successful detection [20]. The performance of these algorithms in terms of accuracy, considering the manipulated and unmanipulated WUSTL-IIoT-2018 dataset is presented in Fig. 6.

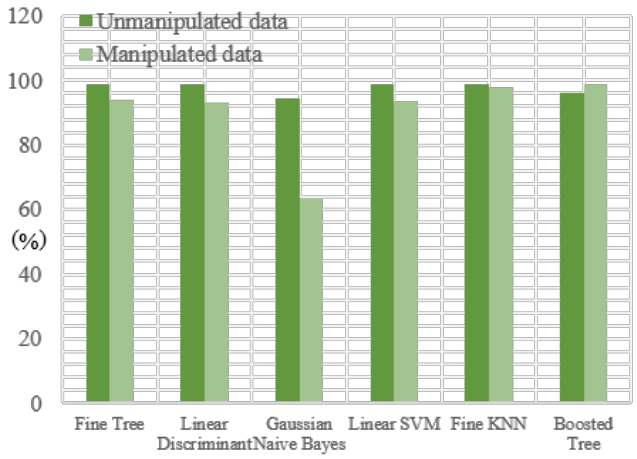

Fig. 6. Performance of the best-performing algorithms on the WUSTL-IIoT-2018 dataset.

The accuracy of the Fine Tree, Linear Discriminant, and Linear SVM all slightly declined with the manipulated data, as shown by the figure in Fig. 6. The Fine KNN showed little degradation while the Gaussian Naive Bayes experienced a significant decline. Surprisingly, the Booted tree's accuracy rose from 96.1% with the unmanipulated data, to 98% with the manipulated data.

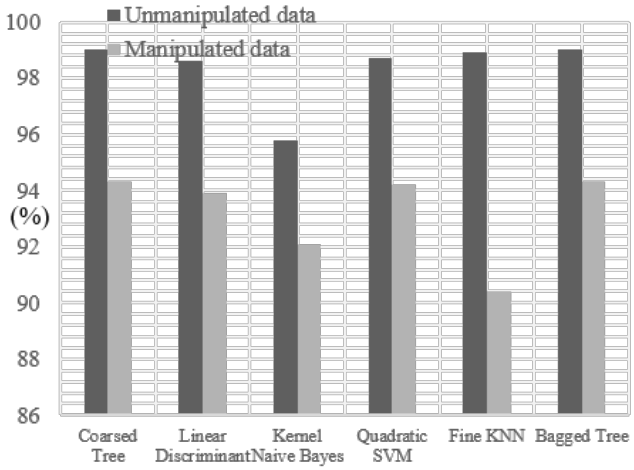

Fig. 7. Performance of the best-performing algorithms on the WUSTL-IIoT-2021 dataset.

The performance analysis of the best-performing algorithms is shown in Fig. 7. The Coarse Tree, Linear Discriminant, Quadratic SVM, and Bagged Tree classifiers experienced a moderate decrease in their performance when tested on manipulated data. On the other hand, the Fine KNN classifier showed a significant drop in accuracy, while the Kernel Naïve Bayes performed poorly on both unmanipulated data and manipulated data.

Algorithms	**Unmanipulated CIC-WUSTL-IIoT-2018**		**Manipulated CIC-WUSTL-IIoT-2018**	
	Prediction speed (obs/sec)	Training time (sec)	Prediction speed (obs/sec)	Training time (sec)
Fine Tree	920000	9.5195	72000	10.952
Linear Discriminant	260000	5.8716	530000	1.6454
Gaussian Naïve Bayes	140000	5.5346	390000	0.9483
Linear SVM	140000	89.756	73000	3266.1
Fine KNN	44000	111.6	49000	307.66
Booted Tree	120000	144.93	33000	340.99

Fig. 8. Prediction Speed and Training Time Comparison on the WUSTL-IIoT-2018 dataset.

The simulation results comparing prediction speed and training time on manipulated versus unmanipulated data are shown in Table 1. The Fine Tree, Linear SVM, and Booted Tree exhibited a decrease in prediction speed on the manipulated data, resulting in an increase in training time. Conversely, the Linear Discriminant and Gaussian Naïve Bayes classifiers demonstrated an increase in prediction speed accompanied by a reduction in training time on the manipulated data. The Fine KNN classifier, on the other hand, demonstrated an increase in both prediction speed and a decrease in training time on the manipulated data.

Algorithms	Unmanipulated CIC- WUSTL-IIoT-2021		Manipulated CIC- WUSTL-IIoT-2021	
	Prediction speed (obs/sec)	Training time (sec)	Prediction speed (obs/sec)	Training time (sec)
Coarse Tree	32000	2.2018	23000	3.6068
Linear Discriminant	120000	4.7506	190000	3.4319
Kernel Naïve Bayes	50	1998.8	41	10265
Quadratic SVM	58000	19.178	49000	81.972
Fine KNN	1100	127.64	1725	169.85
BaggedTree	30000	431.84	23000	431.84

Fig. 9. Prediction Speed and Training Time Comparison on the WUSTL-IIoT-2021 dataset.

The simulation results comparing prediction speed and training time on manipulated versus unmanipulated data (WUSTL-IIoT-2021) are presented in Table II. On the manipulated data, the Coarse Tree, Kernel Naïve Bayes, and Quadratic SVM exhibited a decrease in prediction speed, leading to an increase in training time. Conversely, the Linear Discriminant demonstrated an improvement in prediction speed and a decrease in training time on the manipulated data. In contrast, the Bagged Tree showed a reduction in prediction speed while maintaining a consistent training time across both datasets.

V. Conclusion

This paper exposes the potential risk of power systems to cyber-attacks where a malicious actor could alter the data. In intrusion detection systems, which are commonly used to defend networks against adversarial attacks, machine learning or deep learning techniques are extensively used due to their high accuracy and quick detection rates. The effectiveness of the top machine learning algorithms on manipulated datasets is examined in this paper. The experimental assessment made use of the WUSTL-IIoT-2021 dataset as well as the WUSTL-IIoT-2018 dataset. We evaluate the performance of several machine learning algorithms on both unmanipulated and manipulated datasets. According to experimental findings, except for the Boosted Tree algorithm, all algorithms' accuracy dropped, prediction speed decreased, and training time increased.

References

[1] K. Kumar, and V. Bhatnagar, "Machine Learning Algorithms Performance Evaluation for Intrusion Detection," Journal of Information Technology Management 13, no. 1 (2021): 42-61.

[2] X. Fu, N. Zhou, L. Jiao, H. Li, and J. Zhang, "The robust deep learning–based schemes for intrusion detection in the internet of things environments," Annals of Telecommunications 76, no. 5-6 (2021): 273-285.

[3] E. Degirmenci, I. Ozcelik, and A. Yazici, "Effects of Un targeted Adversarial Attacks on Deep Learning Methods," In 2022 15th International Conference on Information Security and Cryptography (ISCTURKEY), pp. 8-12. IEEE, 2022.

[4] J. A. Abraham, and V. R. Bindu, "Intrusion Detection and Prevention in Networks Using Machine Learning and Deep Learning Approaches: A Review," In 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), pp. 1-4. IEEE, 2021.

[5] S. K. Venkatachary, J. Prasad, and R. Samikannu, "Economic impacts of cyber security in energy sector: A review." International Journal of Energy Economics and Policy 7, no. 5 (2017): 250.

[6] P. A. Oyewole and D. Jayaweera, "Power System Security With Cyber-Physical Power System Operation," in IEEE Access, vol. 8, pp. 179970-179982, 2020, doi: 10.1109/ACCESS.2020.3028222.

[7] D. G. Kyrollos, K. Greenwood, J. Harrold, and J. R. Green, "Detection of false alarms in the NICU using pressure sensitive mat.," In 2021 IEEE Sensors Applications Symposium (SAS), pp. 1-5. IEEE, 2021.

[8] B. Li, B. Zhang, and D. S. Kirschen, "Cyber-Physical Attack Leveraging Subsynchronous Resonance." arXiv preprint arXiv:2207.04149 (2022).

[9] V. D. A. Kumar, "An Effective Comparative Analysis of Data Preprocessing Techniques in Network Intrusion Detection System Using Deep Neural Networks," Smart Intelligent Computing and Communication Technology 38 (2021): 14.

[10] H. He, and Jun Yan, "Cyber-physical attacks and defences in the smart grid: a survey," IET Cyber-Physical Systems: Theory and Applications 1, no. 1 (2016): 13-27.

[11] H. Wang, J. Ruan, B.Zhou, C. Li, Q. Wu, M. Q. Raza, and G. Cao, "Dynamic data injection attack detection of cyber physical power systems with uncertainties," IEEE Transactions on Industrial informatics 15, no. 10 (2019): 5505-5518.

[12] P. Linnartz, A. Winkens, and A. Ulbig, "Assessing the impact of cyber attacks manipulating distributed energy resources on power system operation," arXiv preprint arXiv:2207.07968 (2022).

[13] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, "Machine learning-based network vulnerability analysis of Industrial Internet of Things," in IEEE Internet of Things Journal 6 (2019), pp. 6822-6834.

[14] M. A. Teixeira, T. Salman, M. Zolanvari, R. Jain, N. Meskin, M. Samaka, "SCADA System Testbed for Cybersecurity Research Using Machine Learning Approach," Future Internet 2018, 10, 76,

[15] L. A. C. Ahakonye, C. I. Nwakanma, J. -M. Lee and D. -S. Kim, "Efficient Classification of Enciphered SCADA Network Traffic in Smart Factory Using Decision Tree Algorithm," in IEEE Access, vol. 9, pp. 154892-154901, 2021, doi: 10.1109/ACCESS.2021.3127560.

[16] Y. Chen, C. Fan, and K. Chang, "Manufacturing intelligence for reducing false alarm of defect classification by integrating similarity matching approach in CMOS image sensor manufacturing," Computers and Industrial Engineering 99 (2016): 465-473.

[17] Q. Zhang, X. Chen, Z. Fang, and S. Xia. "False arrhythmia alarm reduction in the intensive care unit using data fusion and machine learning," In 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), pp. 232-235. IEEE, 2016.

[18] M. Qian, J. Luo, Y. Ge, C. Sun, X. Ge, and W. Huang, "Semantic-based false alarm detection approach via machine learning," In 2021 IEEE 21st International Conference on Software Quality, Reliability and Security Companion (QRS-C), pp. 60-66. IEEE, 2021.

[19] B. Taji, A. D. C Chan, and S. Shirmohammadi, "False alarm reduction in atrial fibrillation detection using deep belief networks," IEEE Transactions on Instrumentation and Measurement 67, no. 5 (2017): 1124-1131.

[20] T. Diskin, U. Okun, and A. Wiesel, "Learning to Detect with Constant False Alarm Rate," In 2022 IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC), pp. 1-5. IEEE, 2022.