IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ML-Based Cyber Attack Identification In CP Networks

[1]P. Nagendra [2] P,Chandrakanth

Abstract

It can be challenging to secure the cyber physical systems (CPS) that underpin the Internet of Things (IoT) since security measures taken into account for general information and technology operations (IT/OT) might not function well in a CPS environment. In this paper, we suggested an automatic cyber attack detection and classification framework based on machine learning. For the purpose of ensuring the safety of Internet of Things use cases, it employs a variety of machine learning models. Learning Based - Cyber Attack Detection(LB-CAD) is the approach we suggested. The input of the algorithm consists of an ML pipeline and an IoT integrated use case dataset. Proposed methodology to train classifiers, and subsequently use the generated models to identify and categorize cyberattacks automatically. An empirical investigation is utilized to assess our methods. Performance results revealed that the ML models are capable of detecting and classifying cyber-attacks. In case of binary classification, LSTM(LONG SHORT - TERM MEMORY) showed highest accuracy 78.36%. For multi-class classification K-NN showed highest performance with 82.10% accuracy.

Keywords *– Machine Learning, Cybersecurity, Internet of Things, Binary Classification, Multi-Class Classification*

1. INTRODUCTION

Internet of Things (IoT) devices are gradually existence combined into cyber-physical systems (CPS), including dangerous areas such as dams and power-plants. In these areas, IoT devices (also called Industrial IoT or IIoT) are often part of the Industrial Control System responsible for the reliable operation of the process. ICS can be broadly defined, including Supervisory Control and Data Acquisition (SCADA), Control Systems (DCS), and Programmable Logic Controllers (PLC) and Modbus-based systems. However, the connectivity between ICS or IIoT-based systems and public networks makes them vulnerable to being targeted by cybercriminals. A prime example is the Stuxnet drive in 2010, which targeted Iran's nuclear enrichment centrifuges, causing major effect to the equipment.

Another example is the 2011 pump issue that caused an Illinois water utility to fail. BlackEnergy3 is another Ukrainian nuclear power project that instigated chaos affecting around 230,000 people in 2015 [4]. While security solutions planned for information technology (IT) and operational technology (OT) systems are mature, they may not be directly applicable to ICS. For example, this may be due to tight integration between the physical control centre and the connected systems. Therefore, a steadiness level approach is required to evaluate the body's behaviour and control its working capacity [1].

The security objects of ICS are mainly in terms of availability, integrity, and privacy, unlike most IT/OT systems (which are often critical in determining sharing, integrity, and availability) [5]. Due to the tight connection between the response

control loop and the physical system, a (successful) cyberattack on ICS can have serious consequences and potentially affect people and our environment. This reinforces the importance of developing robust security measures to detect and block access to ICS [1]. While hybrid-based methods are effective in detecting malicious activity, they are unreliable due to frequent network updates and result in different types of detection systems (IDS) [7]. In addition, traditional attack detection and strategies rely solely on network metadata analysis. Our contributions in this paper are as follows.

1. We proposed a ML framework for automatic cyber-attack detection and classification in IoT integrated use case.
2. We proposed an algorithm known as Learning Based - Cyber Attack Detection(LB-CAD).
3. We built a prototype to evaluate our framework and underlying algorithm to know the best performing ML models.

The remainder of the paper is structured as follows. Section 2 reviews literature on existing systems. Section 3 presents the proposed cyber-attack detection methodology. Section 4 presents results of experiments. Section 5 concludes our work and gives directions for future work.

2. RELATED WORK

The Internet of Things can be challenging because security solutions designed for common information/operational technology (IT/OT) do not work well in CPS settings. Therefore, this article presents a Multi-level collaborative effort to find the attack and create a custom design for CPS and more specifically in the business management system (ICS). In the first step, a decision tree combined with a new deep representation learning model is designed for attack detection in a heterogeneous ICS environment. In the second layer, a deep neural network is designed to facilitate the attack. The proposed model was evaluated using real data in water pipes and treatment plants.The Internet of Things (IoT) flashed the Fourth Industrial Revolution (Industry 4.0). This provides great benefits by connecting people, processes and information. However, cybersecurity is emerging as a major challenge for IoT-enabled cyber-physical systems, from connected devices used for business management, to big data generated by products using large IoT. Evolutionary computing will play an important role in cybersecurity along with other smart tools such as antivirus tools for IoT security architectures, data mining/fusion in IoT-enabled cyber-physical systems, and data-driven cybersecurity. This article provides an overview of the security challenges in IoT-enabled cyber-physical systems and the contribution of evolutionary computing and other artificial intelligence to these challenges. The

In the era of standalone systems, security is an important module in environment flexible computing. New types of computing will emerge, such as cognitive heuristics, due to improvements in computing power and connection speed. This approach provides easy and fun human-cantered service anytime, anywhere, on any device. Recently, cyber-physical computing as cyber- physical systems for smart cities, human-computer interaction, smart services and connected devices has been studied worldwide.However, the strategy carefully defines the basis of CPS security.PageRank changes the way you rank and changes the cognitive power of search engines. The proposed system, called SecureCPS, is trained to use real-time data to record interactions on facial recognition websites. The eye area is marked using an illumination algorithm. From the literature [1]- [25], it is understood that there is need for detection

of cyber-attacks in IoT integrated use cases and also classify the attacks in order to take preventive measures.

5. PROPOSED SYSTEM

The search strategy consists of two phases, the representation phase and the detection phase. Using augmented supervised machine learning methods on an unbalanced dataset often yields machine learning models that learn the most class models and miss native class features. “Most researchers try to solve this challenge by creating new models or removing some models to stabilize the data and then importing the data into the machine learning model. However, creating or removing patterns is not a necessary solution in ICS/IIoT security applications. Due to the sensitivity of ICS/IIoT systems, the design has to be checked in the real network, which is impossible as the design can be harmful to the network and have a major impact on the environment or human life. Also, the analysis of the output pattern is time consuming. Also, removing normal data from the dataset is not a solution, as the number of attack patterns in ICS/IIoT datasets is usually less than 10% of the dataset and most of the known dataset is discarded by removing 80% of the dataset.

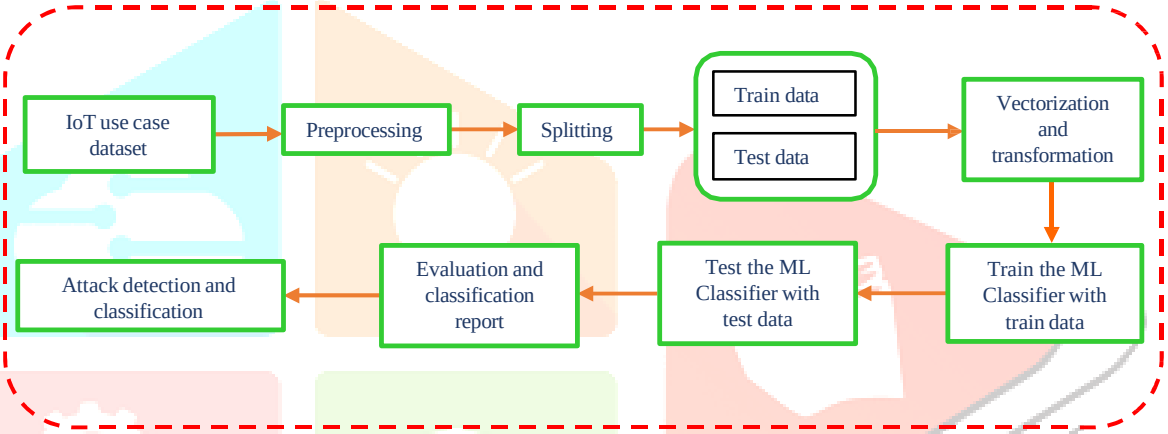

Figure 1: Proposed system for automatic attack detection and classification

To avoid the aforementioned problems with inconsistent data, this work introduces a new machine learning approach that enables models to solve data-consistent text problems without changing, creating or deleting patterns. The model consists of two groups of unsupervised autoencoders, each responsible for finding instances from a class. The model output is a good representation of its input, as each model tries to extract the abstract structure of one class without considering the others.Stacked autoencoders consist of three decoders and encoders with inputs and outputs. Then the ML model will be able to perform binary classification and multi-label classification with better performance.”The proposed process improves the accuracy of results.Although our experiments showed that the performance of the technique was very good, the accuracy of the measurement indicates that it is a difficult task and deserves further investigation.We believe this test can further inform a new set of methods used in different treatments to predict depression and other variables.

Algorithm: Learning Based Cyber Attack Detection and Classification (LBCADC)

Input: IoT use case dataset D, ML pipeline M

Output: Attack detection and classification results R, performance statistics P

1. Begin
2. (T1, T2)←Pre-process(D)
3. For each model m in M
4. m←TrainModel(T1)
5. Save model m
6. (R,P)←TestData(M, T2)
7. Display R
8. End

Algorithm 1: Learning Based - Cyber Attack Detection(LB-CAD)

As presented in Algorithm 1, it takes IoT use case dataset D, ML pipeline M as inputs and produce attack detection and classification results. Then the algorithm proceeds with many machine learning modelswith activities such as model creation, model compilation and model training. Once the model is trained with T1, it is subjected to persisting to reuse in future. In testing phase, the model M is reused to test unlabelled data (T2). Then the algorithm computes result and also performance statistics and display the same. Based on confusion matrix, the evaluation of the proposed algorithm is compared with the state of the art.

4. EXPERIMENTAL RESULTS

A learning machine called DACA-IOT is used to perform the subtraction, custom classification based on machine learning algorithms, and produce the results. It is multi-vector protection that can work in the network and control the operation. The DACA-IOT framework provides tools not only to detect attacks, but also to establish the necessary security conditions for the establishment of the network as an exit mode against the network by the type of attack. This tool has several units that focus on addition, packaging and subtraction, custom classification based on machine learning algorithms, and result generation.

BINARY CLASSIFICATION			
Models	**Precision**	**Recall**	**F1-Score**
Linear Regression	68	96	97
K Nearest Neighbour Classifier	58	97	98
Random Forest	68	98	98
Decision Tree Classifier	66	98	98
LSTM	82	98	98

Table 1: Performance of models in binary classification

As presented in Table 1, the performance of different ML models is evaluated for their ability in binary classification.

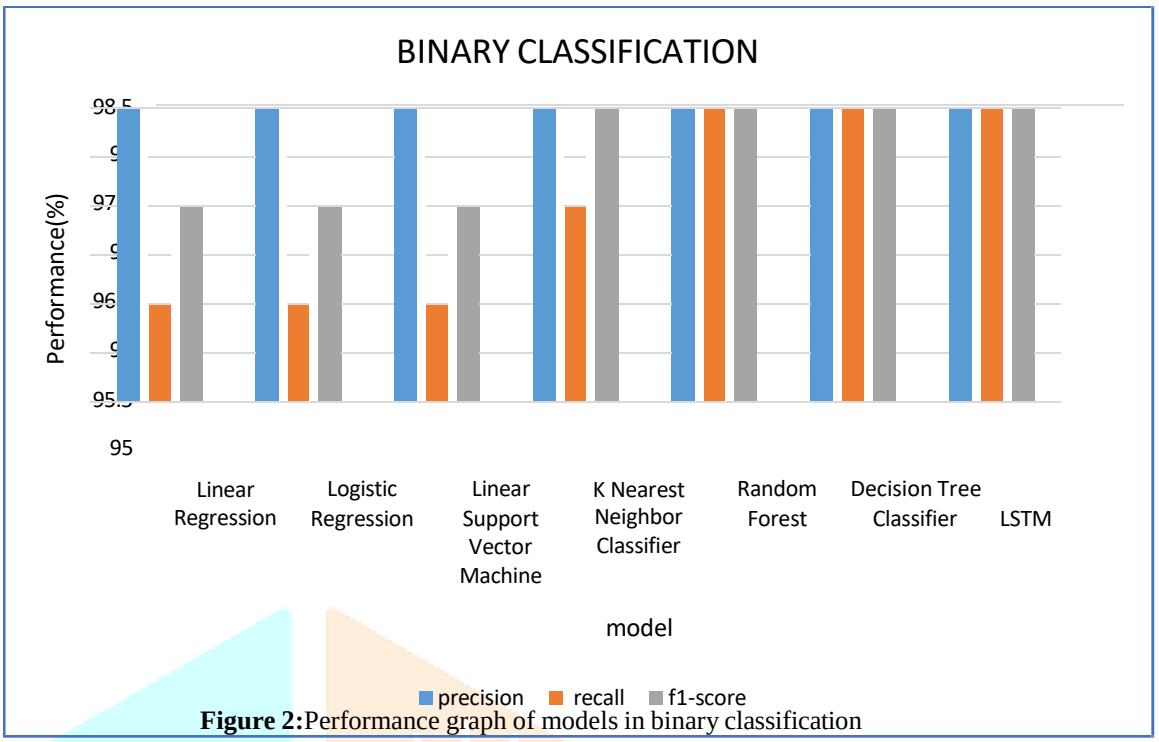

Figure 2: Performance graph of models in binary classification

As presented in Figure 2, binary classification performance is provided for many ML models in terms of precision, recall and F1-Score. With respect to binary classification, LONG SHORT - TERM MEMORY achieved highest accuracy with 98.36%.

MULTI CLASS CLASSIFICATION			
	precision	**recall**	**f1-score**
Linear Regression	0.01	0.01	0.01
K Nearest Neighbour Classifier	78	61	65
Random Forest	67	64	68
Decision Tree Classifier	57	97	97
LSTM	82	98	97

Table 2: Performance of models in Multi Class classification

As presented in Table 2, performance of various ML models is provided when they are evaluated for multi-class classification.

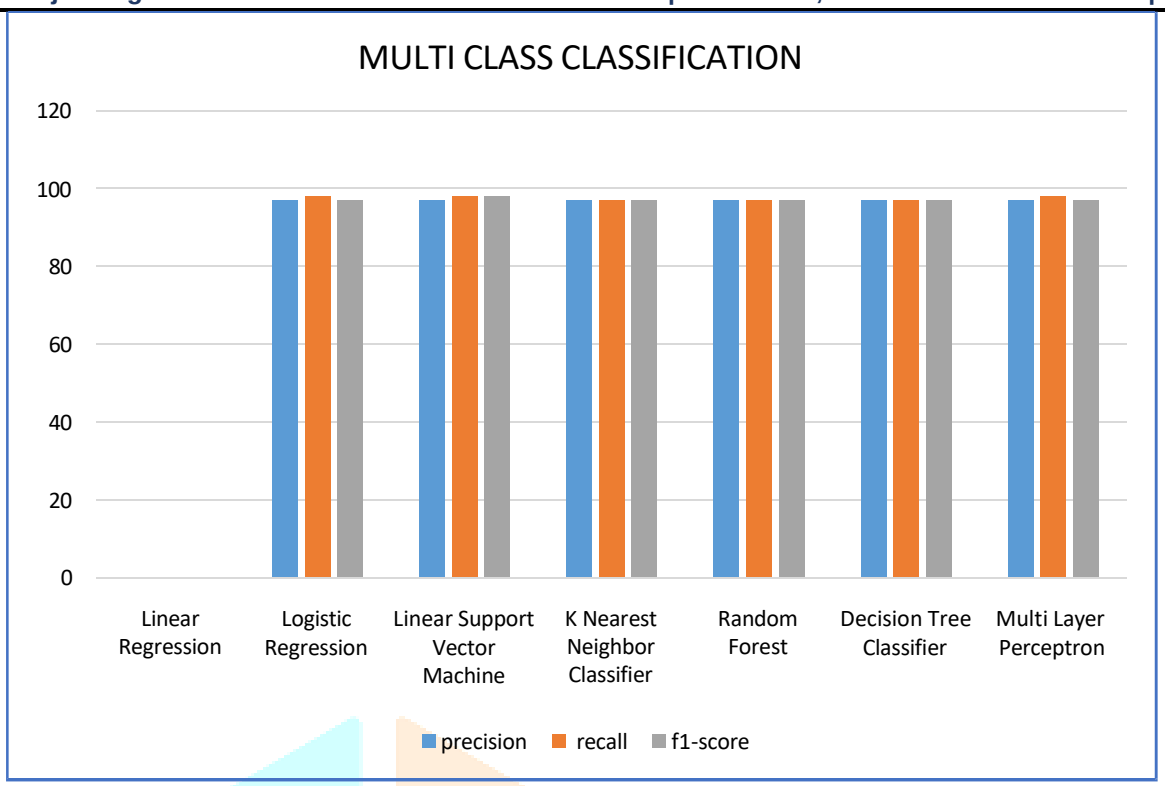

Figure 3: Performance graph of models in Multi Class classification

As presented in Figure 3, multi-class classification performance is provided for many ML models in terms of precision, recall and F1-Score. With respect to multi-class classification, K-NN achieved highest accuracy with 97.59%.

Accuracy Comparison		
	Binary Classification	**Multi-Class Classification**
Linear Regression	74.8	71.3
K Nearest Neighbour Classifier	82.3	64.37
Random Forest	68.64	57.32
Decision Tree Classifier	68.09	97.2
LSTM	78.36	78.54

Table 3: Accuracy comparison of models

As presented in Table 3, performance of all ML models is provided in terms of accuracy comparison which reflects their ability in attack detection and classification.

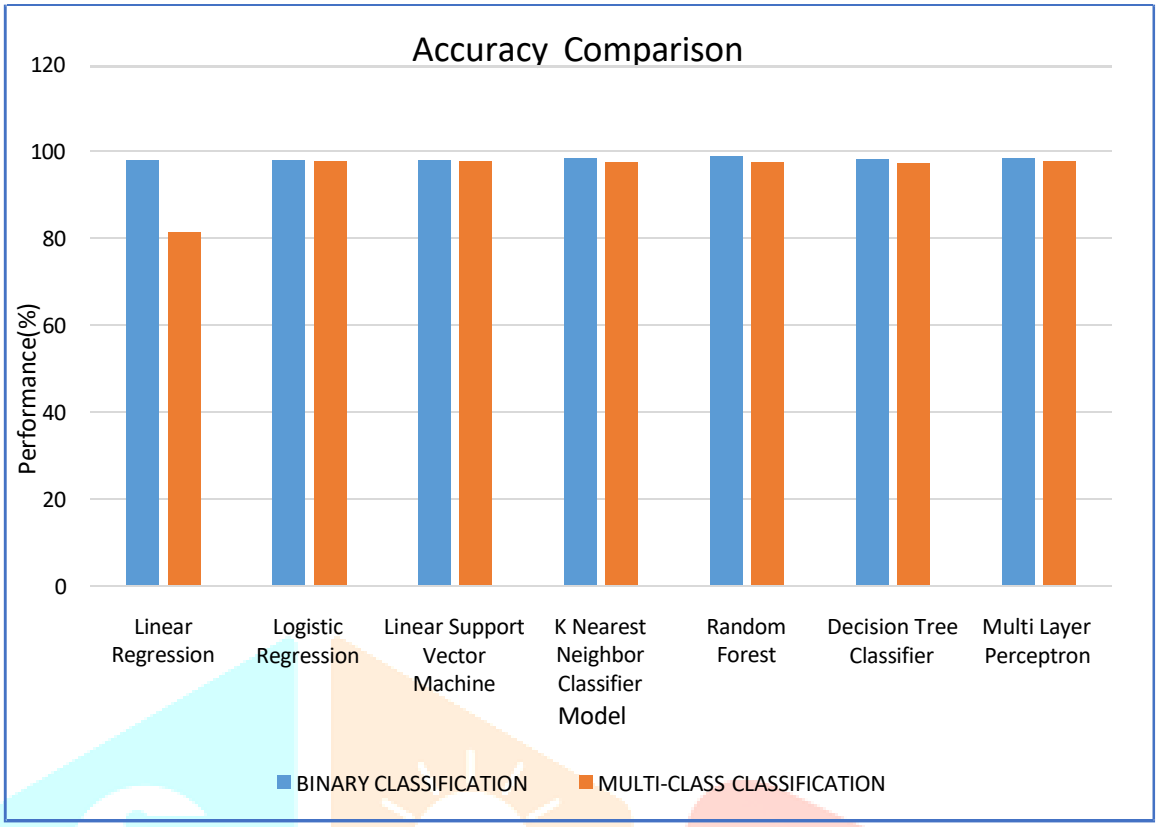

Figure 4: Accuracy Comparison graph of models

As presented in Figure 4, accuracy of various ML models is compared for both binary classification and multi- class classification. Experimental results revealed that the ML models are capable of detecting and classifying cyber-attacks. In case of binary classification, LSTM(LONG SHORT - TERM MEMORY) showed highest accuracy 98.36%. For multi-class classification K-NN showed highest performance with 97.59% accuracy.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a machine learning based framework for automatic detection and classification of cyber-attacks. It makes use of number of ML models to realize cybersecurity of IoT use cases. We proposed an algorithm known as Learning Based - Cyber Attack Detection(LB-CAD). The algorithm takes IoT integrated use case dataset and a ML pipeline as input. It follows a supervised learning approach to train classifiers and then use the resultant models to detect cyber-attacks automatically and also classify them. Our methodology is evaluated with empirical study. Experimental results revealed that the ML models are capable of detecting and classifying cyber-attacks. In case of binary classification, LSTM(LONG SHORT - TERM MEMORY) showed highest accuracy 98.36%. For multi-class classification K-NN showed highest performance with 97.59% accuracy. In future we intend to improve our framework with deep learning methods.

References

[1] F. Zhang, H. A. D. E. Kodituwakku, J. W. Hines, and J. Coble, "Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data," IEEE Transactions on Industrial Informatics, vol. 15, no. 7, pp. 4362–4369, 2019.

[2] R. Ma, P. Cheng, Z. Zhang, W. Liu, Q. Wang, and Q. Wei, "Stealthy Attack Against Redundant Controller Architecture of Industrial CyberPhysical System," IEEE Internet of Things Journal, vol. 6, no. 6, pp. 9783–9793, 2019.

[3] E. Nakashima, "Foreign hacker's targeted U.S. water plant in apparent malicious cyber-attack, expert says." [Online]. Available: https://www.washingtonpost.com/blogs/checkpointwashington/post/foreign- hackers-broke-into-illinois-water-plant-controlsystem-industry-expert-says/2011/11/18/gIQAgmTZYN blog.html

[4] G. Falco, C. Caldera, and H. Shrobe, "IIoT Cybersecurity Risk Modeling for SCADA Systems," IEEE Internet of Things Journal, vol. 5, no. 6, pp. 4486–4495, 2018.

[5] J. Yang, C. Zhou, S. Yang, H. Xu, and B. Hu, "Anomaly Detection Based on Zone Partition for Security Protection of Industrial Cyber-Physical Systems," IEEE Transactions on Industrial Electronics, vol. 65, no. 5, pp. 4257–4267, 2018.

[6] S. Ponomarev and T. Atkison, "Industrial control system network intrusion detection by telemetry analysis," IEEE Transactions on Dependable and Secure Computing, vol. 13, no. 2, pp. 252–260, 2016.

[7] J. F. Clemente, "No cyber security for critical energy infrastructure," Ph.D. dissertation, Naval Postgraduate School, 2018.

[8] C. Bellinger, S. Sharma, and N. Japkowicz, "One-class versus binary classification: Which and when?" in 2012 11th International Conference on Machine Learning and Applications, vol. 2, 2012, pp. 102–106.

[9] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[10] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.

[11] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, "Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things," IEEE Internet of Things Journal, vol. 6, no. 4, pp. 6822–6834, 2019.

[12] I. A. Khan, D. Pi, Z. U. Khan, Y. Hussain, and A. Nawaz, "HML-IDS: A hybrid-multilevel anomaly prediction approach for intrusion detection in SCADA systems," IEEE Access, vol. 7, pp. 89 507–89 521, 2019.

[13] T. K. Das, S. Adepu, and J. Zhou, "Anomaly detection in industrial control systems using logical analysis of data," Computers & Security, vol. 96, p. 101935, 2020.

[14] J. J. Q. Yu, Y. Hou, and V. O. K. Li, "Online False Data Injection Attack Detection With Wavelet Transform and Deep Neural Networks," IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 3271–3280, 2018.

[15] M. M. N. Aboelwafa, K. G. Seddik, M. H. Eldefrawy, Y. Gadallah, and M. Gidlund, "A machine-learning- based technique for false data injection attacks detection in industrial iot," IEEE Internet of Things Journal, vol. 7, no. 9, pp. 8462–8471, 2020.

[16] W. Yan, L. K. Mestha, and M. Abbaszadeh, "Attack detection for securing cyber physical systems," IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8471–8481, 2019.

[17] A. Cook, A. Nicholson, H. Janicke, L. Maglaras, and R. Smith, "Attribution of Cyber Attacks on Industrial Control Systems," EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, vol. 3, no. 7, p. 151158, 2016.

[18] L. Maglaras, M. Ferrag, A. Derhab, M. Mukherjee, H. Janicke, and S. Rallis, "Threats, Countermeasures and Attribution of Cyber Attacks on Critical Infrastructures," ICST Transactions on Security and Safety, vol. 5, no. 16, p. 155856, 2018.

[19] M. Alaeiyan, A. Dehghantanha, T. Dargahi, M. Conti, and S. Parsa, "A Multilabel Fuzzy Relevance Clustering System for Malware Attack Attribution in the Edge Layer of Cyber-Physical Networks," ACM Transactions on Cyber-Physical Systems, vol. 4, no. 3, pp. 1–22, 2020.

[20] U. Noor, Z. Anwar, T. Amjad, and K.-K. R. Choo, "A machine learning-based FinTech cyber threat attribution framework using high-level indicators of compromise," Future Generation Computer Systems, vol. 96, pp. 227–242, 2019.

[21] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," Chemometrics and Intelligent Laboratory Systems, vol. 2, no. 1, pp. 37 – 52, 1987, proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.

[22] A. N. Jahromi, J. Sakhnini, H. Karimpour, and A. Dehghantanha, "A deep unsupervised representation learning approach for effective cyber-physical attack detection and identification on highly imbalanced data," in Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering, ser. CASCON '19. USA: IBM Corp., 2019, p. 14–23.

[23] T. Morris, Z. Thornton, and I. Tunipseed, "Industrial control system simulation and data logging for intrusion detection system research," in 7th Annual Southeastern Cyber Security Summit, 2015.

[24] J. Goh, S. Adepu, K. N. Junejo, and A. Mathur, "A dataset to support research in the design of secure water treatment systems," in Critical Information Infrastructures Security, G. Havarneanu, R. Setola, H. Nassopoulos, and S. Wolthusen, Eds. Cham: Springer International Publishing, 2017, pp. 88–99.

[25] S. N. Shirazi, A. Gouglidis, K. N. Syeda, S. Simpson, A. Mauthe, I. M. Stephanakis, and D. Hutchison, "Evaluation of anomaly detection techniques for scada communication resilience," in 2016 Resilience Week (RWS), 2016, pp. 140–145.