Article

Not peer-reviewed version

Detection of Man-in-the-Middle (MitM) Cyber-Attacks in Oil and Gas Process Control Networks using Machine Learning Algorithms

Ugochukwu Onyekachi Obonna , Felix Kelechi Opara , Christian Chidiebere Mbaocha , Jude-Kennedy Chibuzo Obichere , Isdore Onyema Akwukwegbu , Miriam Mmesoma Amaefule , Cosmas Ifeanyi Nwakanma *

Posted Date: 12 July 2023

doi: 10.20944/preprints202307.0747.v1

Keywords: Amorphous Cyber-attacks; Process Control Network; Anomaly Detection; Machine Learning; Man-in-the-Middle Attacks; SCADA

 doi:10.20944/preprints202307.0747.v1

Article

Detection of Man-in-the-Middle (MitM) Cyber-Attacks in Oil and Gas Process Control Networks Using Machine Learning Algorithms

Ugochukwu Onyekachi Obonna [1], Felix Kelechi Opara [1], Christian Chidiebere Mbaocha[1], Jude-Kennedy Chibuzo Obichere [2], Isdore Onyema Akwukwaegbu [1], Miriam Mmesoma Amaefule [3] and Cosmas Ifeanyi Nwakanma [4,*]

1 Department of Electrical/Electronic Engineering, Federal University of Technology, Owerri, 340110, Nigeria
2 Department of Mechatronics Engineering, Federal University of Technology, Owerri, 340110, Nigeria
3 Department of Mathematics, Federal University of Technology, Owerri, 340110, Nigeria
4 ICT Convergence Research Center, Kumoh National Institute of Technology, Gumi, 39177, South Korea
* Correspondence: cosmas.ifeanyi@kumoh.ac.kr

Abstract: In recent times, the process control network (PCN) of oil and gas installation has been subjected to amorphous cyber-attacks which include Denial-of-Service (DoS), Distributed-Denial-of-Service (DDoS), Man-in-the-Middle (MitM) attacks, and this may have been caused majorly by the integration of open network to Operation Technology (OT) as a result of low-cost network expansion. The connection of the OT to the internet for firmware updates, third-party support, or vendor interventions, has exposed the industry to attacks. The inability to detect these unpredictable cyber-attacks exposes the PCN and a successful attack can lead to devastating effects. This paper reviews the different forms of cyber-attacks in PCN of oil and gas installations and proposes the use of machine learning algorithms to monitor data exchanges between the sensors, controllers, processes, and the final control elements on the network so as to detect anomalies in such data exchanges. Python 3.0 Libraries, Deep-Learning Toolkit, MATLAB, and Allen Bradley RSLogic 5000 PLC Emulator software were used in the simulation of process control. The outcome of the experiments show the reliability and functionality of the different machine learning algorithms in detecting these anomalies with significant precise attack detections identified using a coarse tree algorithm.

Keywords: Amorphous Cyber-attacks; Process Control Network; Anomaly Detection; Machine Learning; Man-in-the-Middle Attacks; SCADA

---

1. Introduction

The Oil and Gas industry is termed critical infrastructure due to the fact that it is a major contributor to the world's energy needs, disruption to its operation could lead to a major impact on the consumers and can lead to devastating effects ranging from catastrophic process safety incidence which may lead to loss of lives, destruction of assets and destruction of the environment, to economic issues to host nations. The choice of standard Information Technology (IT) open systems, their associated communication protocols, and their preference over proprietary dedicated Operational Technology (OT) systems has exposed PCN to insecure communications which have given room to cyber-attacks [1]. The May 2021 Darkside Ransomware attack on the Colonial Pipeline in the USA disrupted and stopped the transportation of gasoline and jet fuel when the computerized equipment managing the pipeline was attacked. After gaining access to the company network of the Colonial Pipeline, Darkside Ransomware was deployed against the company's IT network by intruders [2]. The process variables in the Process Control Network (PCN) serve as inputs to the controllers which make real-time decisions on the final control elements to ensure a continuous and safe operation of the plant. A real-time adjustment or modification of the input variables results in the controller affecting the change in the operating conditions of the logic solvers which eventually results in altered outputs to the final control elements. There is a need to ensure secure communication between the field sensors, the controllers, and the final control elements [3].

Like all other sectors of the economy, continuous digital growth has impacted the Oil and Gas industry. Industrial Control Systems (ICS) are used to operate in isolation, without bridging over Information Technology (IT) infrastructures. Industry 4.0 enabled the integration of multiple industrial technologies in ICT, the engineers can now be able to monitor operations remotely, as well as maintain Supervisory Control and Data Acquisition (SCADA) systems in real-time. This digital revolution has exposed the once air-gapped OT infrastructures to a myriad of new attack surfaces and vectors [4–6]. With the advancement in the Industrial Internet of Things (IIOT), early identification and prevention of attacks that can lead to PCN disasters can be achieved by continuous monitoring using algorithm-based smart monitoring systems [7–9].

ICS operational technology networks can be penetrated by malicious cyber-attackers. Even though there are Intrusion detection systems (IDS), firewalls, demilitarized zones, and data diodes that help in isolating ICS operational technology networks, these security measures cannot be assumed sufficient to stop all malicious penetrations of the air-gapped OT networks. Hackers can access the network through compromised software updates, insider attacks, infected thumb drives, and spear phishing attacks to penetrate heavily isolated and air-gapped OT networks. The Stuxnet malware is a famous example of a worm that penetrated an air-gapped network by exploiting a USB thumb drive autorun vulnerability [10].

Several supervised machine learning algorithms have shown good results in the detection of signature-based attacks which normally are detected by Intrusion detection systems (IDS) but behavior-based attacks which can be termed anomalies or outliers have been difficult to detect or predict based on the dynamic attack strategies deployed by the attackers [11–14]. The choice of the machine learning algorithm to use is influenced by some key factors which include: accuracy, computational capability, prediction speed, false alarm rates, and their application to real-time systems [5,15].

The following objectives are achieved in this research:

1. The process control network ensures effective communication between sensors, controllers, and the final control elements [3]. There is a need to identify and mitigate false data signals that may be introduced as man-in-the-middle (MitM) attacks [16].
2. Disgruntled employees pose a considerable threat to the OT as they can become insider threats with good knowledge of the production facility. Intentional malicious insider attacks usually have a huge impact with a high percentage of success [17].
3. Application of different machine learning algorithms for the detection and prevention of amorphous cyber-attacks on these oil and gas facilities using real-time SCADA dataset.
4. The oil and gas industry in Nigeria is faced with a myriad of challenges ranging from pipeline vandalism, theft, illegal bunkering, and now intrusion attacks [18,19]. This work is focused on the detection and prevention of amorphous cyber-attacks on the networks

The paper is organized as follows: Section I is the introduction, Section II is the review of related works, Section III is the comparison of different machine learning algorithms, Section IV is the results and discussion and Section V is the conclusion and recommendation for future work. Acronyms used in this article are listed in the abbreviations section.

2. Related Works

The integration of standard open network technology has continuously exposed process control networks to malicious cyber-attacks. The need arises to ensure secured communication between the process sensors, the controllers, and the final control elements [3,20]. The connection of the PCN to the internet has also contributed to the growth of cyber-attack incidents with dangerous consequences [21]. The deployment of off-the-shelf IT equipment with its inherent vulnerabilities and associated failures

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2023 doi:10.20944/preprints202307.0747.v1

has also contributed to the exposure of the PCN to cyberattacks [22]. Unstructured and unpredictable attacks are termed outliers to signature-based detections. These nonconforming patterns are termed anomalies, detection of their kind of activities could be done using unsupervised machine learning algorithms [23].

Authors [11] noted that signature-based IDS are disadvantageous as they are unable to detect unknown attacks [11]. The constant dynamic modes of attacks used by the attackers are the major challenge of the work done by Authors [24] used machine learning classifiers as an effective IDS where data was pre-processed to remove unrelated attributes from the dataset [24]. Authors [13] proposed unsupervised machine learning techniques as a solution to unknown attacks including zero-day attacks [13]. Several IDS solutions exist but they cannot detect these un pattern attacks which may be in the form of DoS, DDoS, MitM, or even zero-day attacks [25]. Author [26] reviewed different machine learning capabilities and concluded that the effectiveness and efficiency of a machine learning algorithm-based solution depend on the features and characteristics of the data as well as the performance of the algorithm [26].

Author [16] explained the different forms of MitM which include session hijacking, IP spoofing, and replay attack in which any of the attack forms will lead to the attacker taking over the communication between the sensors and the controllers with the intention of disrupting the process control [27]. Author [28] explained that data trustworthiness, reliability, and availability are necessary for the actualization of cyber-physical systems example smart cities with robust system architecture for secured high bandwidth systems and low-latency diffusion [28]. While supervised machine learning is taught by example and uses labeled data to detect known attacks [29,30], unsupervised machine learning can analyze huge volumes of data to identify hidden patterns, clusters, and outliers, thereby can be very effective in detecting anomalies in datasets which include process upsets, shutdowns or faulty equipment as well as attacks [12,23,31–33]. Deep learning algorithms have shown great results in supervised and unsupervised machine learning applications using very large datasets, timely learning ability, produced great accuracy, and increased prediction speed with negligible false alarm rates [34–36]. Author [36] with the NSL-KDD dataset showed the application of deep learning methods in detecting APT attacks with high detection accuracy.

3. Materials and Methods

3.1. Intrusion Detection Using Machine Learning Models

The study reveals the different forms of unpattern attacks on the PCN with their resultant's effects on the people, assets, and the environment as depicted in Figure 1. The compromise of the intercommunication between the sensors, controllers, and the final control elements could lead to devastating outcomes which may range from fatality to environmental impact. The study reviewed the application of different machine learning algorithms in the modeling of these attacks using the 68,722 real-time SCADA datasets from the oil and gas industry. The performance of the different machine learning algorithms which include: Isolation Forest, k-nearest neighbor (kNN), Python Outlier detection (PyOD) which incorporates Interquartile Range (IQR), kNN, Local Outlier Factor (LOF), Long short-term memory, Support vector machines (SVM) and Decision Tree algorithms were all applied. The 68,722 real-life SCADA data was extracted from an oil and gas facility.

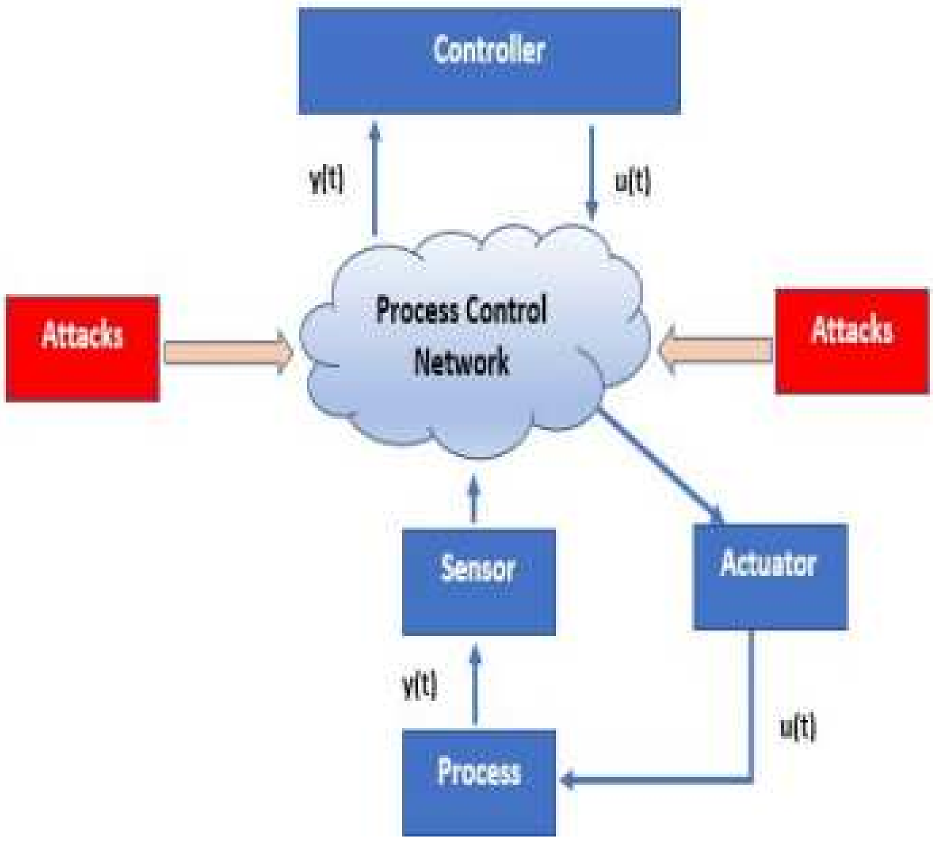


Figure 1. Interconnection of the PCN components under attack

In order to be able to simulate the impact of amorphous cyber-attacks on the oil and gas industry, a 3-phase separator is selected as a case study (see Figure 2). Usually, the natural crude oil flowing from the wellbore which contains entrapped gas and water is fed into a vessel called a three-phase separator. This gravity vessel separates the crude into oil, water, and gas based on their densities [37–40]. In this study, a three-phase separator is used as a case study for ease of computation and simulation to showcase the effect of false data injection in SCADA.

Figure 2 shows a three-phase separator that receives crude oil from the well bore through the shutdown valve and separates the received crude oil into gas, oil, and water. The 3-phase separator has three outlets namely: Gas outlet, Crude oil outlet, and Water outlet respectively. The process variables measured from the vessel include supply pressure, discharge pressure, pressure in the vessel, level of oil with water, level of oil, the temperature of the supplied fluid, vessel temperature, and temperature of the individual discharge lines, while the flow was measured on the respective outlet lines. To prevent process upset and its escalation, there is need for the continuous monitoring of the multivariable inputs with consideration to their interactions in the vessel during the retention time. The 68,722 dataset used in this study simulation, is the 3-phase separator vessel pressure data. The outcome of the simulations using the different machine learning algorithms on the same dataset is documented in the results session. A detailed overview of the system model of this research is shown in Figure 3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2023 doi:10.20944/preprints202307.0747.v1

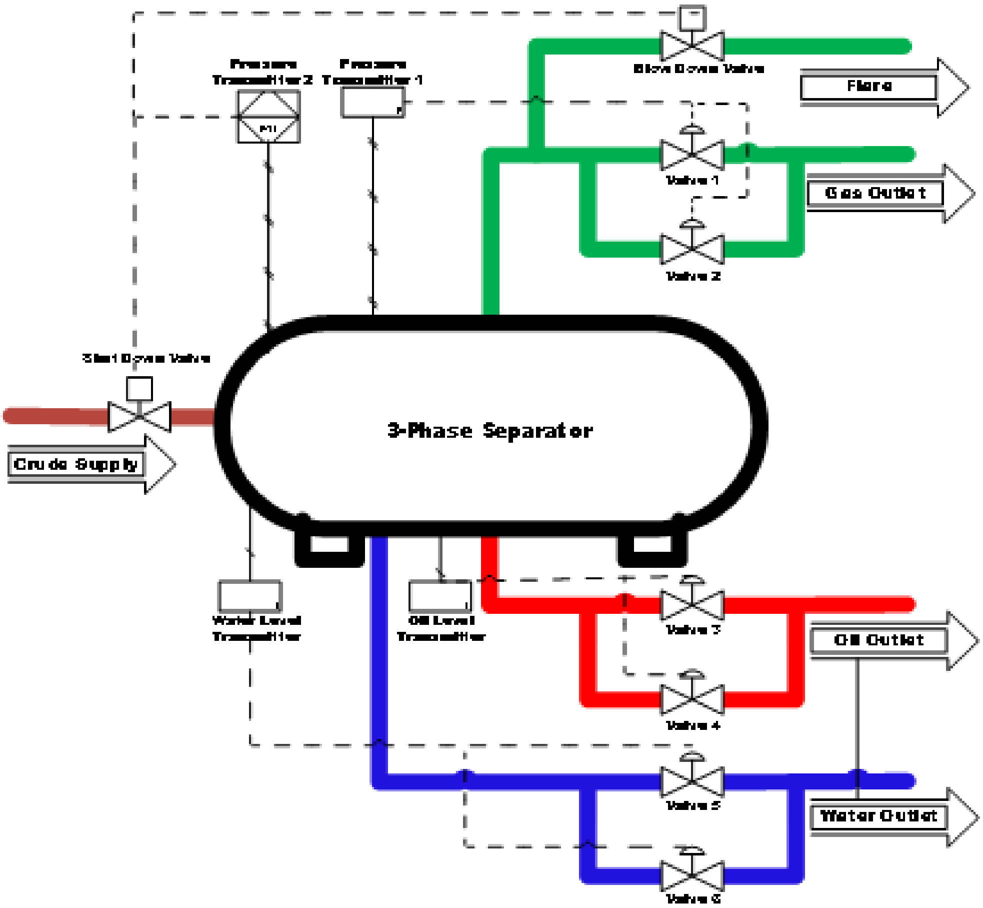


Figure 2. A three-phase separator

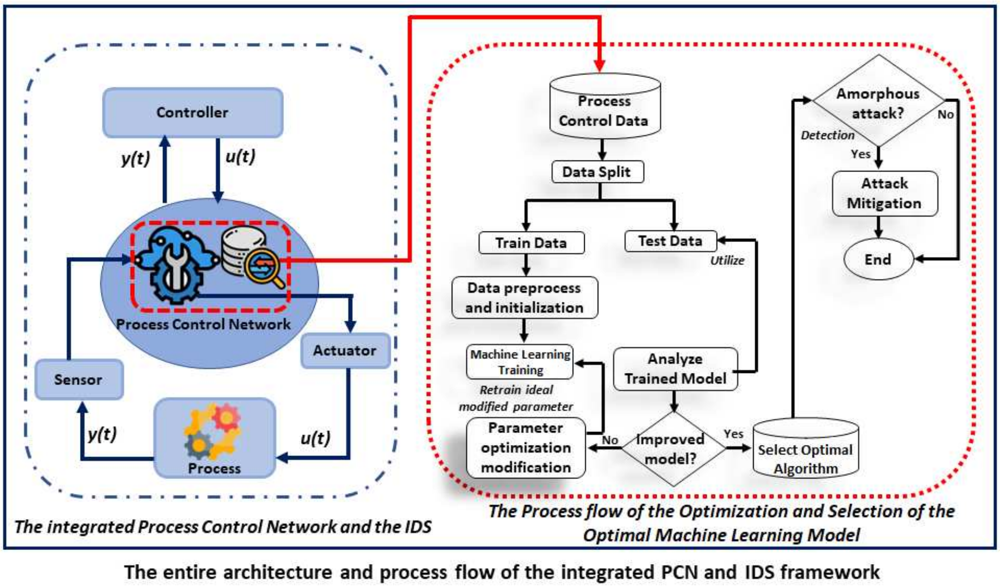
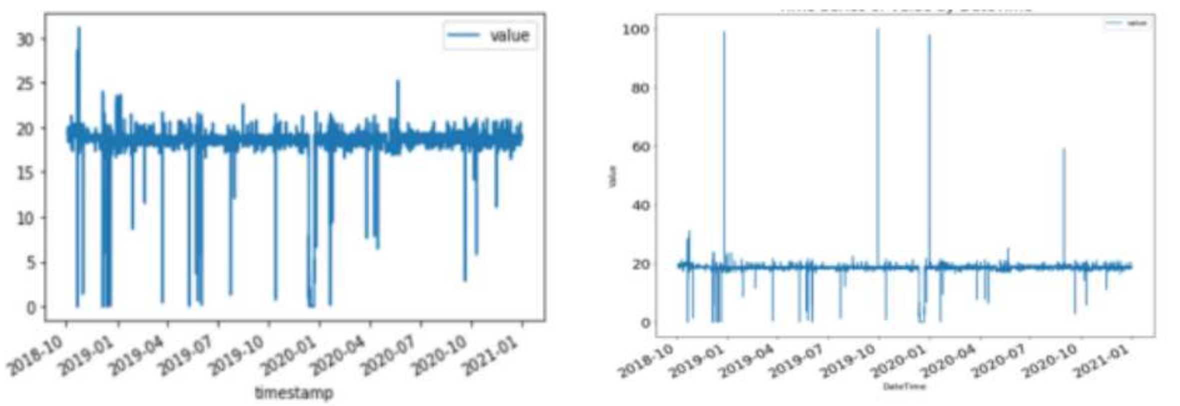


Figure 3. System Model showing the Experimentation and Selection of the Optimimal Machine learning Algorithm

4. Result Discussion and Performance Evaluation

The extracted real-time 68,722 pressure values which is an essential process variable from the SCADA system were plotted against the Date and Time. Pressure is a critical process variable in this process as over-pressurization could lead to explosion and under-pressurization could lead to the implosion of the process vessel, either with catastrophic results which will impact adversely the people, assets, and the environment. The features of the extracted real-time data plotted in Figure 4a show that it does not contain extremely high or extremely low values of pressure for the period under review. For the purpose of simulating the Man-in-the-Middle (MitM) attack, extreme values of pressure were injected into the dataset on specific dates and times. Figure 4b shows the plot of SCADA pressure against the date and time with the anomalies injected.

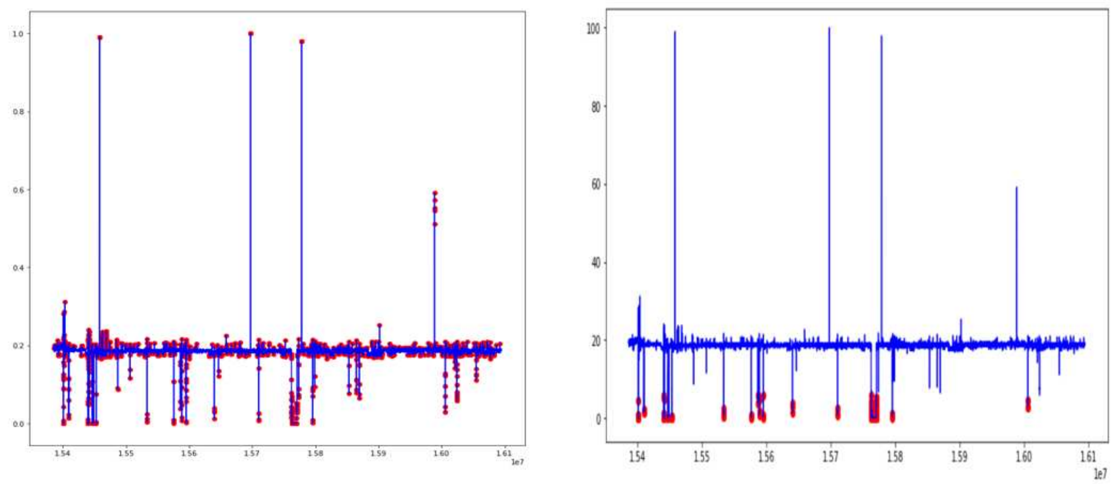


(a) Plot of Pressure (value) against Date and Time with 68,722 raw data samples without anomaly *(y-axis is the pressure while the x-axis is the Date and Time)*

(b) Plot of Pressure (value) against Date and Time, 68,722 data samples with anomalies injected *(y-axis is the pressure while the x-axis is the Date and Time)*

Figure 4. Visualization of the extracted pressure values from the SCADA with and without anomalies.

In Figure 5a, with the contamination parameter set to 0.1, the Isolation Forest Algorithm showed high sensitivity in detecting changes in the pressure values for the period under review including the extreme high-pressure values and detected all as anomalies. This can be termed high False Alarm Rates (FAR). With the contamination parameter set to 0.01, the Isolation Forest was able to detect as anomalies the extreme low-pressure values only with reduced FAR, but it was unable to identify the extremely high anomalies in the dataset and this makes this algorithm for the purpose of real-time detecting MitM attacks as shown in Figure 5b.

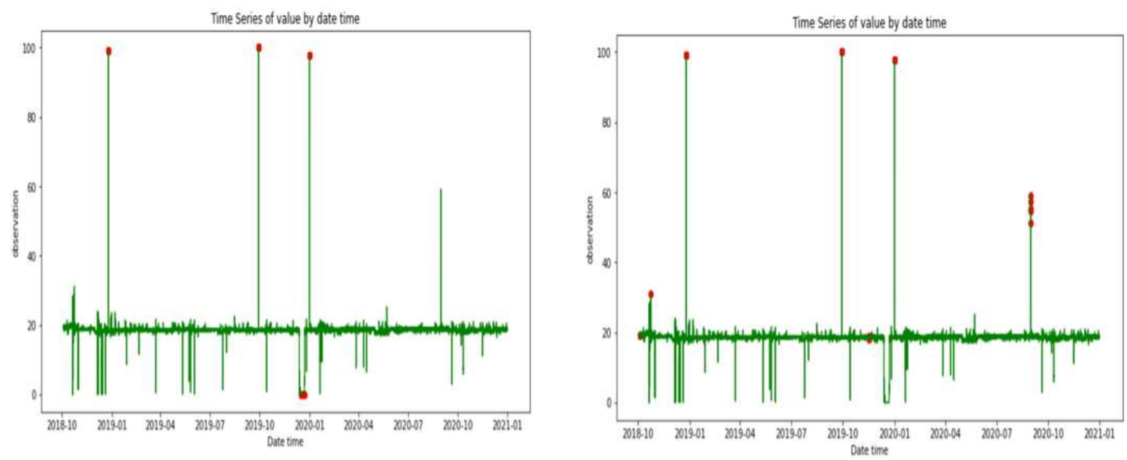


(a) Plot of Isolation Forest Algorithm anomaly detection with 68,722 dataset, contamination parameter set to 0.1 *(y-axis is the pressure while the x-axis is the Date and Time)*

(b) Plot of Isolation Forest Algorithm anomaly detection with 68,722 dataset, contamination parameter set to 0.01 *(y-axis is the pressure while the x-axis is the Date and Time)*

Figure 5. Effect of contamination parameter on he isolation forest algorithm.

In Figure 6a, with step set to 34361, batch size of 32 and 20 epochs, the Long Short-Term Memory (LSTM) algorithm some of the extreme pressure values for the period under review. Changing the batch size to 128 as in Figure 6b, the algorithm detected all the extreme high-pressure values as anomalies though with FAR. The algorithm was unable to identify the extremely low anomalies in the dataset which makes it unreliable for the purpose of real-time detection of MitM attacks.



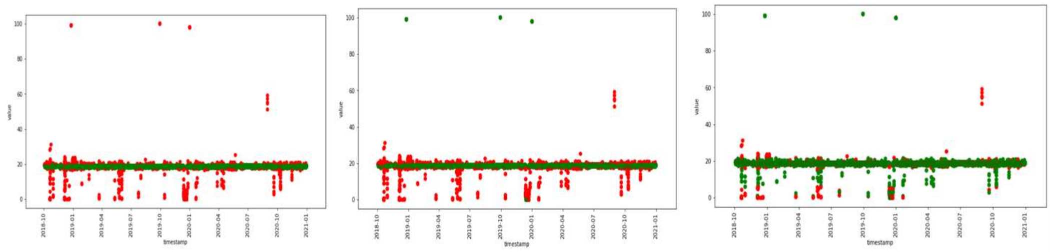
(a) Plot of LSTM Algorithm anomaly detection with a dataset of 68,722, time step of 34361, batch size of 32 and 20 epochs *(y-axis is the pressure while the x-axis is the Date and Time)*

(b) Plot of LSTM Algorithm anomaly detection with a dataset of 68,722, time step of 34361, batch size of 128 and 20 epochs *(y-axis is the pressure while the x-axis is the Date and Time)*

Figure 6. Effect of batch size variation on the LSTM algorithm.

Figure 7a–c show the plot of Python Outlier Detection (PyOD) which incorporates Inter Quartile Range (IQR), k-nearest neighbor (kNN), and Local Outlier Factor (LOF). The results of this algorithm show high sensitivity in detecting pressure value changes by all three algorithms. While IQR was able to detect extreme high-pressure and low-pressure with high FAR, kNN and LOF were unable to detect extreme high-pressure values correctly. Their accuracy is about 70% with high FAR which makes them unsuitable for the detection of MitM attacks.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2023 doi:10.20944/preprints202307.0747.v1



(a) Plot of Inter Quartile Range (IQR) Algorithm anomaly detection with a dataset of 68,722 *(y-axis is the pressure while the x-axis is the Date and Time)*

(b) Plot of k Nearest Neighbor (kNN) Algorithm anomaly detection with a dataset of 68,722 *(y-axis is the pressure while the x-axis is the Date and Time)*

(c) Plot of Local Outlier Factor (LOF) Algorithm anomaly detection with a dataset of 68,722 *(y-axis is the pressure while the x-axis is the Date and Time)*

Figure 7. Plot of Local Outlier factor, KNN performance, and Inter quartile range result

Applying the same 68,722 real-time SCADA pressure dataset to several other machine learning algorithms and comparing their performance metrics which are: accuracy, Receiver Operator Characteristics (ROC), confusion matrix, training time, mis-classification error (MCE) and prediction speed, their outcome is as shown in Table 1. Based on these combined machine learning metrics as shown in Table 1, it was concluded that the coarse tree algorithm has significant performance and can detect MitM attacks effectively with negligible FAR.

Table 1. Behavior of SCADA Pressure Dataset using different Machine Learning algorithms

Algorithm	Accuracy (%)	Training Time (ms)	MCE	Prediction Speed (obs/sec)
Decision Trees				
Fine Tree (FT)	100	1.1708	0	1200000
Medium Tree (MT)	100	1.0781	0	1300000
Coarse Tree (CT)	100	0.45488	0	1000000
Optimizable Tree	100	21.323	0	1300000
Discriminnat Analysis				
Linear Discriminant (LDR)	100	1.843	24	1100000
Quadratic Discriminat (QDR)	99.2	1.1597	518	1600000
Optimizable Discrimiant	100	25.029	24	1600000
Logistic Regression (LR)	100	3.205	N/A	1100000
Naive Bayes				
Gaussian Naive Bayes (GNB)	99.2	1.4947	518	1400000
Kernel Naive Bayes (KNB)	100	65.633	8	4500
Optimizable NB	100	918.96	8	3800
Support Vector Machines (SVM)				
Linear SVM	100	7.3065	25	780000
Quadratic SVM	100	383.79	17	1500000
Cubic SVM	80.2	1657.3	13588	930000
Fine Gaussian SVM	100	7.433	5	610000
Medium Gaussian SVM	100	5.3155	1	760000
Coarse Gaussian SVM	100	5.1452	20	1100000
Optimized SVM	100	7490.9	25	1100000
Nearest Neighbors				
Fine KNN	100	3.6447	0	820000
Medium KNN	100	2.0989	5	460000
Coarse KNN	99.9	3.5228	35	130000
Cosine KNN	99.9	17.422	35	17000
Cubic KNN	100	2.3157	5	380000
Weighted KNN	100	2.1524	0	450000
Ensemble Learning (EL)				
Boosted Trees	99.9	5.0025	35	1200000
Bagged Tree	100	8.5874	0	320000
Subspace Discriminant	100	4.5421	24	260000
Subspace KNN	100	12.777	0	93000
RUSBoosted Tree	100	2.4396	20	960000
Optimized Ensemble	100	232.87	0	530000

 doi:10.20944/preprints202307.0747.v1

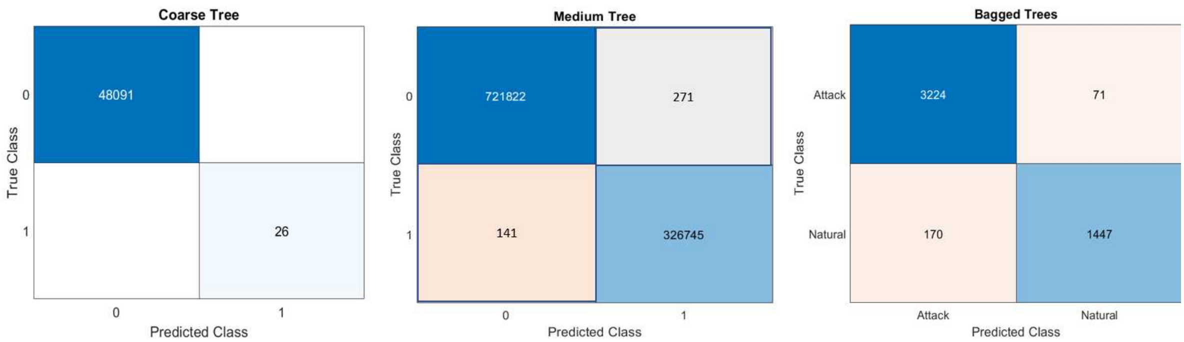
In addition, a thorough comparison was made between the results achieved and that of the other researchers who used WUSTL and ORNL datasets [30,41] in the training of their models. It is important to state that the FAR recorded with the SCADA was zero as compared to other datasets used by other researchers, this is as shown in Table 2.

Table 2. Top and Least Performed Machine Learning Algorithms on various Public Datasets

Datasets /Algorithm	Accuracy (%)	Training Time (ms)	FAR	Prediction Speed (obs/sec)
SCADA Pressure Dataset				
Coarse Tree	100	0.4549	0	1000000
Cubic SVM	80.2	1657.3	13588	930000
WUSTL-SCADA-2018 Datatset				
Medium Tree	100	5.6605	412	4100000
Subspace Discrimiant	93.1	101.64	72009	110000
ORNL POWER GRID Dataset				
Bagged Tree	95.1	4.8021	241	2500
Quadratic Discrimiant	52.4	1.6364	2339	120000

Figure 8a–c shows the plot of Confusion Matrix of the Tree Algorithm with Best Performance using the 68,722 real-time SCADA pressure dataset which shows zero false positives as compared to other WUSTL and ORNL datasets used by other researchers which produced 141 and 170 false positives respectively.

Figure 9a–c shows the plot of Receiver Operator Characteristics (ROC) curve of the best performed Tree Algorithm using the 68,722 real-time SCADA pressure dataset which shows coarse tree produced best result with zero false positives and better Area Under Curve (AUC) while WUSTL and ORNL showed in medium tree and bagged tree respectively with lesser AUC.



(a) Confusion Matrix of the Best Performed Tree Algorithm in SCADA Pressure Dataset

(b) Confusion Matrix of the Best Performed Tree Algorithm in WUSTL-SCADA-2018 Dataset

(c) Confusion Matrix of the Best Performed Tree Algorithm in ORNL (Power Grid) SCADA Dataset

Figure 8. Plot of Confusion Matrix of the Tree Algorithms

 doi:10.20944/preprints202307.0747.v1

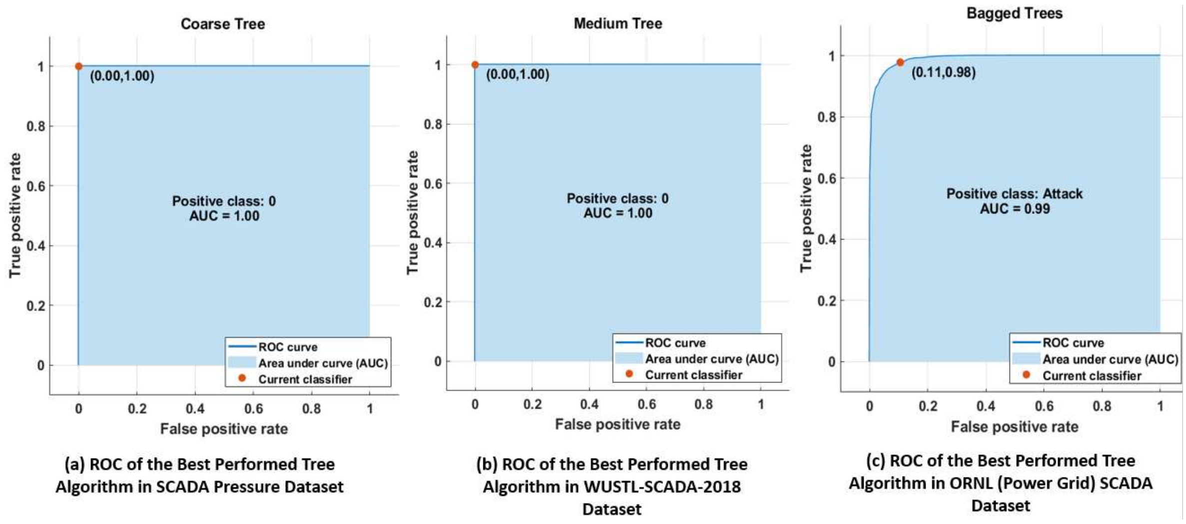


Figure 9. Plot of Receiver Operator Characteristics (ROC) curves

5. Conclusion

The outcome of this study is the evaluation of different machine learning algorithms on the 68,722 SCADA real-time datasets using the following combined machine learning performance metrics: high accuracy, earliest training time, fastest prediction speed, negligible MCE, and less computation power requirement. Based on these combined machine learning performance metrics using the 68,722 datasets, it was concluded that the coarse tree algorithm showed the best performance, and is regarded as the most suitable for the detection of MitM attacks in a process control network of an oil and gas installation. This study can be improved upon by evaluating more machine learning algorithms as well as the use of more real-time SCADA datasets which may go a long way in detecting other forms of cyber-attacks.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, U.O.O, C.I.N., and F.K.O.; methodology, All authors contributed equally; software, U.O.O., C.I.N., and M.M.A.; validation, U.O.O., C.I.N., F.K.O., C.C.M., J.C.O. and I.O.A.; formal analysis, U.O.O, and C.I.N.; investigation, All authors contributed equally; resources, U.O.O. and C.I.N; data curation, U.O.O., C. I. N, and M.M.A.; writing—original draft preparation, all authors contributed; writing—review and editing, U.O.O., and C. I.N.; visualization, U.O.O., C. I. N and M.M.A.; supervision, F.K.O, J.C.O., I.O.A. and C.I.N.; project administration, U.O.O., and C.I.N; funding acquisition, U.O.O. and C.I.N. All authors have read and agreed to the published version of the manuscript.".

Funding: This research work received no external funding

Institutional Review Board Statement: "Not applicable"

Informed Consent Statement: "Not applicable"

Data Availability Statement: "Not applicable".

Acknowledgments: The authors wish to thank Love Allen Chijioke Ahakonye of the Networked Systems Laboratory, IT-Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea for assiting in the simulation during the preliminary stage of the project.

Conflicts of Interest: "The authors declare no conflict of interest." "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results".

Sample Availability: "Not applicable".

Abbreviations

The following abbreviations are used in this manuscript:

APT	Advance Persistent Threats
AUC	Area Under Curve
DDoS	Distributed Denial-of-Service
DoS	Denial-of-Service
FAR	False Alarm Rates
ICS	Industrial Control Systems
ICT	Information and Communication Technology
IDS	Intrusion detection systems
IIoT	Industrial Internet of Things
IoT	Internet of Things
IP	Internet Protocol
IQR	Interquartile Range
IT	Information Technology
kNN	k-Nearest Neighbours
LDR	Linear Discriminant Regression
LOF	Local Outlier Factor
LSTM	Long Short-Term Memory
MATLAB	Matrix Laboratory
MCE	Misclassification error
MitM	Man-in-the-Middle
ORNL	Oak Ridge National Laboratories
OT	Operation Technology
PCN	process control network
PLC	Programmable Logic Controller
PyOD	Python Outlier detection
ROC	Receiver Operator Characteristics
SCADA	Supervisory Control and Data Acquisition
SVM	Support Vector Machines
USB	Universal Serial Bus
WUSTL	Washington University in St. Louis

References

1. Smurthwaite, M.; Bhattacharya, M. Convergence of IT and SCADA: Associated Security Threats and Vulnerabilities. *IOP Conference Series: Materials Science and Engineering* **2020**, *790*, 012041. https://doi.org/10.1088/1757-899X/790/1/012041.
2. CISA.; FBI. DarkSide Ransomware: Best Practices for Preventing Business Disruption from Ransomware Attacks. *Cybersecurity Advisory* **2021**, *July*, 1–1.
3. Irmak, E.; Erkek, İ. An overview of cyber-attack vectors on SCADA systems. 2018 6th International Symposium on Digital Forensic and Security (ISDFS), 2018, pp. 1–5. https://doi.org/10.1109/ISDFS.2018.8355379.
4. Stergiopoulos, G.; Gritzalis, D.A.; Limnaios, E. Cyber-Attacks on the Oil & Gas Sector: A Survey on Incident Assessment and Attack Patterns. *IEEE Access* **2020**, *8*, 128440–128475. https://doi.org/10.1109/ACCESS.2020.3007960.
5. Nwakanma, C.I.; Ahakonye, L.A.C.; Njoku, J.N.; Eze, J.; Kim, D.S. "Effective Industrial Internet of Things Vulnerability Detection Using Machine Learning. Proc. 5th Int. Conf. Inf. Technol. Educ. Dev. Chang. Narrat. Through Build. a Secur. Soc. with Disruptive Technol. ITED 2022, 2022, pp. 1–8. https://doi.org/10.1109/ITED56637.2022.10051622.
6. Ahakonye, L.A.C.; Nwakanma, C.I.; Lee, J.M.; Kim, D.S. Agnostic CH-DT Technique for SCADA Network High-Dimensional Data-Aware Intrusion Detection System. *IEEE Internet of Things J.* **2023**, pp. 1–1. https://doi.org/10.1109/jiot.2023.3237797.

7. Ogu, R.E.; Achumba, I.E.; Okoronkwo, C.D.; Chukwudebe, G.A.; Chukwuchekwa, N. An IoT Solution for Air Quality Monitoring and Hazard Identification for Smart City Development. 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON), 2022, pp. 1–5. https://doi.org/10.1109/NIGERCON54645.2022.9803129.
8. Ogu, R.E.; Chukwudebe, G.A.; Achumba, I.E.; Chukwuchekwa, N.; Ezenugu, I.A. A Robust IoT-based Air Quality Monitoring Node for Multi-Location Deployment. *Int. J. Eng. Res. Technol.* **2022**, *11*, 146–151.
9. Ahakonye, L.A.C.; Nwakanma, C.I.; Lee, J.M.; Kim, D.S. SCADA intrusion detection scheme exploiting the fusion of modified decision tree and Chi-square feature selection. *Internet of Things* **2023**, *21*, 100676. https://doi.org/10.1016/j.iot.2022.100676.
10. Alves, T.; Das, R.; Morris, T. Embedding Encryption and Machine Learning Intrusion Prevention Systems on Programmable Logic Controllers. *IEEE Embedded Systems Letters* **2018**, *10*, 99–102. https://doi.org/10.1109/LES.2018.2823906.
11. Ramotsoela, D.; Hancke, G.; Abu-Mahfouz, A. Attack detection in water distribution systems using machine learning. *Hum. Cent. Comput. Inf. Sci.* **2019**, *9*, 1–22.
12. Zoppi, T.; Ceccarelli, A.; Bondavalli, A. Unsupervised Algorithms to Detect Zero-Day Attacks: Strategy and Application. *IEEE Access* **2021**, *9*, 90603–90615. https://doi.org/10.1109/ACCESS.2021.3090957.
13. Pu, G.; Wang, L.; Shen, J.; Dong, F. A hybrid unsupervised clustering-based anomaly detection method. *Tsinghua Sci. Technol.* **2021**, *26*, 146–153. https://doi.org/10.26599/TST.2019.9010051.
14. Rosa, L.; others. "Intrusion and anomaly detection for the next-generation of industrial automation and control systems. *Futur. Gener. Comput. Syst.* **2021**, *119*, 50–67. https://doi.org/10.1016/j.future.2021.01.033.
15. Ahakonye, L.A.C.; Nwakanma, C.I.; M., L.J.; Kim, D.S. Efficient Classification of Enciphered SCADA Network Traffic in Smart Factory Using Decision Tree Algorithm. *IEEE Access* **2021**, *9*, 154892–154901. https://doi.org/10.1109/ACCESS.2021.3127560.
16. Melnick, J. Top 10 Most Common Types of Cyber Attacks. *netwrix* **2018**, *May 15*.
17. Tang, S.; Liu, Z.; Wang, L. Power System Reliability Analysis Considering External and Insider Attacks on the SCADA System. IEEE Power Eng. Soc. Transm. Distrib. Conf., 2020, pp. 1–4. https://doi.org/10.1109/TD39804.2020.9299922.
18. Hunga, M.O.; Adishi, E. Oil Theft, Illegal Bunkering and Pipeline Vandalism: It's Impact on Nigeria Economy, 2015 -2016. *IIARD Int. J. Econ. Bus. Manag.* **2017**, *3*, 2489–65.
19. Wilson, G. The Nigerian State and Oil Theft in the Niger Delta Region of Nigeria. *J.Sustain. Dev. Africa* **2014**, *16*, 69–81.
20. Mohammed, A.S.; Saxena, N.; Rana, O. Wheels on the Modbus - Attacking ModbusTCP Communications. WiSec '22: Proceedings of the 15th ACM Conference on Security and Privacy in Wireless and Mobile Networks, 2022, pp. 288–289. https://doi.org/10.1145/3507657.3529654.
21. Jang-jaccard, J.; Nepal, S. A survey of emerging threats in cybersecurity. *"Journal of Computer and System Sciences* **2014**, *80*, 973–993. https://doi.org/10.1016/j.jcss.2014.02.005.
22. Amin, S.; Litrico, X.; Sastry, S.; Bayen, A. Cyber security of water SCADA systems- part II: Attack detection using enhanced hydrodynamic models. *IEEE Trans. Control Syst. Technol.* **2013**, *21*, 1679–1693. https://doi.org/10.1109/TCST.2012.2211874.
23. Zoppi, T.; Ceccarelli, A.; Capecchi, T.; Bondavalli, A. Unsupervised Anomaly Detectors to Detect Intrusions in the Current Threat Landscape. *ACM/IMS Trans. Data Sci.* **2021**, *2*, 1–26. https://doi.org/10.1145/3441140.
24. Abrar, I.; Ayub, Z.; Masoodi, F.; Bamhdi, A.M. A Machine Learning Approach for Intrusion Detection System on NSL-KDD Dataset. 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2020, pp. 919–924. https://doi.org/10.1109/ICOSEC49089.2020.9215232.
25. Kulugh, V.E.; Mbanaso, U.M.; Chukwudebe, G.A. Cybersecurity Resilience Maturity Assessment Model for Critical National Information Infrastructure. *SN Comput. Sci.* **2022**, *3*. https://doi.org/10.1007/s42979-022-01108-x.
26. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 1–21. https://doi.org/10.1007/s42979-021- 00592-x.
27. Khraisat, A.; Alazab, A.A. A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges. *Cybersecur.* **2021**, *4*, 1–31. https://doi.org/10.1186/s42400-021-00077-7.

28. Okafor, K.C.; Ndinechi, M.C.; Misra, S. Cyber-physical network architecture for data stream provisioning in complex ecosystems. *Trans. Emerg. Telecommun. Technol.* **2022**, *33*, 1–31. https://doi.org/10.1002/ett.4407.
29. Wakefield, K. A guide to the types of machine learning algorithms:SAS UK. *SAS Institute UK* **2021**.
30. Teixeira, M.A.; Salman, T.; Zolanvari, M.; Jain, R.; Meskin, N.; Samaka, M. SCADA System Testbed for Cybersecurity Research Using Machine Learning Approach. *Future Internet* **2018**, *10*. https://doi.org/10.3390/fi10080076.
31. El Naqa, I.; Murphy, M.J. What Is Machine Learning? *Machine Learning in Radiation Oncology* **2015**, pp. 3–11.
32. Ndubuaku, M.U.; Anjum, A.; Liotta, A. Unsupervised anomaly thresholding from reconstruction errors. *Lect. Notes Comput. Sci.* **2019**, *11874LNCS*, 123–129. https://doi.org/10.1007/978-3-030-34914-1_12.
33. Zoppi, T.; Ceccarelli, A.; Salani, L.; Bondavalli, A. On the educated selection of unsupervised algorithms via attacks and anomaly classes. *Journal of Information Security and Applications* **2020**, *52*. https://doi.org/10.1016/j.jisa.2020.102474.
34. Joloudari, J.H.; Haderbadi, M.; Mashmool, A.; Ghasemigol, M.; Band, S.S.; Mosavi, A. Early detection of the advanced persistent threat attack using performance analysis of deep learning. *IEEE Access* **2020**, *8*, 186125–186137. https://doi.org/10.1109/ACCESS.2020.3029202.
35. Al-Abassi, A.; Karimipour, H.; Dehghantanha, A.; Parizi, R.M. An Ensemble deep learning-based cyber-attack detection in industrial control system. *IEEE Access* **2020**, *8*, 83965–83973. https://doi.org/10.1109/ACCESS.2020.2992249.
36. Bierbrauer, D.A.; Chang, A.; Kritzer, W.; Bastian, N.D. Anomaly Detection in Cybersecurity: Unsupervised, Graph-Based and Supervised Learning Methods in Adversarial Environments. *Cryptogr. Secur. (cs.CR); Artif. Intell. (cs.AI); Mach. Learn.* **2021**. https://doi.org/10.1109/ACCESS.2021.3127560.
37. Song, S.; others. Dynamic Simulator for Three-Phase Gravity Separators in Oil Production Facilities. *ACS Omega* **2023**, *8*, 6078–6089. https://doi.org/10.1021/acsomega.2c08267.
38. Abdu Sabir, B.M.; Elamin, I.H.; Sadiq, H.R. Dynamic Modelling and Simulation of A Three-Phase Gravity Separator. *J. Karary Univ. Eng. Sci.* **2022**, *11*, 1–19. https://doi.org/10.54388/jkues.v2i2.190.
39. Wu, F.; Huang, K.; Li, H.; Huang, C. Analysis and Research on the Automatic Control Systems of Oil-Water Baffles in Horizontal Three-Phase Separators. *Processes* **2022**, *10*. https://doi.org/10.3390/pr10061102.
40. Jonach, T.; Jordan, C.; Haddadi, B.; Harasek, M. Modelling and Simulation of 3-Phase Separators in the Oil and Gas Industry with Emphasis on Water Quality. *Chem. Eng. Trans.* **2022**, *94*, 1009–1014. https://doi.org/10.3303/CET2294168.
41. Morris, T. Industrial Control System (ICS) Cyber Attack Datasets. Center for Cybersecurity Research and Engineering (CCRE), The University of Alabama in Huntsville, https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets, accessed May 29, 2023.