National College of Ireland

Intrusion detection in Industrial OT environment by combination of different machine learning techniques.

MSc Research Project
Cybersecurity

Arindam Ghoshal
Student ID: 20194587

School of Computing
National College of Ireland

Supervisor: Vikas Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Arindam Ghoshal

Student ID: 20194587

Programme: MSC Cyber Security **Year:** January 2021

Module: Research in Computing

Supervisor: Vikas Sahni

Submission Due Date: 07/01/2022

Project Title: Intrusion detection in Industrial OT Environment by combination of different Machine learning techniques.

Word Count: 7238 **Page Count** 21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:

Date:

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Intrusion detection in Industrial OT Environment by combination of different Machine learning techniques.

Arindam Ghoshal
20194587

Abstract

The frequency of attack on industrial systems have taken a sharp rise in recent times, as the traditional control systems have evolved and have incorporated parts of modern-day Information Technology into their architecture. Meanwhile the complexity of industrial system keeps us far from defending them largely from intrusion attacks. Hence more development in the security detection systems need to take place to protect such system from modern day cyber-attacks. Although Intrusion detection system (IDS) are being used these days to secure such environment but not much research has taken place in this field. This research would throw light on whether Intrusion detection system's performance can be enhanced with the help of combining the intrusion detection rate of multiple machine learning algorithms like Random Forest, K-Nearest Neighbour (KNN) and Multilayer perceptron (MLP) for identifying the attack vectors in industrial OT environment. This research produced best result with Random Forest when ran in isolation and slightly better result than Random Forest when combined with the other algorithms.

1 Introduction

Since the 1970s, and particularly with the emergence of the internet in the 1980s, the importance of securing and maintaining private information has grown.

The IT sector has seen a change in cyber threats over the years. Today, as civilization's dependence on computers has grown and technology advanced, attacks have gotten far more sophisticated. Until recently, cyber-attacks were almost entirely contained inside the area of information technology, impacting what we would refer to as "ordinary" PCs.

Since Stuxnet attack in 2010 (Kushner, 2013) the security of industrial control system grabbed global attention and importance of securing the operational technology (OT) environment got highlighted. In initial days the OT environment used to work in silo with respect to IT environment, and it was referred to air gap model, but it has change since industrial revolution 4 took place and it has become more connected with traditional IT and hence has become more vulnerable to cyber-attacks and ever since 2010 cyber-attacks in Industrial Control Systems has only increased. Hence research on IDS for OT has become more important to able to create optimised threat detection systems to defend industries from probable cyber intrusions.

1.1 Motivation

The primary motivation behind choosing this topic of research has been the lack of study on this field in contrast to the traditional IT environment. The gap in research has left OT with lot of vulnerabilities which could lead to security breaches very easily. Although Machine learning had been used recently by multiple researchers to design IDS systems for OT, but not many of them have worked on multiple algorithms to compare the best possible results or combined the different algorithms to find optimised predictability.

1.2 Research Question

- The purpose of this thesis was to conduct a comparison analysis of different machine-learning algorithms in order to evaluate their performance for OT dataset (Borges Hink *et al.*, 2014)[1].
- Secondly, to find out whether an ensemble-based machine learning model which combines the output or predictions of different models, works better in identifying cyber threats than the models consisting of individual algorithms.

1.3 Summary of contents

The research report comprises of sections such as **related research** where the historical studies on IDS for OT by different researchers has been discussed and opportunities for further studies has been highlighted.
Methodology section describes the procedures engaged in the study to reach the result of the research with detailed description being mentioned in the **design detail** section. Python codes and tools used for the study along with the evaluation of the outcome of experiments has been mentioned in the **implementation, evaluation, and discussion** section respectively. Further discussion on future work have been provided in the conclusion section.

2 Related Work

This section deals with review of historical researches which includes various research papers, articles, and conferences for identification of most relevant machine learning algorithms and its implementations which could result into a good detection rate in Intrusion Detection System for Industrial environment using operational technology. Few of them has been mentioned and quoted over the span of the paper.

[1] http://www.ece.uah.edu/~thm0009/icsdatasets/binaryAllNaturalPlusNormalVsAttacks.7z

2.1 Literature Review

(Anton, Sinha and Dieter Schotten, 2019) published their research on detection of attacks in the OT environment with the help of machine learning algorithms such and SVM and Random Forest and got a good amount of accuracy with both algorithms where they were able to establish Random Forest had edge on the output over SDM. Although the results were quite satisfactory but in OT environment the accuracy makes a lot of difference and there was still a scope of improvement with the outcome.

(Beaver, Borges-Hink and Buckner, 2013) had published their work on establishing the use of Machine learning algorithms for network intrusion detection in operational technology field. Various classifying algorithms were used in this study which includes
Naïve Bayes, Random Forests, SVM, J48, NNGe and OneR.
This was a great progress in the OT environment threat detection, but data set was primitive, and algorithms were simple as well, so there are scopes of better datasets and complex algorithms to work on for better output.

(Sawas, Khani and Farag, 2021) described solutions based on deep learning using convolutional neural networks (CNN) where CNN mapped the original one-dimensional data to a two-dimensional matrix representing the CNN input using a feature mapping approach based on Mahala Nobis distance. The technique employed a CNN to classify critical temporal patterns in SCADA data and to identify time windows susceptible to network assaults. The single-dimensional CNN model had an 89 percent detection rate in this implementation on the Swat dataset. Although the accuracy was commendable but still, we have a scope of improvement with the detection accuracy.

(Feng, Li and Chana, 2017) suggested a model comprised of a pair of detectors for detecting anomalies in the gas distribution SCADA system for data injection or denial-of-service (DoS) attacks. The first detector monitored the database's package signature. In SCADA systems, databases are used to store network models and communication model signatures. If the Bloom filter does not contain the package's signature, the package is considered anomalous. The second detector gets the normal package after it has been filtered by the Bloom filter and monitors the following step's activity. This anomaly detection technique, which was used in a SCADA system for a natural gas pipeline, successfully identified assaults 92% of the time. Despite the high accuracy rate of attack detection, the 35-minute training time necessary to train the LSTM is extremely lengthy, which looks to be a downside of the proposed model.

As per research done by (Altunay *et al.*, 2021), unsupervised feature learning, which is a component of deep learning methods, facilitates the discovery of significant features within a large dataset. To learn SCADA data characteristics, architectures including convolutional neural networks, autoencoders, deep belief networks, and LSTM were employed. In the classification process, these structures made use of extreme learning machines, deep belief networks, and multilayer detectors. Based on the analysis of the experiments done, it was concluded that deep learning methods could provide novel approaches to the process of attack detection on SCADA systems. It was concluded that in the future, the effective application of deep learning methods could ensure the safety of industrial control systems. On the contrary although deep learning approaches are very accurate in detecting anomalies, their lengthy training period and selection of data set remains a challenge.

(Sewak, Sahay and Rathore, 2018) have shown in their research that although Deep learning seems more resource longing and complex as compared to standard machine learning approaches, classical machine learning algorithm outperforms Deep learning methods in specific circumstances. This research compared the performance of the classical RF and DNN with 2, 4, and 7-layers architectures and concluded that the classical RF surpasses the DNN in terms of accuracy.

(Ahmed and Hamad, 2021) described an intrusion detection system based on artificial neural network model. High-dimensional Modbus data was trained on a multi-layer perceptron algorithm with binary classification, and then labelled as normal or malicious. They constructed a multi-layer perceptron and binary-based IDS and observed anomaly detection accuracy using a simulated network dataset. It was discovered that the IDS's anomaly detection accuracy was quite high.
But it lacked the ability to identify Denial of Service attacks and future improvement can be the addition of time stamps to the fields to learn the average time packets arrive, which would help to understand the abnormality in packet flow.

This article by (Robles-Durazno *et al.,* 2020) presented a real-time anomaly intrusion detector for a water supply system model. The energy consumption of the components was monitored and recorded during the attacks to create a new training data set and testing. For further testing, the taught machine learning algorithms were created and deployed online throughout the control system's operation. They compared the performance attained by physical and digital training and assessment. The collected findings demonstrated that KNN and SVM could beat rest of the models in terms of accuracy and false-positive and false-negative warnings.

A technique detection of cyber threats on Industrial Control Systems was proposed in the article by (Zhang *et al.,* 2019), who used a multilayer data-driven approach. The suggested intrusion detection system was organized around the notion of defense in depth, and it utilized both supervised model and unsupervised model for intrusion detection, according to the authors. Their experimental setup comprised of a supervisory control and data acquisition system (SCADA) as well as a testbed. The collection comprised of network traffic and host system statistics gathered by the Windows performance monitor, including malicious and legitimate traffic. The malicious traffic comprised of packet sniffing utilizing Man in the middle (MITM) attack, denial of service, data exfiltration, fake data injection and manipulation, and simultaneous cyber-attacks. Taking into consideration their gathered data, the researchers obtained a true positive rate of 98.84 percent for KNN, followed by 98.27 percent for bagging, 97.69 percent for random forest, and 94.80 percent for decision tree.
Even though their technique yielded promising results, they continued to study on the information gathered from the network, and hence their set of assumptions were based on the network rather than OT system dataset.

(Yang, Cheng and Chuah, 2019) published research to define prominent temporal patterns of SCADA traffic and identify time periods where network threats were occurring, the suggested approach used a convolutional neural network (CNN) algorithm. They also devised a re-training strategy to handle previously undiscovered network attack instances, allowing SCADA operators to augment their neural network models with site-specific traces of attacks. The suggested approach proved well in attaining high detection accuracy and giving the potential to manage zero-day threats, according to the results utilizing genuine SCADA traffic data sets.

But in this experiment the only protocol used for OT environment was DNP3 and secondly the attack type was more network specific, hence improvement could be done with more protocol types in data set and taking mixed bag of attack vectors.

(Wang *et al.*, 2021) proposed a model based on a deep residual Convolution Neural Network (CNN) which could avoid gradient explosion or disappearance and assure more accuracy. This research could avoid the disadvantages of traditional machine learning algorithms which makes IDS less capable of detecting Zero-day attacks and long training hours of data sets for deep learning methodologies. The use of transfer learning in conjunction with a modification of the residual CNN structure in this research ensured the detection of unknown attacks. This paper revealed that CNN can be used with other anomaly detection algorithm to give better results. Secondly the data set used here is KDD CUP which is mostly ICT network flow data, and it could be replaced by OT dataset for better results.

(Khan *et al.*, 2019) used the Bloom filter and KNN in an automated multi-level intrusion detection system.
Upon pre-treatment and dimension reduction of the observed data, the first level Bloom filter was used to develop a set of authentic network packet signature. A KNN-based algorithm was engaged at the second level to identify possible zero-day attacks. Additionally, the suggested technique recognized that the accuracy of classifiers could be enhanced by balancing the dataset. The recovered features, which were the outcome of the DFR-advised technique, in conjunction with the Bloom filter and KNN, confirmed the reliability of the insights between the system's regular and irregular behavior.
In general, the suggested IDS could achieve an adequate degree of efficiency while keeping the cost of computation low.
To enhance the DR and efficiency of the suggested methodology, deep learning approaches and more datasets are recommended.

To help with the implementation of futuristic SCADA systems, the authors (da Silva *et al.*, 2016) of this study described the advantages of adopting Software-Defined Networking (SDN). A network-based intrusion detection system (NIDS) for SDN-based SCADA systems was also presented. This system leveraged SDN to record network information and monitored communication between power grid components. One-Class Classification (OCC) algorithms analyzed network device information collected on a regular basis using SDN. OCC methods have the benefit of not relying on known attack signatures to identify malicious traffic, which is important since attack traces in SCADA networks are rare and not publicly released by utilities. This study's findings showed that OCC algorithms are very accurate, with an estimated 98 percent detection rate for SCADA-targeted cyber-attacks.

(Almalawi *et al.*, 2016) published article which described a new intrusion detection technique for detecting SCADA-specific threats. This was accomplished by the use of a data-driven clustering approach for process parameters, that intuitively determined a system's regular and vulnerable conditions. It then derived proximity-based detection rules for monitoring purposes from the indicated states. The efficacy of the suggested technique was evaluated via tests on eight data sets including the values of process parameters. An average accuracy around 98 percent was achieved in recognizing critical situations in this approach, hence easing SCADA system monitoring.
The main drawback of this experiment was, the detection capability lagged the traditional machine learning approach and secondly the data sets comprised of mix of real life and test

datasets, hence we could not compare with other experiments carried out with different machine learning algorithms.

(Derhab *et al.*, 2019) discussed the security of commands in industrial IoT against forgery and command misrouting. The proposed security architecture consisted of two components: (1) an IDS using RSL-KNN, that integrated Random Subspace Learning (RSL) and K-Nearest Neighbor (KNN) for detecting and securing against forged commands directed at industrial control processes; and (2) a Blockchain-based Integrity Checking System (BICS) that could protect from threats which impacts the OpenFlow policies of software defined network enabled industrial control systems.
Overall, the suggested security solution performed well. Overall accuracy stood up to 96.73% with BICS detecting fraudulent flow up to 100%.
However, the detection accuracy was found to be lesser compared to standard machine learning algorithms and deep learning algorithms. Further authors have expressed their future work would include other deep learning algorithms and more datasets to improve the accuracy.

(Lee, Kim and Jung, 2008) proposed an intrusion detection system architecture that detected intrusions in various stages. The proposed system recognized the distinctive intrusion signals associated with each intrusion step that comprise of an intrusion method and calculated the total incursion based on their sequence. This system had a low chance of mistake, and it also reduced system overhead by delegating intrusion signal detection to autonomous agents at each level. This article used the Hidden Markov Model algorithm to identify intrusions using the misuse detection approach, which has the disadvantage of being unable to detect new or updated intrusion strategies.

2.2 Summary of Historical Research work

The table below consolidates the summary of the literature review of the related research works:

S. No	Authors	IDS system for Operational Technology (mostly between 2015-2021)	Review comments	Accuracy and other evaluation parameters
1	S. D. D. Anton, S. Sinha and H. Dieter Schotten (Anton, Sinha and Dieter Schotten, 2019)	Anomaly-based Intrusion Detection in Industrial Data with SVM and Random Forests	Here Random Forest algorithm showed improvement over SVM but there is scope of improvement in accuracy.	90% to 92% for SVM and 99% in case of RF
2	Beaver, J.M., Borges-Hink, R.C. and Buckner, M. A (Beaver, Borges-Hink and Buckner, 2013)	An evaluation of machine learning methods to detect malicious SCADA communications	Random Forest outperformed other supervised classifiers but combination of the same could be tested for better accuracy.	Evaluated on Precision and F1 score and was found to be more that 75% for most of the algorithms.
3	D. Lee, D. Kim and J. Jung (Lee, Kim and Jung, 2018)	Multi-Stage Intrusion Detection System Using Hidden Markov Model Algorithm	The method used HMM model which is unable to detect newer intrusion attacks.	Detection rate between 95% to ((% over different number of Hosts.

4	M. Kravchik and A. Shabtai, A (Ahmed and Hamad, 2021)	Detecting Cyber-Attacks in Industrial Control Systems Using Convolutional Neural Networks	CNN produced a moderate result, but opportunity of improvement was there and combination with other models could be tried.	Varied for 62 % to 97% accuracy for different algorithms.
5	C. Feng, T. Li and D. Chana (Feng, Li and Chana, 2017)	Multi-Level Anomaly Detection in Industrial Control Systems via Package Signatures and LSTM Network	Long-Short-Term Memory provided good results, but training time is lengthy, which is a serious concern for any IDS.	Accuracy of 92% was achieved.
6	H. C. Altunay, Z. Albayrak, A. N. Özalp and M. Çakmak (Altunay *et al.*, 2021)	Analysis of Anomaly Detection Approaches Performed Through Deep Learning Methods in SCADA Systems	Long-Short-Term Memory provided good results, but training time is lengthy, which is a serious concern for any IDS.	92% accuracy was achieved with CNN and LSTM model.
7	J. Vávra and M. Hromada (Vávra and Hromada, 2017)	Evaluation of anomaly detection based on classification in relation to SCADA	Supervised algorithms were chosen over unsupervised algorithms.	True Positive rate varied from 78% to 96%. With ROC with RF giving the best results in terms of TP and precision.
8	M. Sewak, S. K. Sahay and H. Rathore (Sewak, Sahay and Rathore, 2018)	Comparison of Deep Learning and the Classical Machine Learning Algorithm for the Malware Detection	Random Forest superseded DNN in this experiment. Scope of other unsupervised algorithm to be tried to find different result.	Accuracy of ((.7% was achieved with RF where DNN could reach 97.7% at most without any feature selection.
9	A. Hijazi, E. A. El Safadi, and J.-M. Flaus (Ahmed and Hamad, 2021)	A deep learning approach for intrusion detection system in industry network	ANN-based intrusion detection system lacked the ability to identify all attack types, such as DoS. Dataset used is related to IT network.	Accuracy achieved was 99.89%
10	A. Robles-Durazno, N. Moradpoor, J. McWhinnie and G. Russell (Robles-Durazno *et al.*, 2020)	Real-time anomaly intrusion detection for a clean water supply system, utilising machine learning with novel energy-based features	KNN and SVM were having the best output in terms of accuracy. Combination of unsupervised algorithm could be checked.	Accuracy of 99.3% and 97.9 % were achived by KNN and SVM respectlively.
11	F. Zhang, H. A. D. E. Kodituwakku, J. W. Hines, J. B. Coble, W. Hines (Zhang *et al.*, 2019)	Multi-Layer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network System and Process Data	Even though their technique yielded promising results, the assumptions were based on the network rather than OT system dataset.	Highest detection rate achieved was 98.84% with KNN.
13	H. Yang, L. Cheng and M. C. Chuah (Yang, Cheng and Chuah, 2019)	Deep-Learning-Based Network Intrusion Detection for SCADA Systems	CNN was used and favorable results were obtained in detecting zero-day attacks and data was more relevant to TCP/IP prototocol rather than OT specific protocols.	Accuracy achieved was 99.84%

14	W. Wang et al. (Wang *et al.*, 2021)	Anomaly detection of industrial control systems based on transfer learning	CNN was used and favorable results were obtained but the dataset used was KDDCUP99 which is a general IT network cyber threat test dataset.	Accuracy and Precision over 99% were achieved,
15	I. A. Khan, D. Pi, Z. U. Khan, Y. Hussain and A. Nawaz (Khan *et al.*, 2019)	A Hybrid-Multilevel Anomaly Prediction Approach for Intrusion Detection in SCADA Systems	KNN is used in this research. Although the computational cost is low, but detection rate is low as well and hence we need to look for more alternatives to this algorithm.	Accuracy of 97% was achieved.
16	E. G. da Silva, A. S. d. Silva, J. A. Wickboldt, P. Smith, L. Z. Granville and A. Schaeffer-Filho (da Silva *et al.*, 2016)	A One-Class NIDS for SDN-Based SCADA Systems	One-Class Classification (OCC) algorithms and detection rate was favorable, although research focuses on SDN based NIDS only.	Accuracy of 98% was achieved.
17	A. Almalawi, A. Fahad, Z. Tari, A. Alamri, R. AlGhamdi and A. Y. Zomaya (Almalawi *et al.*, 2016)	An Efficient Data-Driven Clustering Technique to Detect Attacks in SCADA Systems	Data Driven Clustering was done in this research. The data set was a mixed bag of different real time and test data; hence it is hard to compare with another research to verify its stand.	Accuracy of 98% was achived.
19	Derhab, Abdelouahid, Mohamed Guerroumi, Abdu Gumaei, Leandros Maglaras, Mohamed Amine Ferrag, Mithun Mukherjee, and Farrukh Aslam Khan (Derhab *et al.*, 2019)	Blockchain and random subspace learning-based IDS for SDN-enabled industrial IoT security	RSL-KNN model was used for the research, further deep learning models could be incorporated to verify whether they work better on the same dataset.	Accuracy of 96.73 % was achieved .for binary and 91.07 % was achieved for multiclass datasets.

Table 1: Summary of research literature review.

2.3 Literature Review Gap

As per the reviewed research works mentioned above, it is evident that there had been lesser amount of research work pursued in the field of threat detection in industrial control system (ICS) area. Aside to that the availability of operational technology data set had been challenge for many researchers and hence studies had been conducted over more generic data sets like KDD'99 consisting of information technology network intrusion cases (W. Wang *et al.*, 2021). Looking into comparative study conducted over research works by (Rakas *et al.*, 2021) it was found that very few research were specifically performed over SCADA or other ICS and not much experimentation has been performed over designs of IDS for industrial control system environment.

Further most of the research conducted with Machine Learning methods used single layered model with few exceptions and practically very few of them deployed the method of combining the IDS output of different ML models to verify if that helps to boost the overall detection rate of the IDS.

Hence this research was focused on developing an IDS for Industrial control System using combination of multiple IDS models comprising of ML algorithms like Random Forest, K-Nearest Neighbor and multilayer perceptron (MLP) and also called NLP natural language processing. Secondly, this study was conducted using specific data set related to industrial control system (SCADA) obtained from Mississippi State University and Oak Ridge National Laboratory (Beaver *et al*. 2014)

3 Research Methodology

The research project is based on development of an IDS solution for Operational Technology environment by combination of different machine learning models to form a hybrid IDS system. The purpose of this report is to enhance the accuracy of the system and thus, a combination of the outputs of different machine learning techniques has been used that involves MLP, K-NN as well as Random Forest (Tama, Comuzzi and Rhee, 2019). Datasets have been selected that contain data on cyber-attacks and those datasets have been used to detect alerts and vulnerabilities (Huang and Lei, 2020).

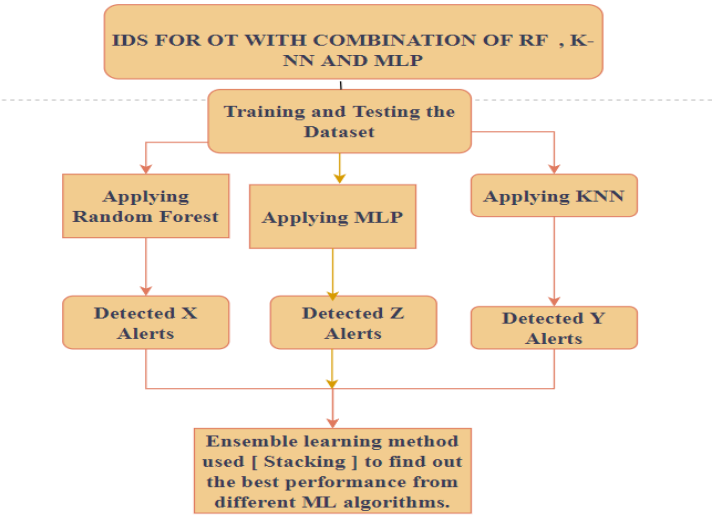

Figure 1: Flow Chart of Process Implementation using Random Forest and CNN

3.1 Data collection

Data has been collected from various datasets based on simulated cyber-attacks and those datasets have been used as this report is based on detecting defects and increasing the accuracy of the IDS system. The data sets used in this report have been obtained from public data repository which has been used for various research conducted on power system cyber-attack (Borges Hink *et al.*, 2014).

3.2 Data pre-processing

The data set obtained was analyzed to ascertain the distribution specific to the research topic.

In this step the various classes of data available were mapped to the identifiable known intrusion classes. This process was aimed at balancing the data set to avoid any classification issues which might impact the outcome of the experiment. Method which was used was over-sampling of the normal observations to balance the data set to avoid too many instances of the malicious cases.

Further the data set has been cleaned by removing the null values from the dataset or replacing them with closest fits to enhance the efficiency of the model. Presence of null or invalid entries data set might degrade the accuracy and predictability of the model drastically.
Lastly, the string values were removed from the data set to make it more interpretable for the machine learning algorithms. Hence the output variable of the data set was converted from string to numerical binary values 0 and 1.

3.3 Data Training

Separation of data was done into two sets, namely training set and testing set, as this is important for evidence-based analysis of data mining models. The data was partitioned into 70-30 ratio for training and testing purpose. Usage of comparable data in testing and training sets ensured lesser discrepancies in data and better efficiency with the experiments. The primary purpose of such partition was to have a reference dataset to validate the output of the data processing, where training data is processed, and output is compared with the known values in the testing data set to determine with the accuracy of predictions.

3.4 Classification Algorithms

This research used four of the machine learning algorithms which are mentioned below. The

- **Random Forest:** Supervised machine learning classifier.
- **K-NN:** Supervised machine learning classifier.
- **Multilayer perceptron (MLP):** un-supervised Deep learning algorithm.
- **Stacked Ensemble Learning:** Hybrid learning method where the prediction of other models are used as inputs for this model.

3.5 Model Evaluation

The model evaluation metrics considered for the research are accuracy, area under the curve, precision, and sensitivity.

Accuracy defines the detection percentage of True positive and true negative over the total number of samples.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Precision defines the percentage of true positive value over the summation of true positive and false positive values.

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

Sensitivity defines the percentage of true positive over true positive and false negative.

$$\text{Sensitivity} = \frac{TP}{(TP + FN)}$$

4 Design Specification

The technique used in this research was based out of stacking of multiple machine learning models to provide a better result. The research uses the Hybrid or ensemble-based machine learning algorithms. Although ensemble consists of different types, bagging is one of the most popular methods used worldwide for studying different datasets and here the similar method has been used to combine the output of multiple classifiers to detect different attack types from the ICS data set consisting of intrusion and normal events. Although a python code has been used here to find the attacks detected and compute the union of the three output data sets obtained as a result from three models.
The ML algorithms used here are Random Forest, K-NN and MLP, which were ran in isolation on the same data set and then finally the detected attacks from each of the experiments done in isolation were combined to produce the final detected attacks set. The final set of detected attacks was then evaluated with respect to the entire input data set to verify the detection rate of the ensemble or combined IDS model.

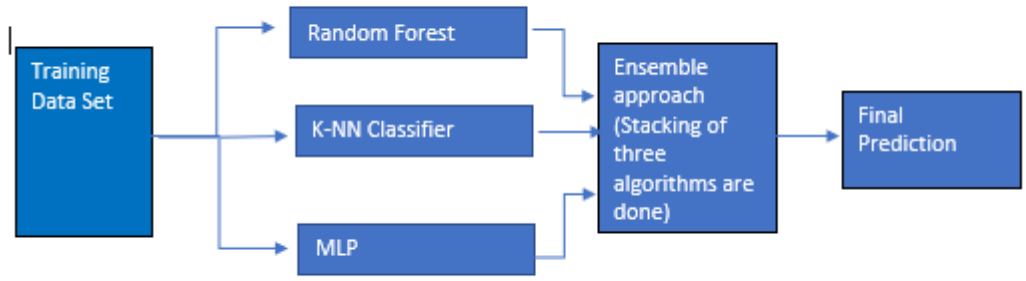

Figure 2: Design of the IDS solution for combining Result of multiple ML models.

4.1 Steps involved in the proposed IDS solution:

Data Preparation
The dataset considered for the experiment consisted of 15 separate datasets collected over different timelines which consisted of 37 power system event scenarios in each. These 15 datasets were merged to form a single dataset.

Data Pre-processing

Data pre-processing stage is one of the very important stages to ensure the input data for the machine learning models are accurate and interpretable to the algorithms to ensure correct analysis of the input and an accurate prediction. Following steps were performed in this stage for this project.

- Dataset was divided into training and test in 70% and 30% ratio.
- Training and test data sets were cleaned to remove unknown values.
- Sting values of marker column were changed to binary values to support interpretation of ML algorithms.
- After the data was cleaned and string values were converted to numerical values the dataset was balanced with RapidMiner tool to avoid overfitting of the models.

Data Classification

- Three ML models were run in isolation to obtain the individual results on prediction parameters like True Positive, False Positive, True Negative and False Negative.
- On the basis of above findings, the accuracy, precision, sensitivity and ROC were calculated. The final dataset obtained was used as an input for all three models in isolation.
- Lastly with the help of a python code the individual rows detected as an attack (true positive) were marked for each of the model.

Final Prediction

Individual model's outputs (True positive detections) were compared with Ensemble function in python code to find the final prediction result (confusion matrix) for the project.

4.2 Algorithms chosen for this research and why

Random Forest

Random Forest is a kind of supervised machine learning method. A random forest is created when numerous decision trees collaborate. Each decision tree functions separately, and the outcomes are then averaged to provide a conclusion. Averaging is accomplished by bootstrap aggregation of all the outputs from several decision trees (Breiman, 2001).

The various advantages of Random Forest (RF) are provided below (Ali *et al.*, 2012).

- Generally, accuracy of RF has been found to be on the higher side for most kind of the datasets.
- Specifically, it had been found to work very well with larger datasets, which is the case here with our dataset as well.
- It consists of different hyperparameters which could be controlled to get good results.
- Overfitting issues are minimal with RF unlike other algorithms.

Aside to that, RF has been found to be used in various research (Zhang *et al*., 2019) related to IDS systems for IT infrastructure and OT with good predictability (S.D.D Anton *et al.*, 2019).

K-NN

K Nearest Neighbour is a basic machine learning method. Machine learning models anticipate output values based on a collection of input values. KNN is a very basic machine learning method that is mostly used for classification. It classifies the data point based on the classification of its neighbours.

The main advantages of using K-NN algorithm are followings:

- The training time for K-NN is very low or almost nil as it uses stored training data during prediction on real-time. This makes K-NN faster than other algorithms.
- Due to its lazy learning characteristic, data can be added seamlessly, and it does not impact its accuracy.
- It is easy to implement.

Aside to that in historical research on IDS designs with other IT (Fauzi, Hanuranto and Setianingsih, 2020) and OT datasets it has stood fairly in prediction of Cyber-attacks (Durazno *et al.,* 2020).

MLP

Multi-layer perceptron is a kind of feedforward artificial neural network (ANN). also termed casually as 'vanilla neural networks' specifically when it is found to have a mono-layer structure.

MLP consists of three layers of nodes at the minimum, which includes an input layer, a hidden layer, and an output layer. Each node, except for the input nodes, is a neuron with a nonlinear activation function. MLP is trained using backpropagation technique. Difference between MLP and a linear perceptron is existence of multiple layers in MLP and its process of activation which is non-linear in nature. The capability of differentiating non-linearly separable date makes it unique. (Rosenblatt, 1961)

The advantages of using MLP are given below[2].

- MLP can work on tabular format of data such as csv format.
- It has proved to be a reliable method for predicting scenarios for different scenarios for long time.

Aside to that during literature review, it was observed, not many of the researchers have used this model, although other neural network models have been used very often. Hence it was chosen to find out its predictability with the selected dataset.

Stacked (Ensemble) Approach

Combination of multiple data mining algorithms is termed as hybrid approach. In this research the combination of outputs from Random Forest, K-nearest neighbor and Multilayer perceptron is used to reach the goal of determining a greater accuracy and precision in detection of cyber-

[2] https://machinelearningmastery.com/when-to-use-mlp-cnn-and-rnn-neural-networks/

attacks in Operational Technology environments. Ensemble approach consists of multiple methods like stacking, bagging, and boosting. This research has used the stacking method, which takes the input from multiple classifiers and then compares and combine to produce a better prediction based on its input from different classifiers.
This method has proved itself in other field of study and traditional IT (Wang *et al.*, 2016) and OT infrastructure but not many historical studies have been found in operational technology environment which have engaged this model (Khoei *et al.*, 2021). Hence this model was selected to find out whether it could help in enhancing the performance of the IDS in ICS/OT arena[3].

5 Implementation

This section of the report consists of the steps that has been performed to reach the result of the research, with the excerpts of the codes and configuration highlights.

5.1 Software and Hardware Used:

The Integrated Development Environment software (IDE), that was used for this project is Google Colaboratory (Colab). The details of the Software and hardware used for the project has been provided in the research configuration manual.

5.2 Dataset used for the analysis

The data set used for this research has been collected from Mississippi State University and Oak Ridge National Laboratory which included fifteen sets, each having 37 power system event scenarios. The binary datasets were in the CSV format, which was also compatible with Weka. The 37 scenarios were separated into eight natural events, one without an event, and others with attacks (28) (Borges Hink *et al.*, 2014).

5.3 Datafiles used for the analysis

The files which have been used for this project are mentioned below. Research configuration manual could be referred for details on the files mentioned below.

- 20194587_DissertationF.ipynb
- Data1 to Data15
- Final_df_unsampled (1).csv

5.4 Libraries imported and used for this thesis

- Pandas

[3] http://theprofessionalspoint.blogspot.com/2019/02/advantages-and-disadvantages-of-knn.html

- NumPy
- Sklearn
- Matplotlib
- Seaborn

5.5 Data Pre-processing

This stage comprised of the following steps.

- 15 datasets were merged into a single dataset for analysis.
- Cleaning of dataset has been performed.
- String values were transformed to numerical values.
- The dataset was balanced by up sampling the benign events with help of RapidMiner data processing tool.

Details on above steps with relevant codes have been described in the configuration manual.

5.6 Algorithms Used

This stage is comprised of running four ML analysis models to find out the parameters like accuracy, precision, and sensitivity. The algorithms used in this section are mentioned below.

- Random Forest
- K-Nearest Neighbour
- Multilayer perceptron (MLP)
- Stacked Ensemble Learning

In this section the predictions by top three models have been combined by taking the outputs and feeding it into the ensemble technique (stacking algorithm), that compared the output and computed the best possible result from all three inputs and provided us with the final predictions.

6 Evaluation

This section of the report displays the results of the various experiments carried out to find the accuracy, precision and sensitivity of the models explained in the above sections. Here the model has been trained using 70% of the dataset and then the algorithm was tested over 30% of the test dataset to find the statistics provided below.

6.1 Experiment done using Random Forest (RF)algorithm.

The first model has been trained by Random Forest algorithm and acceptable level of scores for the evaluation parameters have been obtained.

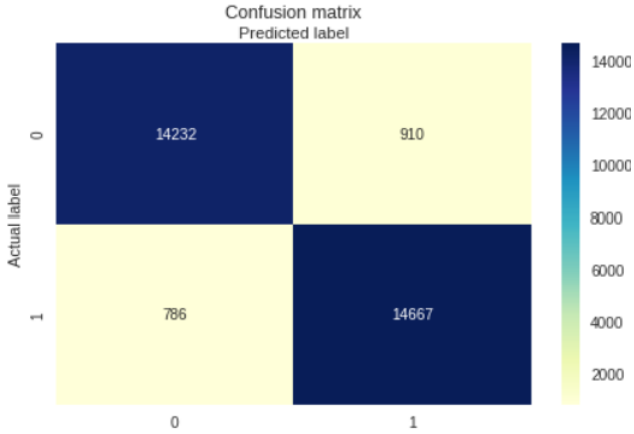

Figure 11: Confusion matrix for Random Forest.

Evaluation Parameters			
	Accuracy	Precision	Sensitivity
Random Forest	94.46	93.99	94.77

Figure 12: Evaluation parameter scores (RF

6.2 Experiment done using K-NN algorithm.

The second experiment has been performed with K-NN algorithm and scores of the evaluation parameters are provided below along with confusion matrix for clarity.

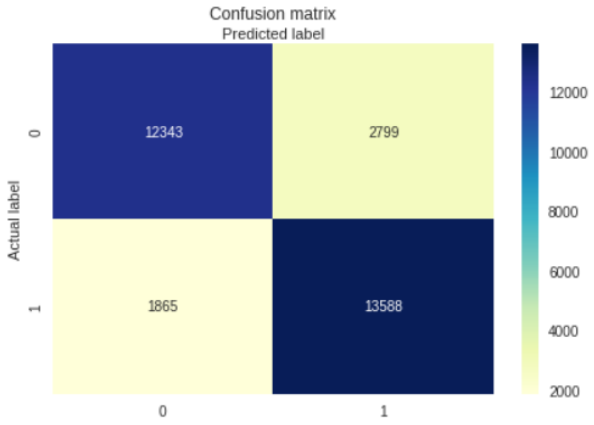

Figure 14: Confusion matrix for K-NN.

Evaluation Parameters			
	Accuracy	Precision	Sensitivity
K-NN	84.76	81.51	86.87

Figure 15: Evaluation parameter scores (K-NN)

6.3 Experiment done using MLP algorithm.

The third experiment has been conducted with the Multilayer perceptron algorithm to check if it performs well with the OT dataset, and it did not seem to do well except for the sensitivity.

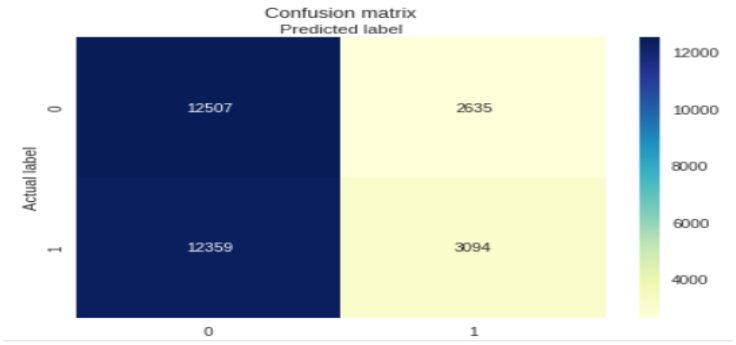

Figure 17: Confusion matrix for MLP.

Evaluation Parameters			
	Accuracy	Precision	Sensitivity
MLP	49.49	49.80	98.60

Figure 18: Evaluation parameter scores (MLP)

6.4 Experiment done using Stacking (Ensemble Learning) model.

The last experiment has been performed with stacking algorithm, which is a type of ensemble learning, where the training data from previous three models were fed as input to compare and combine to produce the best possible prediction.

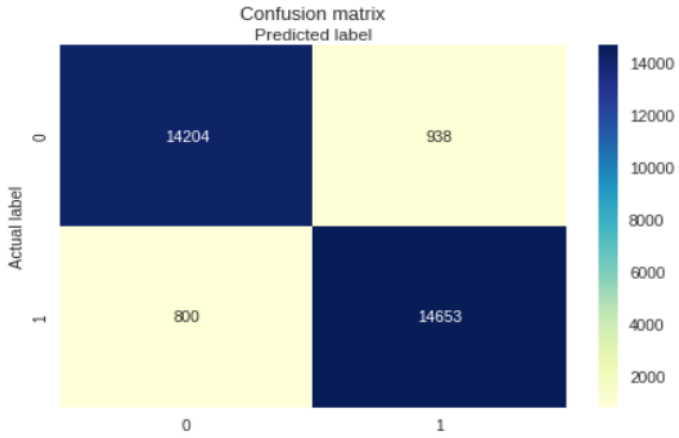

Figure 20: Confusion matrix for Stacking Model.

Evaluation Parameters			
	Accuracy	Precision	Sensitivity
Stacking Model	94.48	93.98	94.83

Figure 21: Evaluation parameter scores (Stacking model)

7 Discussion

The above section discussed about the experiments conducted with various data analysis models comprising of ML algorithms like Random Forest, K-NN, MLP and Stacked ensemble learning to predict accuracy, precision, and sensitivity. The results when compared revealed

that Stacked ensemble learning model produced marginally better result over Random Forest, and better than other two models using K-NN and MLP algorithms. The performance of the models has been provided below in a comparison bar chart to demonstrate the difference among them.

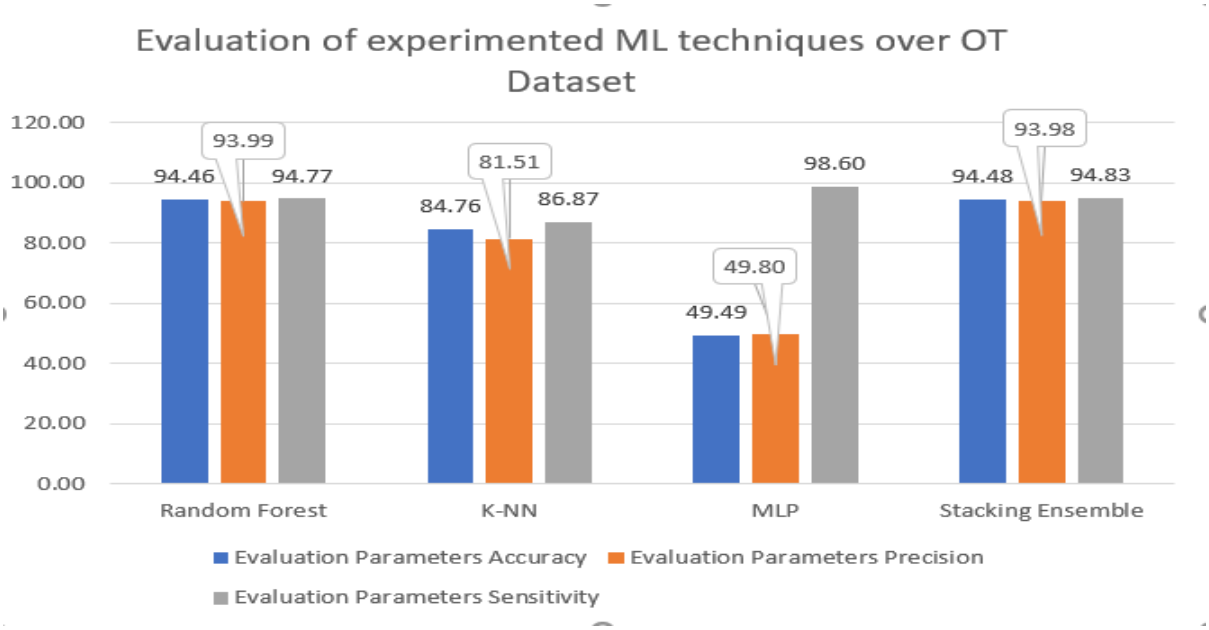

Figure 22: Comparison of accuracy, precision, and sensitivity.

From the comparison it is evident that for the used dataset (Beaver et al.2013), among the 4 models Stacking ensemble and Random Forest have performed better than others ensuring better accuracy and precision which would provide consistent level of true positive alerts to make a sample IDS perform efficiently, although there are scopes of further improvement with the predictability of the models.

The only unsupervised ML model used in this research was MLP (a variation of Artificial Neural Network) and it did not perform well. It can be observed from the confusion matrix provided above, that MLP had the highest number of false positive detections. The section where it shines is its sensitivity, which would ensure any minor fluctuations or changes in environment might get detected by the IDS.

8 Conclusion and Future Work

8.1 Conclusion

The thesis was conducted with industrial control system dataset (Borges Hink *et al.*, 2014) which was simulated in lab environment, and it consisted of 14233275 events among which, 70% had were used for training the data and rest 30% for the test. The experiments conducted over the datasets included analysis with four different ML models and the algorithms used were Random Forest, K-nearest neighbour, Multilayer perceptron, and Stacked ensemble learning.

The best results were delivered by Stacked Ensemble learning, although marginally over Random Forest classifier. Hence primarily it can be concluded as per this research hybrid models perform better than individual algorithms. Secondly, considering this research and previous studies (S.D.D Anton et al. 2019) (Borges Hink *et al.*, 2014)it could also be derived that for ICS datasets Random Forest performs better than other isolated classifiers.

8.2 Future Work

The highest accuracy obtained from this research, for the given dataset (Borges Hink *et al.*, 2014) was 94.48% and hence more studies with variations in datasets, analysis techniques and variation of combination of different models are required to be researched on, to achieve a higher level of accuracy for the IDS for OT.
Secondly, it had been difficult to collect real-life OT dataset in the limited timeline of the dissertation and as a future work the research would consider using real world ICS datasets to work on with more variety of analysis models to find out the best fit for higher performance.

References

Ahmed, L.A.H. and Hamad, Y.A.M. (2021) 'Machine Learning Techniques for Network-based Intrusion Detection System: A Survey Paper', in *2021 National Computing Colleges Conference (NCCC). 2021 National Computing Colleges Conference (NCCC)*, pp. 1–7. doi:10.1109/NCCC49330.2021.9428827.

Ali, J. *et al.* (no date) 'Random Forests and Decision Trees'.

Almalawi, A. *et al.* (2016) 'An Efficient Data-Driven Clustering Technique to Detect Attacks in SCADA Systems', *IEEE Transactions on Information Forensics and Security*, 11(5), pp. 893–906. doi:10.1109/TIFS.2015.2512522.

Altunay, H.C. *et al.* (2021) 'Analysis of Anomaly Detection Approaches Performed Through Deep Learning Methods in SCADA Systems', in *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pp. 1–6. doi:10.1109/HORA52670.2021.9461273.

Anton, S.D.D., Sinha, S. and Dieter Schotten, H. (2019) 'Anomaly-based Intrusion Detection in Industrial Data with SVM and Random Forests', in *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM). 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pp. 1–6. doi:10.23919/SOFTCOM.2019.8903672.

Beaver, J.M., Borges-Hink, R.C. and Buckner, M.A. (2013) 'An Evaluation of Machine Learning Methods to Detect Malicious SCADA Communications', in *2013 12th International Conference on Machine Learning and Applications. 2013 12th International Conference on Machine Learning and Applications*, pp. 54–59. doi:10.1109/ICMLA.2013.105.

Borges Hink, R.C. *et al.* (2014) 'Machine learning for power system disturbance and cyber-attack discrimination', in *2014 7th International Symposium on Resilient Control Systems (ISRCS)*. *2014 7th International Symposium on Resilient Control Systems (ISRCS)*, pp. 1–8. doi:10.1109/ISRCS.2014.6900095.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, 45(1), pp. 5–32. doi:10.1023/A:1010933404324.

Derhab, A. *et al.* (2019) 'Blockchain and Random Subspace Learning-Based IDS for SDN-Enabled Industrial IoT Security', *Sensors*, 19(14), p. 3119. doi:10.3390/s19143119.

Fauzi, M.A., Hanuranto, A.T. and Setianingsih, C. (2020) 'Intrusion Detection System using Genetic Algorithm and K-NN Algorithm on Dos Attack', in *2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS)*. *2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS)*, pp. 1–6. doi:10.1109/ICORIS50180.2020.9320822.

Feng, C., Li, T. and Chana, D. (2017) 'Multi-level Anomaly Detection in Industrial Control Systems via Package Signatures and LSTM Networks', in *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pp. 261–272. doi:10.1109/DSN.2017.34.

Khan, I.A. *et al.* (2019) 'HML-IDS: A Hybrid-Multilevel Anomaly Prediction Approach for Intrusion Detection in SCADA Systems', *IEEE Access*, 7, pp. 89507–89521. doi:10.1109/ACCESS.2019.2925838.

Khoei, T.T. *et al.* (2021) 'Ensemble Learning Methods for Anomaly Intrusion Detection System in Smart Grid', in *2021 IEEE International Conference on Electro Information Technology (EIT)*. *2021 IEEE International Conference on Electro Information Technology (EIT)*, pp. 129–135. doi:10.1109/EIT51626.2021.9491891.

Kushner, D. (2013) 'The real story of stuxnet', *IEEE Spectrum*, 50(3), pp. 48–53. doi:10.1109/MSPEC.2013.6471059.

Lee, D., Kim, D. and Jung, J. (2008) 'Multi-Stage Intrusion Detection System Using Hidden Markov Model Algorithm', in *2008 International Conference on Information Science and Security (ICISS 2008)*. *2008 International Conference on Information Science and Security (ICISS 2008)*, pp. 72–77. doi:10.1109/ICISS.2008.22.

Robles-Durazno, A. *et al.* (2020) 'Real-time anomaly intrusion detection for a clean water supply system, utilising machine learning with novel energy-based features', in *2020 International Joint Conference on Neural Networks (IJCNN)*. *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. doi:10.1109/IJCNN48605.2020.9207462.

Rosenblatt, F. (1961) *PRINCIPLES OF NEURODYNAMICS. PERCEPTRONS AND THE THEORY OF BRAIN MECHANISMS*. CORNELL AERONAUTICAL LAB INC BUFFALO NY. Available at: https://apps.dtic.mil/sti/citations/AD0256582 (Accessed: 3 January 2022).

Sawas, A.M., Khani, H. and Farag, H.E.Z. (2021) 'On the Resiliency of Power and Gas Integration Resources Against Cyber Attacks', *IEEE Transactions on Industrial Informatics*, 17(5), pp. 3099–3110. doi:10.1109/TII.2020.3007425.

Sewak, M., Sahay, S.K. and Rathore, H. (2018) 'Comparison of Deep Learning and the Classical Machine Learning Algorithm for the Malware Detection', in *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pp. 293–296. doi:10.1109/SNPD.2018.8441123.

da Silva, E.G. *et al.* (2016) 'A One-Class NIDS for SDN-Based SCADA Systems', in *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*. *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, pp. 303–312. doi:10.1109/COMPSAC.2016.32.

Tama, B.A., Comuzzi, M. and Rhee, K.-H. (2019) 'TSE-IDS: A Two-Stage Classifier Ensemble for Intelligent Anomaly-Based Intrusion Detection System', *IEEE Access*, 7, pp. 94497–94507. doi:10.1109/ACCESS.2019.2928048.

Vávra, J. and Hromada, M. (2017) 'Evaluation of anomaly detection based on classification in relation to SCADA', in *2017 International Conference on Military Technologies (ICMT)*. *2017 International Conference on Military Technologies (ICMT)*, pp. 330–334. doi:10.1109/MILTECHS.2017.7988779.

Wang, W. *et al.* (2021) 'Anomaly detection of industrial control systems based on transfer learning', *Tsinghua Science and Technology*, 26(6), pp. 821–832. doi:10.26599/TST.2020.9010041.

Yang, H., Cheng, L. and Chuah, M.C. (2019) 'Deep-Learning-Based Network Intrusion Detection for SCADA Systems', in *2019 IEEE Conference on Communications and Network Security (CNS)*. *2019 IEEE Conference on Communications and Network Security (CNS)*, pp. 1–7. doi:10.1109/CNS.2019.8802785.

Zhang, F. *et al.* (2019) 'Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data', *IEEE Transactions on Industrial Informatics*, 15(7), pp. 4362–4369. doi:10.1109/TII.2019.2891261.