Journal of Theoretical and Applied Information Technology
30th June 2020. Vol.98. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CYBER SECURITY BASED MACHINE LEARNING ALGORITHMS APPLIED TO INDUSTRY 4.0 APPLICATION CASE: DEVELOPMENT OF NETWORK INTRUSION DETECTION SYSTEM USING HYBRID METHOD

[1]SARA TAMY,[2]HICHAM BELHADAOUI,[3]NABILA RABBAH,[4]MOUNIR RIFI

[1,2,4]Laboratory of Network, Computing, Telecommunications and Multimedia, ESTC, Hassan II University, BP. 8012, Casablanca, Morocco

[3]Laboratory of Structural Engineering, Intelligent Systems and Electrical Energy, ENSAM, Hassan II University, BP. 20000, Casablanca, Morocco

E-mail: [1]saratamy@yahoo.fr

ABSTRACT

The implementation of industry 4.0 is a complex process that involves several steps and management of a project involving all the company's resources: production, supply chain, engineering, maintenance, human resources, information systems and many others.
faced of the risks weighing on companies, and even more on their sensitive data, the need to create a cybersecurity strategy for the Industry 4.0 is more essential than ever.

It has been apparent for several years that industrial systems are vulnerable to computer attacks. This can be explained because they were not designed with security constraints, particularly because of their physical isolation from Internet. However, they now face a variety of attackers with different objectives and abilities. In this paper, we present our strategy of cybersecurity, based on machine learning algorithms, applied in the context of industry 4.0. For this purpose, Unified Threat Management based on machine learning algorithms (ML-UTM) was used. to touch all layers of pyramid of Computer Integrated Manufacturing we propose to put an ML-UTM between layer 4 and 3, and an Industrial Unified Threat Management based on machine learning algorithms (ML-IUTM) between layer 3 and 2, another ML-IUTM between layer 2 and 1. Then we will cite the works based on the use of filtering device between layers 1 and 0.

This paper describes a machine learning approach to build an efficient and accurate network intrusion detection system, which is one of the features of UTM, using a hybrid method. Thus, we have combined the different machine learning algorithms, namely Support vector machine (SVM), One rule (OneR), K-nearest neighbor (K-NN) and Random forest (RF) with Particle Swarm Optimization (PSO) method using a real data set (Gas pipeline), and according to the results of our analysis, we have selected the best optimized classifier.

The experimental results have demonstrated the reliability and efficiency of the proposed approach. The PSO method can provide various advantages to K-NN and RF classifiers such as higher accuracy, lower MSE and faster time to build model. After analyzing and comparing all these results it was found that the NIDS based RF optimized by PSO give the best performances, with accuracy of 99.30%, F-measure of 99,30% and MSE that has been reduced to 0,0034.

Keywords: *Unified Threat Management, Intrusion Detection System, Machine Learning, Particle Swarm Optimization, industry 4.0*

Journal of Theoretical and Applied Information Technology
30th June 2020. Vol.98. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1. INTRODUCTION

The fourth industrial revolution also called industry 4.0 has created a new market in which customers, companies and machines are interconnected. It involves production processes that combine internet of thing (IoT) and other digital technologies, such as robotics, 3D printing, augmented reality and artificial intelligence (AI), to exploit information from Big data and digital models [1]. Nevertheless, despite all these advantages, it is also entails new challenges. The security of data is more imperative than ever. Thus, data have to be protected against misuse and unauthorized access.

The implementation of new technologies has to go hand in hand with the creation of an environment in which data can be stored, backed up, sent, shared securely and in which customers and manufacturers are protected. With the fourth industrial revolution, the architecture has fundamentally changed. the industrial network is connected to internet and to corporate network. therefore, cybersecurity approach has to be applied. In this paper, we propose to use Unified Threat Management (UTM) based on machine learning techniques (ML), as it is an all-in-one security solution with many features including: firewall, spam filtering, antivirus, intrusion detection or prevention system (IDS or IPS), and application content filtering [2].

Previous research has examined the application of machine learning techniques for the prediction and classification of intrusions in industrial networks. However, these studies focus on the particular impacts of specific machine learning techniques and not on the optimization of these techniques using optimized methods.

The main contributions of this paper are:

- proposing a global security strategy to guarantee the security of all layers of pyramid of Computer Integrated Manufacturing, so as to use Unified Threat Management based on machine learning algorithms (ML-UTM) and Industrial Unified Threat Management based on machine learning algorithms (ML-IUTM).
- Optimization of Network Intrusion Detection System which is one of the features of the UTM by using different classification algorithms namely Support vector machine, One rule, K-nearest neighbor and Random forest optimized by Particle Swarm Optimization (PSO).
- Comparison of different data mining algorithms on the gas pipeline dataset.
- Identification of the best performance based algorithm for intrusion detection.

Paper outline: In This paper we focus on performance assessment of machine learning algorithms such as SVM, k-NN, OneR and RF in detecting attacks in SCADA systems. section II presents the evolution of industrial network. Related works is discussed in section 3, section 4 presents problematics (security problems and problems related to real time and determinism). In section 5 we describe our proposed approach for network security. Then, we present an application case which include the result of classifiers performances in section 6, and we compare them with the results based on our approach which aims to use the PSO method for optimization of algorithms. Finally, we conclude our work in section 7 and propose directions for future work.

2. EVOLUTION OF INDUSTRIAL NETWORK

The first industrial revolution was mechanical production with steam and hydraulics. The second one is mass production using electricity. The third is automated production with programmable logic controllers (PLCs) and supervisory systems (SCADA for Supervisory Control and Data Acquisition). Now, The Fourth Industrial Revolution is here. This is marked by the pursuit of automation and the central role of communication. The communication between human, system and machine, even the product itself [2]. Figure.1 shows Evolution of industrial network.

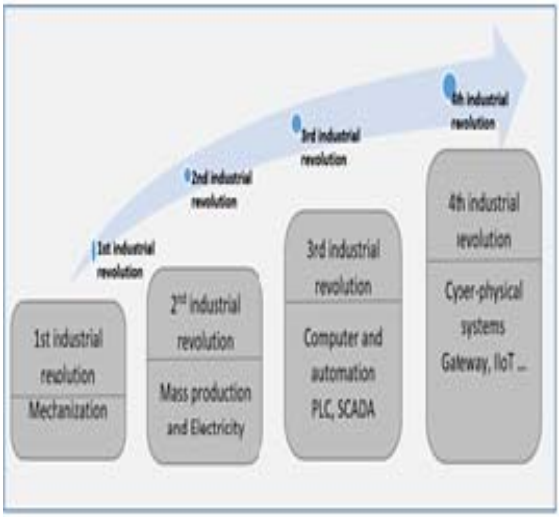

Figure.1 Evolution of Industrial Network.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. RELATED WORKS

with the arrival of industry 4.0, the industrial network architecture has fundamentally changed, the industrial network is connected to the Internet to various corporate networks, and to wireless communication systems [3]. Therefore, we have a strong increase in the number of network attacks, thus the detection of network anomalies has become a dynamic and primordial research. Several papers [4-7], discuss the machine learning techniques for detecting anomalies in SCADA systems.

Paper [4] presents a rule-based intrusion detection system (IDS) using a deep packet inspection (DPI) approach, including signature-based approaches and models adapted to SCADA systems. The rules proposed by the authors are based on signatures and can precisely detect many known malicious attacks. They also proposed model-based detection as a complementary technique for detecting unknown attacks. Its intrusion detection approaches proposed for SCADA networks are implemented and verified using Snort rules.

the authors of paper [5] have presented an anomaly detection method for the ICS. they have proposed a hybrid model that takes advantage of the anticipated and consistent nature of communication patterns. Their approach takes several steps; in which they have first applied some preprocessing techniques to standardize the data. then, they have applied dimensionality reduction algorithms to improve the process of anomaly detection. Third, an edited nearest neighbor rule algorithm have been employed to balance the dataset. Fourth, they have created a signature database by employing the Bloom filter. Finally, they have combined their package contents level detection with another instance based learner to make a hybrid method for anomaly detection to detect new attacks. The experimental results with a real large-scale dataset generated from a gas pipeline SCADA system show that their approach HML-IDS exceeds the reference models with an accuracy rate of 97%.

In [6], the authors have presented an unsupervised SCADA data driven anomaly detection approach in order to detect attacks in SCADA systems. It has been made automatically through identifying both the consistent and inconsistent states of SCADA data, and then extraction of proximity based detection rules for each behavior to detect inconsistent states. Their method allows automatically the identification of consistent and inconsistent states, with high accuracy. However, they have chosen to implement a full virtual SCADA lab and simulate a water distribution system as a supervised infrastructure, and they have applied their simulation only with Modbus/ TCP protocol.

In our previous work [7], we have applied supervised machine learning algorithms for predicting attacks in SCADA networks. The objective was to analyze and evaluate several classifiers based on recall, error rate, time to build the model, and accuracy of prediction in order to select the most efficient one. As a result, Random Forest is the best classifier used to predict attacks on SCADA networks using a "10% Random Sample Gas Pipeline" dataset.

In paper [8], the authors have presented an intelligent technique, the Grey Wolf Optimization (GWO), which is one of the newest optimization methods, inspired by nature and used for the selection of key feature. They have used the KDD CUP 99 dataset, and the features are optimally selected using the GWO algorithm. Thus, they have reduced the datasets from 41 to 24 features. After obtaining optimal feature sets, the authors have applied different classification methods, namely K-nearest neighbor, Support vector machine and Generalized regression neural network, to classify the data into the normal class or intrusion class. this optimization method increases the accuracy rate for GWO-KNN from 75.8% to 77.9% and GWO-SVM from 71.32% to 76.05% and finally for GWO-GRNN from 73.48% to 75%.

The authors in [9] have proposed a honey-pot based approach which is used in the network security for the real time intrusion detection and prevention system. Their approach consists of three groups: the honeypot server application, the monitor application, and the IDS application. They have presented a honey-pot based intrusion detection and prevention system (IDPS) which was able to show the network traffic on servers visually in real-time animation, and reduce the cost of information security in an enterprise network. The proposed system can also detect zero-day attacks using the intrusion detection configuration, which gives it superior performance compared to other IDSs. This system also helps to reduce the level of false positives in IDSs.

As can be seen from the previous works mentioned above, intrusion detection systems are widely used to detect and predict intrusions into SCADA networks, in order to protect it against threats and vulnerabilities. Nowadays, machine

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

learning techniques are widely used to build an efficient intrusion detection system, such as Regression, Classification, Clustering, Ensemble Methods and many others.

The majority of previous studies have investigated the machine learning methods to predict and classify intrusions into industrial networks. However, these studies have focused on the particular impacts of several machine learning techniques and have not exploited optimization methods.

There are several machine learning algorithms that are known by their good performance in the learning process. Thus, the choice of the appropriate algorithm is very important to implement an efficient intrusion detection system.

In this paper, we have used the algorithms: Support vector machine, One rule, K-nearest neighbor and Random forest, because they are the most used and popular ones, and also for the fact that some of them (Support vector machine and Random Forest) have been tested in our previous work [7], and they have given good results. The idea now is to present a hybrid approach for intrusion detection based on the optimization of these algorithms by the Particle Swarm Optimization method, which has the advantage of being effective on a wide variety of problems in different areas, with no need to change the basic principle of method's algorithm.

4. PROBLEMATICS

4.1. Security Problems

Threats in the industrial system are caused by malicious acts or accidents, irregular procedures and technical malfunctions. These threats have historically been limited to internal elements of the industrial system, principally staff members, operators or technical support personnel [2].

Today, with the fourth industrial revolution, all industrial systems are connected to Internet, various corporate networks and wireless communication systems. Thus, there are many ways to access and control the industrial system network, such as: Internet connections, compromised virtual private networks, weak authentication protocols and industrial system components or buffer overflow attacks on industrial system control servers, accessible through PLCs and human-machine interfaces [10].

4.2. Problems Related to Real Time and Determinism.

Industry 4.0 allows more flexibility in production infrastructure. It also makes it easier to launch production in a reactive way and make decisions in real time, through remote connection. For the analysis, supervision or storage of the history of each site, there is a tolerance. The problem arises if you want to control a machine or make a decision remotely and in real time.

One of the constraints that real time must satisfy is determinism. It is necessary to take into consideration the respect of timing, and the priority of events, in case of emergency a hierarchy is established between the different treatments, to define the most important ones [11].

4.3. Overview of Attacks

This section presents a non-exhaustive list of attacks or failures that have affected industrial systems in the past, in order to give the reader a better idea of their causes and consequences and to show the different forms that an attack can take.

Maroochy Shire [12]: An ex-employee of Hunter Watertech, a subcontractor of Maroochy Shire County Council, discharges the contents of a waste water tank into the wild. He has used material from HWT, namely a RTU PDS Compact 500, a laptop computer with HWT software for reprogramming RTUs and a radio transceiver. The HWT RTU PDS Compact 500 communicates via MODBUS and DNP3 protocols. the attack have caused a many defects including:

- alarms that were not reported to the SCADA;
- SCADA which was not able to communicate with the pumps;
- pumps that didn't work.

As a result of these defects, a tank overflowed.

Jeep Cherokee [13-14]: Charlie Miller and Chris Valasek, security researchers at Uber, show on two demonstrations how to take total control of connected cars. Both demonstrations were covered by journalist Andy Greenberg. The first demonstration in 2013 was made on a Ford Escape vehicle using a computer directly connected to the car's control systems. The second demonstration in 2015 was carried out on a Jeep Cherokee via Internet using a zero-day vulnerability. In both cases, they were able to inject CAN bus packets containing malicious commands. They were able to turn on the car's radio, disable the brakes or accelerator, control the steering wheel, etc.

Aurora [12], [15]: In 2007, the Aurora project conducted by the National Laboratory of Idaho demonstrated the need to protect industrial equipment from computer attacks. This test was conducted on a diesel generator controlled by circuit breakers. Researchers sent opening and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

closing commands to the circuit-breakers in an unsynchronized way. These orders had physical consequences on the generator's rotation mechanisms, causing its destruction. This demonstration was performed by operators legitimately connected to the system which can be seen as malicious operators. However, the generator circuit breakers are controlled via the MODBUS protocol, which does not provide any security mechanism. Thus, the same attack could have been carried out by an attacker present on the network and able to send arbitrary commands. This type of attack can be avoided by rejecting flows that do not guarantee a certain delay between the opening and closing commands.

Stuxnet [16-17]: Between 2005 and 2010, Iran was the victim of an attack on its nuclear programme. The stexnet was introduced via an infected USB key by a subcontractor. Malicious binaries were signed with compromised Realtek Semiconductor Corps certificates to avoid raising suspicions. Once in the system, the worm spread, particularly via the network and removable media, to a computer used to program Siemens Simatic S7-315 PLCs. Once one of these machines was reached, the worm used it to modify the program executed by the PLC. This allowed him to intercept the messages sent and received by the PLC and replace them.

Before any action was taken, the virus monitored traffic to determine if the target was in a certain target state. This observation phase could last from thirteen days to three months. Then the virus entered the attack phase by modifying the values of the variables controlling the rotation speed of the centrifuges. Thus, Stuxnet has sabotaged the system by slowing down and accelerating the centrifuges at different times.

5. PROPOSED APPROACH FOR NETWORK SECURITY

In order to protect the entire network and guarantee a very high level of security, we propose to use ML-UTM (Machine Learning based Unified Threat Management) between layer 4 and 3, and an ML-IUTM (Machine Learning based Industrial UTM) between layer 3 and 2, another ML-IUTM between layer2 and 1. between layer 1 and 0 the authors have already proposed an applicative filter, we will cite some of the work afterwards. Figure2 show our proposed strategy for network security.

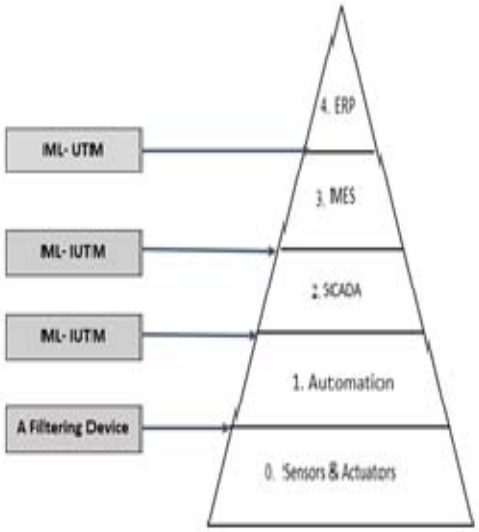

Figure2.Proposed strategy for Network Security

5.1. Unified Threat Management (UTM)

UTM is a solution that has been proposed, based on security approach all in one that provides firewall, intrusion detection system (IDS), intrusion prevention system (IPS), antimalware, content filtering and VPN.

UTMs may also include more advanced features such as identity-based access control, Quality of Service (QOS), load balancing, SSL and SSH inspection, and application aware [18].

UTMs are usually purchased as cloud services or network appliances.

Due to real-time exploitation of the logs transmitted by the firewall, we are at any time warned when a threat has been detected.

Among the features presented in UTM, we have Intrusion detection system and malware detection that have the same purpose which is protect a computer system from attackers and ensure that the security policy is respected. Historically, IDSs have focused on detecting exploits that make it possible to bypass the security policy of a computer system. To this end, the detection is based on two approaches which are: Scenarios approach, in this case the IDS uses a database of signatures, and attempts to associate data obtained from the system's information sources with that already known [19]. and the Behavioral Approach which was the first approach proposed and developed. Anderson proposes to detect violations of the system security policy by observing user behavior and comparing it with a behavior model considered normal, called a profile [20].

There are three types of IDS: The Network Based Intrusion Detection System (NIDS), The Host Based Intrusion Detection System (HIDS) and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Hybrid IDS, which uses the both NIDS and HIDS to have more relevant alerts.

In intrusion detection, four interesting indicators are often used:

- True Positive;
- False positives;
- True negative;
- False negatives.

5.2. Applicative Filter

Up to now we have only talked about IT part, for the strictly industrial part we will not treat it because it is not our field but we will cite some works:

The ARAMIS project can be highlighted [21]. The objective of this project is to propose a device to physically separate the networks of industrial systems and to filter exchanges taking into account the specific business constraints.

This device has to reject any flow identified as illegal, and therefore potentially malicious. As an embedded device, it has to respect memory constraints in addition to the time constraints of industrial communications.

In paper [12], the researchers used two filter input languages, called "low level" language and a Python API. They also proposed API functionalities to generate rules in low level language. More generally, they have shown how an applicative filter can help to contain a variant of the Maroochy Shire attack by filtering the commands that have stopped the pump.

6. APPLICATION CASE: DEVELOP AN NIDS BASED HYBRID METHOD

6.1. Machine Learning Algorithms used for Anomaly Detection:

Machine learning techniques can be used to build models that separate normal and abnormal behavior patterns in a data set [22]. Generally, using an unsupervised algorithm can hamper the detection of abnormal patterns, as several approaches of anomaly detection involve labeled observation samples of normal and/or abnormal behaviors. Therefore, researchers usually use a supervised or semi-supervised learning algorithm to detect abnormalities.

In this application case, we aim to develop a Network Intrusion Detection System (NIDS) based on machine learning techniques namely Support Vector Machine (SVM), K-Nearest Neighbors (KNN), One Rule (OneR), Random Forest (RF) optimized by PSO method to detect attacks targeting SCADA systems.

6.1.1. Support Vector Machine

Support Vector Machines (SVM) are one of the most used and powerful classifiers nowadays, it is a type of feedforward neural network using kernelized learning algorithms [22-23]. SVM can be used to classify linearly and non-linearly separable models, it is a precursor to finding the optimal separation hyperplane using linear SVM.

The concept is to maximize the margin between the closest points of the classes to find the optimal hyperplane of separation between two classes. consider the case of a linear discriminant function obtained by linear combination of the input vector $x=(X1,..Xn)^T$, with weight vector $w=(W1,...Wn)^T$:

$$f(x)= w^Tx+w0. \quad (1)$$

A discriminating hyperplane will satisfy:
x is class 1 if $w^Tx+w0\geq 0$;
x is class -1 if $w^Tx+w0 < 0$.
$f(x)=0$ is a hyperplane, called a separating hyperplane (Figure3).

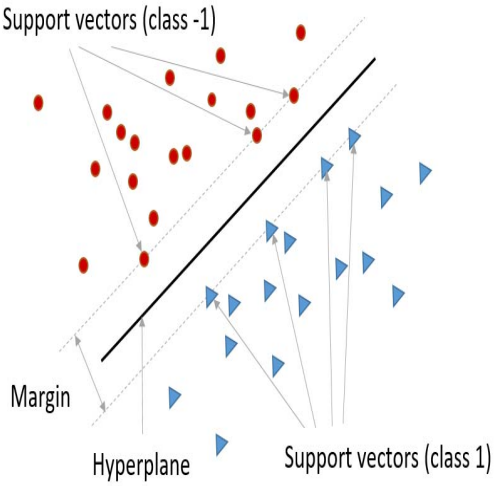

Figure3.Support Vectors Machine [24]

6.1.2. K-Nearest Neighbors

The k-nearest neighbors (KNN) is one of the popular and simplest methods. It is used for both classification and regression. The aim of KNN algorithm is to use a database in which the training examples are expressed as data points in the problem feature space and separated into several separate classes. Thus, to predict the target class of a new sample point x, first, it is projected in the considered feature space. after that the distances between x and the Kth nearest examples are calculated. then, x is classified using a majority vote of its neighbors. Similarity is defined using usual metrics, such as Euclidean distance, Hamming distance and Mahalanobis distance [25].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6.1.3. One Rule

One Rule (OneR) is a simple and powerful classifier which creates one rule for each predictor in training data and selects the rule with the smallest error rate for use as "one rule".

The One Rule algorithm operates as follows [26]:

1. From clustered set, generate a set of rule for each value of each feature predictor as in following steps:
   - Count the frequency of appearance of each value in the target class.
   - Identify the most frequent class.
   - Create a set of rules assign this class to this value of the predictor attribute.
   - Calculate the error rate occurs in the rules set for each attribute predictor.
2. Choose the best attribute predictors that have a smallest error rate as a classification rules.

6.1.4. Random Forest

Random Forest (RF) is an ensemble classifier used to improve accuracy. It is based on the standard machine learning technique known as decision tree. This technique, developed by Leo Breiman, enhances the accuracy of classification by embedding randomness in the construction of each individual tree or classifier [27]. The training phase of Random Forest is performed as follows [22]:

- k different and bootstrapped samples, each of size n, are drawn as separate decision trees (Figure4).
- Each of these decision trees is separately trained, and the set of these trees is the final classifier.

In the test phase, a novel observation runs through each decision tree, based on the decisions made at each node. Then, the predicted class label is provided by the decision which have the most votes from the different classifiers.

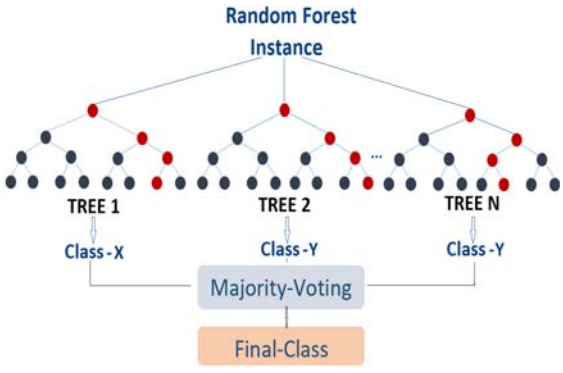

Figure4.Random Forest Simplified

6.2. Particle Swarm Optimization

The Particle Swarm Optimization (PSO) is a swarm intelligence technique which differs from the well-known scalable computational algorithms, like Genetic Algorithms, in that the population is not controlled by operators inspired by human's DNA procedures [28].

R.Eberhart and J.Kennedy, have been inspired by our living world in order to put in place a metaheuristic: optimization by swarm of particles. The method is based on collaboration of individuals between them: each particle moves and at each iteration, one of them closest to the optimal communicates its position to the others to allow them to modify their trajectory. The idea consist that a group of unintelligent individuals can have a complex global organization [28].

In particle swarm optimization, after the initialization of the population, each particle updates its velocity and its position in each iteration based on their own experience (pbest) and the best experience of all the particles (gbest). At the end of each iteration, the performance of all particles will be evaluated by the following functions:

$$v\,i\,[t+1] = w.\,v\,i\,[t] + c1\text{rand1}(p\,i,best[t] - p\,i\,[t] + c2\text{rand2}\,(p\,g,best[t] - pi\,[t]) \quad (2)$$

$$pi\,[t+1] = p\,i\,[t] + v\,i\,[t+1] \quad (3)$$

Where, i = 1,2,3,….. , N, N represent the a number of swarm population. $v\,i\,[t]$ is the velocity vector in $[t]th$ iteration. $pi[t]$ represent the current position of the ith particle. $p\,i,best[t]$ is the previous best position of ith particle and $p\,g,best[t]$ is the previous best position of whole particle. W is used to control the pressure of local and global search. $c1$ and $c2$ are acceleration constant. rand1 and rand2 are random number between 0 and 1.

We present bellow the flowchart of the PSO algorithm on Figure5.

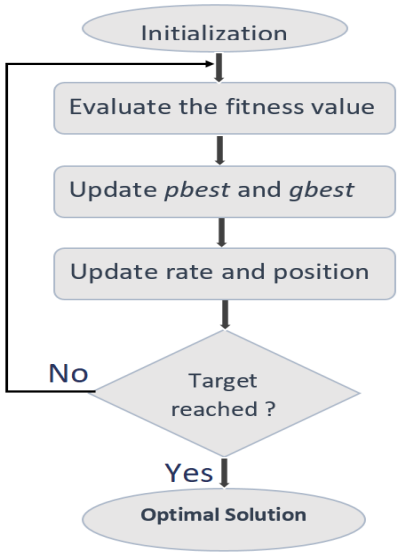

Figure 5. The flowchart of the PSO algorithm

Journal of Theoretical and Applied Information Technology
30th June 2020. Vol.98. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6.3. Experiment

In this section, we focus on the performance of classifiers for "Gas Pipeline" data set (web 1). This dataset is from a laboratory scale gas pipeline, and contains 97019 instances and 27 attributes (Table1). We have used 70% of dataset for training and 30% for testing. The full list of attack vectors for the gas pipeline dataset is shown in Table 2.

The experiment is conducted in a PC equipped with an Intel(R) Core™ i7 – 2760QM CPU, 16 GB RAM, and Windows 10 Professional operating system.

For simulation we have used WEKA (Waikato Environment for Knowledge Analysis) which is a collection of machine learning algorithms designedto facilitate the application of machine learning techniques to a variety of real-world problems, including tools for data preparation, classification, regression, clustering, association rule extraction and visualization (web2). And to get more accurate and highly detailed results we have used python programming language on Anaconda distribution which is a free software platform including the complete Python environment (v2.7 and 3.5) and many libraries (numpy, pylab...). it is provided with the SPYDER (Scientific Python Development Environment) work environment (web3).

Table 1: Dataset attributes

Gas Parameters	Type of attribute	Gas Parameters	Type of attribute
Command_address	Real	resest	Real
Response_address	Real	deadband	Real
Command_memory	Real	cycletime	Real
Response_memory	Real	rate	Real
command_memory count	Real	setpoint	Real
response_memory count	Real	control_mode	Real
comm_read_function	Real	control_scheme	Real
comm_write_fun	Real	pump	Real
resp_read_fun	Real	solenoid	Real
resp_write_fun	Real	crc_rate	Real
sub_function	Real	measurement	Real
command_length	Real	time	Real
resp_length	Real	result	Real
gain	Real		

Table 2: Attacks

Attack Name	Abbreviation
Naïve Malicious Reponse Injection	NMRI
Complex Malicious Response Injection	CMRI
Malicous State Command Injection	MSCI
Malicous Parameter Command Injection	MPCI
Malicous Function Code Injection	MFCI
Denial Of Service	DOS
Reconnaisance	Recon

6.4. Results and discussion

6.4.1. Metrics

In This section we will describe the metrics, evaluate the machine learning method used, and discuss the results. all the metrics are based on True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The metrics can be presented by the following equation:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (5)$$

$$Specificity = \frac{TN}{FP+TN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$F-measure = \frac{2*Precision*Sensitivity}{Precision+Sensitivity} \quad (8)$$

The performance of the classifiers is compared using the weighted average of the classifiers, and we were based on Precision, Recall and F-Measure. The precision average value of the best performance without optimization it's for RF with 98.80% and SVM with 95% than K-NN with 91.60%, and finally OneR with 83.60% (Table3). After the optimization by PSO we find that the best performance of precision it is for RF with 99.30 % than K-NN with 96.60 % shown inTable4.

Journal of Theoretical and Applied Information Technology
30th June 2020. Vol.98. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table3.Comparison of precision, Recall and F-Measure of classifiers

	SVM			K-NN			OneR			RF		
Class	**Precision**	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**
Normal	94,00%	98,00%	96,00%	93,20%	93,90%	93,50%	89,00%	88,70%	88,80%	99,10%	99,00%	99,10%
NMRI	100,00%	0,10%	0,20%	12,60%	11,70%	12,20%	98,10%	95,80%	96,90%	95,90%	95,80%	95,80%
CMRI	93,80%	98,30%	96,00%	92,60%	91,90%	92,30%	98,80%	98,20%	98,50%	99,40%	99,70%	99,60%
MSCI	95,70%	93,60%	94,60%	90,00%	92,30%	91,20%	0,00%	0,00%	0,00%	95,60%	92,80%	94,20%
MPCI	97,70%	98,10%	97,90%	96,20%	95,50%	95,80%	60,70%	52,40%	56,20%	95,60%	95,80%	95,70%
MFCI	100,00%	95,30%	97,60%	94,20%	94,20%	94,20%	0,00%	0,00%	0,00%	95,90%	95,30%	95,60%
DOS	98,30%	64,70%	78,00%	97,30%	92,00%	94,60%	0,00%	0,00%	0,00%	97,60%	96,70%	97,20%
Recon	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	59,20%	99,60%	74,30%	100,00%	100,00%	100,00%
Weighted Avg.	**95,00%**	**94,80%**	**93,30%**	**91,60%**	**91,70%**	**91,70%**	**83,60%**	**85,40%**	**84,10%**	**98,80%**	**98,80%**	**98,80%**

Table4.Performances of classifiers optimized by PSO method

	SVM			K-NN			OneR			RF		
Class	**Precision**	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**
Normal	92,60%	98,10%	95,30%	96,50%	98,00%	97,30%	89,00%	88,70%	88,80%	99,40%	99,50%	99,40%
NMRI	100,00%	0,10%	0,20%	99,70%	38,70%	55,80%	98,10%	95,80%	96,90%	99,50%	95,80%	97,60%
CMRI	93,90%	98,30%	96,00%	93,90%	98,30%	96,00%	98,80%	98,20%	98,50%	99,40%	99,90%	99,60%
MSCI	0,00%	0,00%	0,00%	95,70%	93,60%	94,60%	0,00%	0,00%	0,00%	95,70%	93,60%	94,60%
MPCI	97,70%	98,10%	97,90%	97,70%	98,10%	97,90%	60,70%	52,40%	56,20%	97,70%	98,10%	97,90%
MFCI	100,00%	95,30%	97,60%	100,00%	95,30%	97,60%	0,00%	0,00%	0,00%	100,00%	95,30%	97,60%
DOS	98,70%	55,40%	71,00%	99,10%	97,60%	98,40%	0,00%	0,00%	0,00%	99,10%	97,50%	98,30%
Recon	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	59,20%	99,60%	74,30%	100,00%	100,00%	100,00%
Weighted Avg	**93,40%**	**93,90%**	**92,00%**	**96,60%**	**96,50%**	**96,10%**	**83,60%**	**85,40%**	**84,10%**	**99,30%**	**99,30%**	**99,30%**

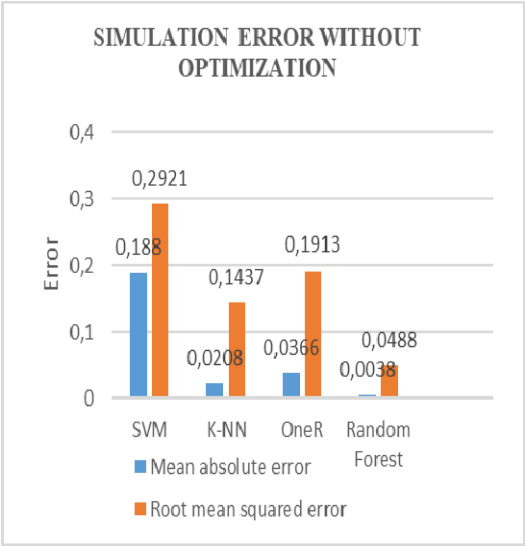

Figure6. Simulation error without optimization

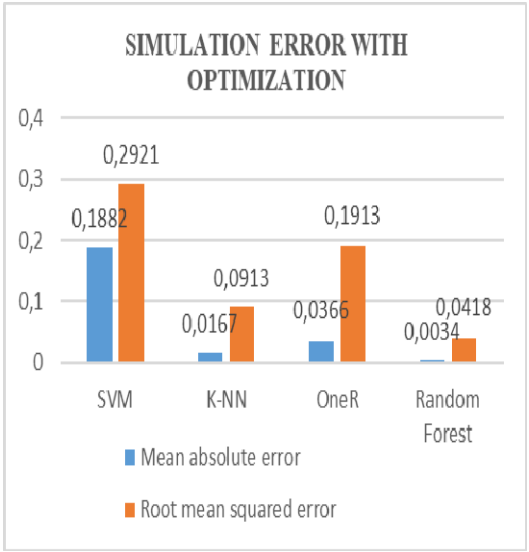

Figure7. Simulation error with optimization

Journal of Theoretical and Applied Information Technology
30th June 2020. Vol.98. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The experiment show that RF has the lowest MSE of 0.003 and RMSE of 0.048, followed by K-NN with MSE of 0.02 and RMSE of 0.143 then OneR with MSE of 0.036 and RMSE of 0.191 and finally SVM with MSE of 0.188 and RMSE of 0.292 (Figure6). The optimization method allows to reduce the error rate of RF and K-NN (Figure7).

Table5.Time to build model of classifiers

	Time to build model	Time to build optimized model
SVM	0,33	0,22
K-NN	494,16	299,1
OneR	0,16	0,16
RF	1,85	1,27

As shown in Table5, it has depicted that OneR has taken the shortest time 0.16 seconds to build the model, followed by SVM with 0.33 seconds, followed by RF with 1.85 then K-NN with 494.16 seconds.

The optimization by the PSO method allows to reduce the time to build the model for all the classifiers except OneR which has taken the same amount of time.

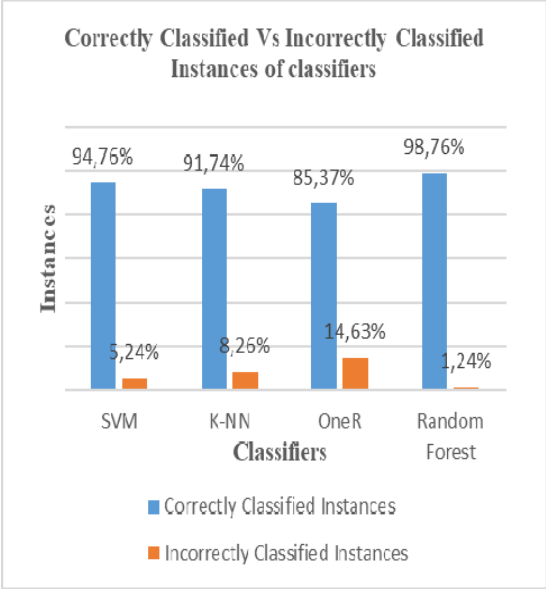

Figure8. Evaluation of classifiers without optimization

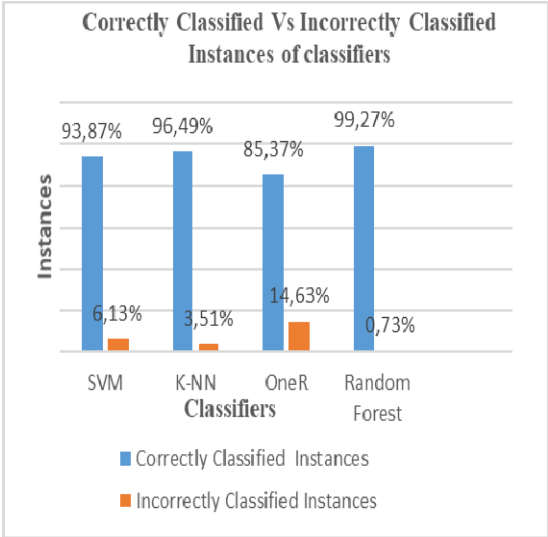

Figure9. Evaluation of classifiers after optimization

The results illustrated in the Figure8 show that Random Forest is the best classifier. it predicts better than other algorithms with accuracy of 98.76%. SVM shows the next higher correctly classified instances with accuracy of 94.76%. The K-NN has achieved 91.74% and OneR has achieved 85.37% accuracy.

Compared to the results of the experiments, we find in Figure9 that classifiers optimized by PSO give more efficient results than the classifiers alone for RF and K-NN. It has optimized the detection of intrusion passing from 98.76% to 99.27% for RF and from 91.74% to 96.49% for K-NN. For SVM and OneR the PSO method did not give better results.

6.4.2. Confusion matrix

The confusion matrix is the useful way to evaluate the classifier, each row of Table 6 and 7 represents rates in an actual class while each column shows predictions.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table6. Experimental result of classifiers without optimization in Confusion Matrix

	Normal	NMRI	CMRI	MSCI	MPCI	MFCI	DOS	Recon
Normal								
SVM	17987	0	300	10	44	0	6	0
K-NN	17227	665	339	24	78	1	13	0
OneR	16266	7	47	0	731	0	0	1296
RF	18172	34	27	10	92	0	12	0
NMRI								
SVM	828	1	0	0	0	0	0	0
K-NN	730	97	0	0	1	0	1	0
OneR	8	794	3	0	0	0	0	24
RF	35	794	0	0	0	0	0	0
CMRI								
SVM	77	0	4563	0	0	0	0	0
K-NN	377	0	4263	0	0	0	0	0
OneR	4	3	4557	0	0	0	0	76
RF	12	0	4628	0	0	0	0	0
MSCI								
SVM	15	0	0	220	0	0	0	0
K-NN	18	0	0	217	0	0	0	0
OneR	233	0	1	0	1	0	0	0
RF	17	0	0	218	0	0	0	0
MPCI								
SVM	43	0	0	0	2249	0	0	0
K-NN	92	3	0	0	2188	9	0	0
OneR	1089	0	3	0	1200	0	0	0
RF	88	0	0	0	2196	7	1	0
MFCI								
SVM	0	0	0	0	8	164	0	0
K-NN	2	0	0	0	8	162	0	0
OneR	166	0	1	0	1	0	0	4
RF	0	0	0	0	8	164	0	0
DOS								
SVM	195	0	0	0	0	0	357	0
K-NN	42	2	0	0	0	0	508	0
OneR	508	0	0	0	44	0	0	0
RF	18	0	0	0	0	0	534	0
Recon								
SVM	0	0	0	0	0	0	0	2042
K-NN	0	0	0	0	0	0	0	2042
OneR	4	5	0	0	0	0	0	2033
RF	0	0	0	0	0	0	0	2042

Table7 Experimental result of classifiers optimized by PSO in Confusion Matrix

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	Normal	NMRI	CMRI	MSCI	MPCI	MFCI	DOS	Recon
Normal								
SVM	18001	0	298	0	44	0	4	0
K-NN	17989	1	298	10	44	0	5	0
OneR	16266	7	47	0	731	0	0	1296
RF	18254	4	30	10	44	0	5	0
NMRI								
SVM	828	1	0	0	0	0	0	0
K-NN	508	321	0	0	0	0	0	0
OneR	8	794	3	0	0	0	0	24
RF	35	794	0	0	0	0	0	0
CMRI								
SVM	78	0	4562	0	0	0	0	0
K-NN	78	0	4562	0	0	0	0	0
OneR	4	3	4557	0	0	0	0	76
RF	4	0	4636	0	0	0	0	0
MSCI								
SVM	235	0	0	0	0	0	0	0
K-NN	15	0	0	220	0	0	0	0
OneR	233	0	1	0	1	0	0	0
RF	15	0	0	220	0	0	0	0
MPCI								
SVM	43	0	0	0	2249	0	0	0
K-NN	43	0	0	0	2249	0	0	0
OneR	1089	0	3	0	1200	0	0	0
RF	43	0	0	0	2249	0	0	0
MFCI								
SVM	0	0	0	0	8	164	0	0
K-NN	0	0	0	0	8	164	0	0
OneR	166	0	1	0	1	0	0	4
RF	0	0	0	0	8	164	0	0
DOS								
SVM	246	0	0	0	0	0	306	0
K-NN	13	0	0	0	0	0	539	0
OneR	508	0	0	0	44	0	0	0
RF	14	0	0	0	0	0	538	0
Recon								
SVM	0	0	0	0	0	0	0	2042
K-NN	0	0	0	0	0	0	0	2042
OneR	4	5	0	0	0	0	0	2033
RF	0	0	0	0	0	0	0	2042

The results, which are listed in Table6 and Table7, show that k-NN and RF classifiers optimized by PSO perform better than this models alone for the prediction and classification of intrusions. Random Forest (RF) is still the best

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classifier, it correctly classifies a large number of normal and malicious packets. The categorical classification report in confusion matrix shows the detection rate of each data type.

7. CONCLUSION AND FUTURE WORK

In this paper, we have explored the issues of cybersecurity in Industry 4.0 and proposed a cybersecurity strategy to protect the overall network. Our approach is based on the use of UTM based on machine learning algorithms. Throughout this article we tried to touch all layers of pyramid of Computer Integrated Manufacturing. By putting an ML-UTM between layer 4 and 3, and an ML-IUTM between layer 3 and 2, another ML-IUTM between layer 2 and 1. then the works based on the use of filtering device between layers 1 and 0 were cited.

As the UTM presents an all-in-one security solution, this paper focuses on the optimization of intrusion detection system, which is one of the UTM's features. for this aim, we have proposed to build an efficient and accurate IDS based on a hybrid system, in order to exploit and benefit from the high performance of the machine learning algorithms, namely SVM, OneR, K-NN and RF, and also from the advantages of the PSO method. These algorithms have been analyzed, optimized and evaluated, using Gas pipeline dataset, in order to confirm the validity of our approach as well as select the best optimized classifier to detect intrusions in the industry 4.0.

The first experiment results show that RF is the best classifier with a precision rate of 98,80%, a recall of 98,80%, F-measure of 98,80% and MSE of 0,0038. After the optimization of algorithms by PSO, the obtained simulation results show that RF is still the best classifier with a precision rate that has been increased to 99.30%, a recall rate of 99,30%, F-measure of 99,30% and MSE that has been reduced to 0,0034.

Our next work will focus on hybrid model using multiple classifiers to optimize intrusion detection.

REFERENCES

[1] L.Wang., & G. Wang, Guanghui. “Big data in cyber-physical systems, digital manufacturing and industry 4.0”. *International Journal of Engineering and Manufacturing (IJEM)*, 2016, Vol. 6, No 4, pp. 1-8.

[2] S. Tamy, H. Belhadaoui, N. rabbah, M. Rabbah, M. Rifi.. "Study of Strategies for Real-Time Supervision of Industrial Network Security." *In: Proc. of the International Conference on Smart Application and Data Analysis for Smart Cities (SADASC'18)*, 2018.

[3] A. Gilchrist. “Designing industrial internet systems”. *In: Industry 4.0. Apress, Berkeley, CA*, 2016. pp. 87-118.

[4] Y. Yang, K. McLaughlin, T. Littler, S. Sezer, B. Pranggono, & H.F. Wang. "Intrusion detection system for IEC 60870-5-104 based SCADA networks." *2013 IEEE power & energy society general meeting. IEEE*, 2013, pp. 1-5.

[5] I.A .Khan, Pi. DECHANG, Z. U. Khan, Y. Hussain, & A. Nawaz. “HML-IDS: a hybrid-multilevel anomaly prediction approach for intrusion detection in SCADA systems”. *IEEE Access, 7*, 2019, pp.89507-89521.

[6] A. Almalawi, X. Yu, Z. Tari, A. Fahad, & I. Khalil, "An unsupervised anomaly-based detection approach for integrity attacks on SCADA systems." *Computers & Security, Vol. 46*, 2014, pp. 94-110.

[7] S. Tamy, H. Belhadaoui, N. rabbah, M. Rabbah, M. Rifi.. "An Evaluation of Machine Learning Algorithms To Detect Attacks in SCADA Network." 2019 *7th Mediterranean Congress of Telecommunications (CMT). IEEE*, 2019, pp. 1-5.

[8] D.Srivastava, R.Singh, & V.Singh. “An intelligent gray wolf optimizer: a nature inspired technique in intrusion detection system (IDS)”. *Journal of Advancements in Robotics,* Vol. 6, No 1, 2019, pp. 18-24.

[9] M. Baykara, &R. Das. “A novel honeypot based security approach for real-time intrusion detection and prevention systems*”. Journal of Information Security and Applications,* Vol. 41, 2018, pp.103-116.

[10] S. Kwon, H. Yoo, T. Shon, & G. Lee. "Scenario-based attack route on industrial control system.". *International Conference on IT Convergence and Security (ICITCS). IEEE*, 2014, pp. 1-3.

[11] J-P. Elloy, "Les contraintes du temps réel dans les systèmes industriels répartis." *Revue générale de l'électricité* 2, 1991, pp.26-30.

[12] M. Puys. Sécurité des systèmes industriels: filtrage applicatif et recherche de scénarios d'attaques. Diss. 2018.

[13] A. Greenberg. "Hackers reveal nasty new car attacks–with me behind the wheel." Forbes 2013.

[14] A. Greenberg. "Hackers remotely kill a jeep on the highway—with me in it." Wired 7 (2015): 21.

[15] Aurora Project. Foia response documents. United States Department of Homeland

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Security. http://s3.documentcloud.org/documents/12125 30/14f00304-documents.pdf. (cf. p 18).2014

[16] N. Falliere, L-O. Murchu, and E. Chien. "W32. stuxnet dossier." White paper, Symantec Corp., Security Response, Vol. 5, No 6, 2011, pp. 29.

[17] R. Langner. "Stuxnet: Dissecting a cyberwarfare weapon." IEEE Security & Privacy Vol. 9, No. 3, 2011, pp. 49-51.

[18] W. Gauvin. "System and method for unified threat management with a relational rules methodology." U.S. Patent No. 7,735,116. 8 Jun. 2010.

[19] P. Biondi. "Architecture expérimentale pour la détection d'intrusions dans un système informatique." 2001.

[20] J-P. Anderson. "Computer security threat monitoring and surveillance." Fort Washington 1980.

[21] ARAMIS. "Architecture robuste pour les automates et matériels des infrastructures sensibles" http://aramis.minalogic.net/. (cf. p 34- 156). 2015-2017.

[22] Lopez Perez, R., Adamsky, F., Soua, R., & Engel, T. "Forget the Myth of the Air Gap: Machine Learning for Reliable Intrusion Detection in SCADA Systems". *EAI Endorsed Transactions on Security and Safety*.2019

[23] SALEH, Ahmed I., TALAAT, Fatma M., et LABIB, Labib M. A hybrid intrusion detection system (HIDS) based on prioritized k-nearest neighbors and optimized SVM classifiers. Artificial Intelligence Review, 2019, Vol. 51, No 3, pp. 403-443.

[24] BAGHAEE, Hamid Reza, MLAKIĆ, Dragan, NIKOLOVSKI, Srete, et al. "Support vector machine-based Islanding and grid fault detection in active distribution networks". *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2019.

[25] KASONGO, Sydney Mambwe et SUN, Yanxia. "A deep learning method with filter based feature engineering for wireless intrusion detection system". *IEEE Access*, Vol. 7, 2019, pp. 38597-38607.

[26] AL SAYYDEHA, Osama Nayel and MOHAMMAD, Mohammad Falah. "Diagnosis of the parkinson disease using enhanced fuzzy min-max neural network and OneR attribute evaluation method". *In: 2019 International Conference on Advanced Science and Engineering (ICOASE). IEEE*, 2019, pp. 64-69.

[27] COURONNÉ Raphael, PROBST, Philipp, and BOULESTEIX, Anne-Laure. "Random forest versus logistic regression: a large-scale benchmark experiment". *BMC bioinformatics*, vol. 19, no 1, 2018, pp. 270.

[28] Y. Khourdifi & M. Bahaj. "Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization". *In International Journal of Intelligent Engineering and Systems*. Vol.12, No.1, 2019

Web1:https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets

Web2: https://www.cs.waikato.ac.nz/ml/weka/

Web3: https://www.anaconda.com/