international journal of critical infrastructure protection 34 (2021) 100446

Contents lists available at ScienceDirect

International Journal of Critical Infrastructure Protection

journal homepage: www.elsevier.com/locate/ijcip

Adaptive anomaly detection system based on machine learning algorithms in an industrial control environment

Jan Vávra [a,b,*], Martin Hromada [a], Luděk Lukáš [a], Jacek Dworzecki [c,d,e]

[a] Tomas Bata University in Zlin, Zlin, Czech Republic
[b] nam. T. G. Masaryka 5555, 760 01 Zlin
[c] University of the Land Forces in Wroclaw, Poland
[d] Academy of the Police Force in Bratislava, Slovak Republic
[e] Pomeranian Academy in Slupsk, Poland

ARTICLE INFO

Keywords:
Cyber Security
Machine Learning
Critical Information Infrastructure
Anomaly Detection
Industrial Control System

ABSTRACT

Technology has become an integral part of contemporary society. The current transition from an industrial society to an information society is accompanied by the implementation of new technologies in every part of human activity. Increasing pressure to apply ICT in critical infrastructure resulted in the creation of new vulnerabilities. Traditional safety approaches are ineffective in a considerable number of cases. Therefore, machine learning another evolutionary step that provides robust solutions for extensive and sophisticated systems. The article focuses on cybersecurity research for industrial control systems that are widely used in the field of critical information infrastructure. Moreover, cybernetic protection for industrial control systems is one of the most important security types for a modern state. We present an adaptive solution for defense against cyber-attacks, which also consider the specifics of the industrial control systems environment. Moreover, the experiments are based on four machine learning algorithms (artificial neural network, recurrent neural network LSTM, isolation forest, and algorithm OCSVM). The proposed anomaly detection system utilizes multiple techniques and processes as preprocessing techniques, optimization techniques, and processes required for result interpretation. These procedures allow the creation of an adaptable and robust system that meets the need for industrial control systems.

1. Introduction

Information and communication technologies (ICT) have seen exponential growth over the last few decades. A considerable number of critical information infrastructure (CII) systems are dependents on these technologies. Moreover, automation, digitization, robotics for partial autonomous work with remote control are commonly implemented in CII. These interconnected systems are an essential part of the revolution, so-called "Industry 4.0″. Frank et al. [1] introduced areas where Industry 4.0 is applied. These areas include integration, energy management, traceability, automation, virtualization, Internet of Things (IoT), and cloud computing.

This technological revolution highly influences the critical information infrastructure and, subsequently, the critical infrastructure (CI) sector. Moreover, the functionality destabilization or loss of CII functionality can seriously impact the environment, the population, the financial sector, or the state's basic functions. Furthermore, CII commonly utilizes an area of information technology known as an Industrial Control System (ICS). ICS are systems designed to support, control, and monitor industrial processes utilized in a considerable number of industrial areas as power plants, dams, water treatment, oil production, chemicals, gas distribution, etc. Therefore, the cybersecurity of ICS and CII will be one of the main issues to solve in the following decades. The first step to create a reliable cybersecurity solution is to analyze the protected system itself. The cybersecurity characteristic of ICS can be described by three criteria: availability, confidentiality, and data integrity. We can conclude that the availability of the CII systems is the most important, up to our knowledge. Moreover, multiple authors concluded the same assumption [2, 3]. In a nutshell, the availability of ICS services is an important element to protect the critical requirements for contemporary society. Thus, the misclassified normal operation as cyber-attacks can limit ICS availability.

* Corresponding author.
E-mail address: janvav3@gmail.com (J. Vávra).

https://doi.org/10.1016/j.ijcip.2021.100446
Received 21 January 2021; Received in revised form 15 March 2021; Accepted 27 April 2021
Available online 24 May 2021

Although the field of ICS cybersecurity is relatively new, there is a considerable number of research papers. We identified relevant shortcomings of the state of the art for cyber-attacks detection. Sokolov et al. [4] presented the basic advantages and disadvantages of applying machine learning algorithms in the ICS environment. They pointed to the better detection capabilities of a group of techniques based on tree structures (e.g., Random Forest). However, the results showed susceptibility to "overfitting" of the created model, and therefore worse generalization properties of this solution according to the authors. The last examined group of machine learning algorithms was an artificial neural network. The authors declared the group as the most accurate one, but at the cost of higher computational complexity, especially with a large dimension of input data. Liu et al. [5] focused on designing an anomaly detection system based on convolutional neural networks to identify significant attributes and anomalies in ICS network traffic. The solution also utilizes an algorithm to define the states of the monitored system. They achieved a more robust solution, which had several shortcomings that needed to be addressed to interpret results and detection capabilities. Kravchik and Shabtai [6] applied a multilayer recurrent neural network to detect cyber-attacks in a wastewater treatment plant. Even though their research results are promising, the authors declared their solution's shortcomings, such as detection system limitation due to small dataset or interpretation of results.

A significant number of authors [4-7] present promising solutions, which do not take into account aspects and criteria for the ICS environment. Therefore, a considerable number of questions arise. One of the key issues for the deployment of machine learning methods is the computation demand of the machine learning detection model, which is usually a time-consuming process. ICS technology commonly consists of technology that is old for several decades. Therefore, the implementation of a novel technological system can be a difficult task in the ICS environment. The second key issue is focused on false alarms. Every anomaly detection system has to reduce the false classification of ICS normal processes in order to maintain system availability and operational continuity. The third issue includes the choice of an unsuitable machine learning area to detect anomalies. The majority of the researchers use a supervised learning group of machine learning algorithms. This anomaly detection system cannot identify unknown cyberattacks; therefore, there must be immense up to date databases with all cyber-attacks to train the classification model. The main shortcomings of the state of the art can be summarized in the following points:

- Detection of unknown cyberattacks,
- Scalability of the anomaly detection system,
- The adaptability of the anomaly detection system,
- High false alarm rate,
- High computational complexity of the anomaly detection system,
- Interpretation of cyberattacks.

The article presents an adaptive anomaly detection system for ICS which addresses all described issues. We considered detected anomalies as possible symptoms of cyber-attacks. The creation of the anomaly detection system is composed of several sections. In the first section, the data preprocessing and feature selection are introduced. The second section includes semi-supervised machine learning algorithms utilize for binary classification, such as anomaly detection. Moreover, four machine learning algorithms were chosen for this research (Artificial Neural Network - ANN, recurrent neural network LSTM, Isolation Forest - IF, and algorithm OCSVM). The adaptation of the anomaly detection system is achieved by hyperparameter optimization of machine learning algorithms (Random Search - RS, Evolution Algorithm - EA, Tree-structured Parzen Estimator - TPE). Additionally, the objective function of optimization algorithms is defined by multicriteria evaluation such as Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). The last section is focused on the interpretation of detected anomalies. We conducted extensive research, including four machine learning algorithms, three optimization techniques, and multiple data preprocessing techniques. Moreover, we use three datasets to conduct experiments [8-10]. Finally, the research aims to select and test the optimal anomaly detection system, which corresponds to the best combination of techniques and algorithms.

The rest of the article is organized as follows. In Section 2 data preprocessing techniques are described. Section 3 is dedicated to the description of machine learning algorithms that are used for cyber-attacks detection. The optimization algorithms which are used to optimize machine learning algorithms via hyperparameters are described in Section 4. The research methods are described in Section 4. Section 5 shows the results of the experiments. Lastly, Section 6 is dedicated to a discussion of the results.

2. Data preprocessing techniques

The data are vital for machine learning algorithms. Moreover, they fundamentally influence the performance of machine learning algorithms. The process of capturing data is commonly imperfect. Therefore, data are regularly captured in different formats and have missing values. Thus, data preprocessing is a critical part of the machine learning model creation. According to the collected datasets, we recognize three main problems. The missing values in datasets due to flaws in the ICS recording process are defined as the first problem. The different scale of features is identified as the second issue. Moreover, the differences between the scale of the dataset feature can negatively affect their importance and, therefore, may result in weak detection capabilities of the machine learning model. The last main problem for machine learning algorithms is the dataset extensive dimensionality, which increases the computational complexity of the anomaly detection system. Therefore, the creation of a machine learning model can be a very time and resources consuming process.

We choose multiple techniques to solve all described data preprocessing issues. Three techniques were chosen to address missing values. The missing values are replaced by value zero, mean of the feature, or feature median [11]. Another three techniques were chosen to address the features scale problem of data. The data scale is changed according to the interval $\langle -1,1\rangle$, $\langle 0,1\rangle$, or the data are standardized [11]. We made a considerable number of experiments to find out the optimal combination of techniques. Thus, nine final combinations were established. The results are published in section 6.

Furthermore, the last question in this section that needs to be solved is focused on the immersive trend in high dimensional data. ICS has become complex systems that generate high dimensional datasets. Moreover, categorical data amplify this trend. In this case, the one-hot encoder transformation [11] is used. The categorical data are transformed into a binary representation where every unique sample represents one new feature in the dataset. Thus, a new, excessively large dataset is created. We adopted the dimension reduction technique (Principal Component Analysis – PCA [12]) to address the issue.

PCA is based on covariance matrices, which express the interdependence between the described features and their standard deviations. The basic idea of PCA is to reduce the number of features while preserving its original information value. So that the newly created attributes contain higher variability than the original attributes. In practice, the original data are projected into a lower dimension (lower number of features). Thus, the principal components are created by a linear combination of the original features. These new features contain variability of the previous dataset even the new dataset has lower dimensionality [12].

3. Machine learning algorithms

This chapter is dedicated to a description of machine learning algorithms, which are cornerstones of the article. The four popular

algorithms were chosen (ANN, LSTM, IF, OCSVM). All of these algorithms are modified to work in semi-supervised learning mode. This mode of machine learning algorithms is based on a combination of supervised learning and unsupervised learning. The input dataset for the creation of the machine learning model includes only the normal processes of the ICS system. The training dataset cannot be contaminated by malicious behavior like cyber-attacks. Moreover, the created model should represent the normal behavior of ICS, and every deviation from it is classified as an anomaly (potentially cyber-attack). The procedure ensures the detection of unknown cyber-attacks. However, on the other hand, semi-supervised learning has a problem with contamination of training dataset, interpretation of results, and poor detection capabilities. To create a semi-supervised learning model, three datasets used. The training and validation dataset with only one class is used to create a machine learning model. Otherwise, the test dataset includes all classes of a dataset. Thus, the normal operation of the system and cyber-attacks are included. Moreover, the data in the test dataset are completely separated from the training and validation dataset. The evaluation of detection capabilities is provided as a result of the test dataset classification. Furthermore, there is a considerable number who examined the possibility of semi-supervised learning algorithms to detect anomalies. [13-15]

Four popular machine learning algorithms are used to examine the possibility of anomaly detection for ICS technology. One of the article's goals is to find an optimal machine learning algorithm according to evaluation metrics. All machine learning algorithms are used in a semi-supervised learning mode. Moreover, the python machine learning library Scikit-learn [16] to adopt machine learning algorithms IF and OCSVM for the needs of the anomaly detection system. Otherwise, the python library Keras [17] is used to create neuron network and recurrent neuron network LSTM.

One-class Support Vector Machines (OCSVM) algorithm is a modified version of the Support Vector Machines (SVM) algorithm. The modification enables OCSVM to work as a semi-supervised learning algorithm. The deterministic algorithm OCSVM creates a hyperplane to separate anomalies from the rest data. [16,18] Moreover, the OCSVM is implemented with a radial kernel. Therefore, there is only a gamma hyperparameter to tune by optimization algorithms.

Machine learning algorithm Isolation Forest (IF) is a modification of the more known Random Forest algorithm (RF). Moreover, IF works in semi-supervised learning mode. The IF was firstly published by Liu et al. [19]. The authors introduced the semi-supervised learning modification of RF to detect anomalies. IF is based on two assumptions. The first assumption depends on the fact that anomalies are very rarely present in the data. The second assumption depends on the difference in value between the attribute of a normal record and an anomalous record attribute. [16,19] Additionally, four hyperparameters should be optimized. They are the following hyperparameters: maximum number of features, number of samples, number of trees, and contamination.

The artificial neural network and recurrent neural network LSTM are originally supervised learning algorithms. Therefore, their modification has to be introduced. The symmetrical autoencoder architecture is utilized for both algorithms. In the past, autoencoder basic use was mainly to remove noise from input signals. However, the autoencoder can be used to create a normal operation model due to its generalization capabilities. The architecture of the basic symmetrical autoencoder is shown in Fig. 1. The main idea of autoencoders is to create a model with identical outputs with inputs (compression and decompression). The dimension of the input layer is the same as the output layer. There is also the middle layer, which is called "bottleneck". The generalization ability of autoencoder relies on the bottleneck, which has a lower number of neurons than input and output layers.

The detection of anomalies is based on the ANN or LSTM model. Every deviation from the model is classified as an anomaly. The error, which represents a deviation between the model and new data, is

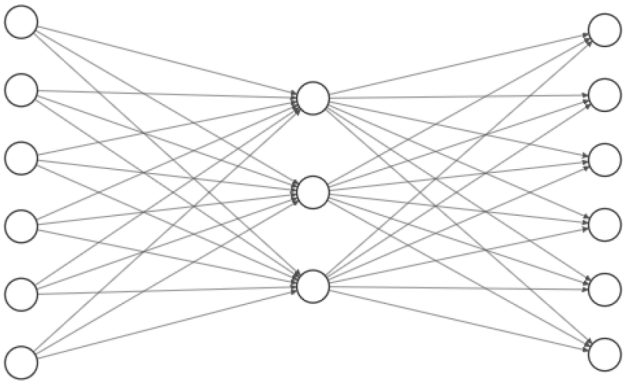

Fig. 1. The architecture of the basic symmetrical autoencoder.

calculated according to Eq. (1), where $x_{i,j}^{mod}$ represents the model for i th feature and j-th data point. The $x_{i,j}^{real}$ represents real data for i th feature and j-th data point.

$$x^{r} = \left| x_{i,j}^{mod} - x_{i,j}^{real} \right| \tag{1}$$

The Mean Squared Error (MSE) formula is used to calculate the mean deviation value for each final data point according to Eq. (2). Moreover, the precision/recall curve is used to calculate the final threshold to distinguish the normal operation of the ICS system from cyber-attacks.

$$MSE = \frac{1}{2}\sum_{i=1}^{n} \left(x_i^{r}\right)^2 \tag{2}$$

The first machine learning algorithm with autoencoder architecture is ANN. ANN function are based on the control center of the nervous system in biological organisms. Moreover, the control center consists of neurons, which are unique cells destined for the storage and transmission of information necessary for the proper functioning of the organism. Rosenblatt published the basis of the ANN in his publication [20], where he introduced the definition of perception. It is a simple example of a feedforward neural network with a single neuron. Furthermore, the shortcomings of perceptron were solved by multilayer perceptron, which is still used today. [17]

The second machine learning algorithm with autoencoder architecture is LSTM. We use the LSTM algorithm presented by Hochreiter and Schmidhuber in the publication. [21] LSTM is a machine learning algorithm that falls into the subgroup of recurrent neural networks. These algorithms work with sequential data, where also the data arrangement in time plays an important role. LSTM is a well-known algorithm that is often used for text or audio classification purposes. Unlike the artificial neural network, the LSTM contains a modified neural cell containing three so-called gateways. These gateways allow to store and transmit information from previous records. [21,22]

The optimization process of ANN and LSTM, including the following hyperparameters. Number of neurons, number of layers, number of neurons for "bottleneck", number of epochs, batch size, "dropout" size, activation function, the optimizer for algorithm training and learning rate. The only difference between hyperparameters for ANN and LSTM is "size of recurrent dropout," used for LSTM model training.

4. Optimization techniques

Optimization is a process where several techniques and procedures are used to find the best, i.e., optimal solution. Additionally, optimization algorithms utilize an iterative process with feedback to find the minimum or maximum objective function according to the assignment. The objective function (OF) consists of a score calculated by the multi-criteria evaluation algorithm TOPSIS for the anomaly detection system. Tzeng and Huang published TOPSIS in the publication [23]. The main

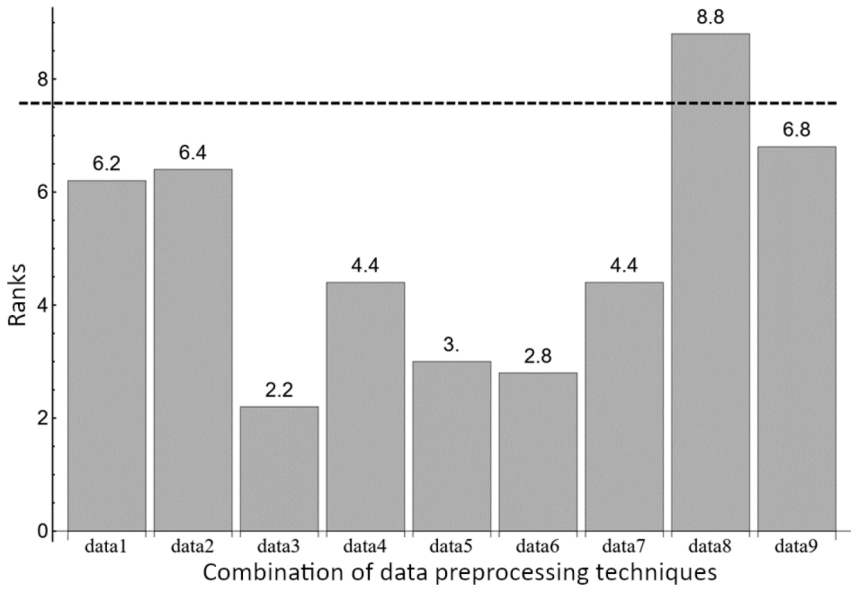

Fig. 2. Friedman test, including Nemenyi critical distance for various combinations of data preprocessing techniques ANN - CA1_1.

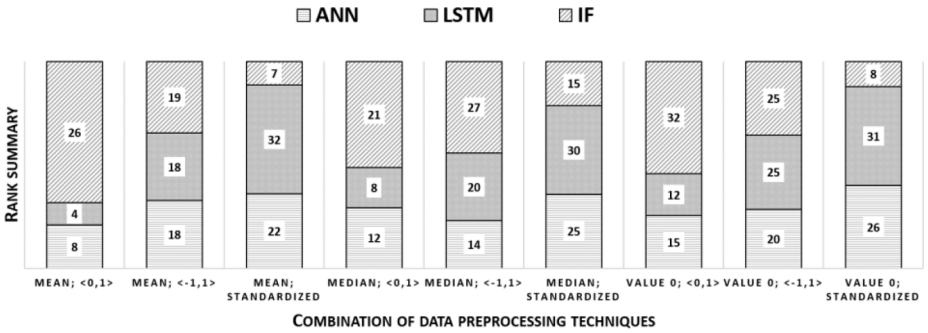

Fig. 3. Summarized rank results of the Friedman test for various combinations of data preprocessing techniques.

idea of TOPSIS is to choose the best variant from the set of all variants. The selection of the best variant is conducted according to five metrics, which are detailed described in Section 5. Thus, the best configuration of hyperparameters is calculated for all machine learning algorithms based on optimization algorithms.

The three well-known optimization algorithms were chosen (GA, RS, TPE). The best-suited optimization algorithm is chosen based on experiments for the anomaly detection system. The results are presented in chapter 6. The first described optimization algorithm is RS. This is a considerably simple optimization algorithm. Moreover, RS is usually the first selection for the optimization of machine learning algorithms. This optimization algorithm is based on the "Grid search" (GS) algorithm. GS systematically calculating all possible combinations of hyperparameter. The downside of the algorithm is its enormous time and computation demands. The RS randomly select hyperparameters from all search space, unlike GS. Therefore, RS more likely selects the more effective solution.

Genetic algorithms (GA) is the second optimization algorithm that is adopted for the hyperparameter optimization task. GA is a robust heuristic-based search algorithm based on Darwin's theory of evolution, published by Holland in publication [24]. The basic idea of GA is based on the assumption that only the most capable individuals will reproduce. Therefore, their attributes will be able to pass on to future generations. In the beginning, the population is created. Moreover, each individual consists of the sets of traits that are hyperparameters in this case. Each individual is evaluated using the OF. Moreover, better individuals are created through the crossover and mutation in each generation. The best solution is obtained in the last generation, where the final individuals converge into one best solution.

TPE is the last selected optimization algorithm for the experiments. TPE is a Bayesian-based optimization algorithm that is based on the Gaussian process. Bergstra et al. firstly introduce TPE in their publication [25]. Additionally, TPE is a common optimization algorithm for hyperparameter optimization tasks. The algorithm is based on Sequential model-based optimization (SMBO). The model is used as a sequential procedure to implement the gradual modification. This iterative process based on OF updated the probability values of the model. The process leads to more accurate results and, therefore, to the optimal solution in the last iteration.

5. Research methods

The core of the experiments is based on three datasets. The dataset consists of recorded ICS communication under cyber-attacks. Lemay and Fernandez introduced the first dataset [8]. Dataset 1 consists of a several-hour record of the ICS network communication. Moreover, the

Table 1
Data preprocessing techniques.

Missing values	Data scaling
mean	<0,1>
median	<−1,1>
value 0	standardized

dataset contains data in pcap format. Thus, the data conversion was needed into CSV format. The data were generated using a series of electrical distribution network simulations. This dataset is primarily used for data preprocessing experiments due to the low number of cyber-attacks in the dataset. Additionally, four cyber-attacks were chosen CA1_1, CA1_2, CA1_3, CA1_4.

The second dataset was captured from a real-world ICS system created by "University of Technology and Design" in Singapore. [9] Dataset 2 represent recorded data from the water treatment system, which produces approximately 20 l of filtered water per minute. The system called SWaT (Secure Water Treatment) consists of several sensors, actuators, PLC, HMI, SCADA workstations. Additionally, six cyber-attacks were chosen CA2_1, CA2_2, CA,2_3, CA2_4, CA2_5, CA2_6 for creation of the anomaly detection system (optimization process), and three cyber-attacks CA2_7, CA2_8, CA,2_9 for anomaly detection system evaluation.

The third dataset was developed at Mississippi State University. [10] The data in dataset 3 was captured from the gas pipeline system. Moreover, the ICS system consists of sensors and actuators which regulate the pipe pressure. Additionally, six cyber-attacks were chosen CA3_1, CA3_2, CA,3_3, CA3_4, CA3_5, CA3_6 for the creation of the anomaly detection system (optimization process), and three cyber-attacks CA3_7, CA3_8, CA,3_9 for anomaly detection system evaluation.

The five metrics were chosen for the evaluation of machine learning models. Moreover, this set includes well-known and robust metrics that suits well ICS environment. Each machine learning model is evaluated using a binary confusion matrix that expresses the relationship between predicted classes and real classes. The first expresses the normal operation of the system (negative class), the second a cyber-attack on the system (positive class).

Each of the three datasets for experiments is divided into a training dataset and a testing dataset. Training dataset always contains only data with negative class. Therefore, only data without cyber-attacks are included in training datasets. The testing dataset contains data of both classes. Moreover, data of normal operation of ICS and cyber-attacks are included. However, each training dataset is unique. None of the data is shared between training and testing datasets. This separation guarantees the validity of results where none of the data from the testing dataset is included in training the machine learning model.

All machine learning algorithms are implemented in a semi-supervised (one-class classification) manner. Therefore, only data with one class is included in the training dataset. Moreover, the training dataset must not be contaminated by cyber-attacks. There is a common approach to evaluate classification models. The k-fold cross-validation technique enables the estimation of how well the classification model predicts new and unseen data. However, this technique cannot be used in the case of one-class classification because the training dataset cannot contain different classes (cyber-attacks) except one (normal operation of the ICS system). Therefore, a different approach was established. Each testing dataset consists of completely unique data of both classes (normal operations and cyber-attacks). Therefore, each testing dataset is contained by unseen data. This applies even between individual testing datasets.

The confusion matrix is a fundamental basis for the calculation of metrics. Additionally, the metrics are as follows:

- **M_{MCC}** - (Matthews Correlation Coefficient) expresses all aspects of the confusion matrix. Moreover, the metric is resilient against the usage of unbalanced datasets.
- **M_{F1}** - (F1 score) is the second popular metric. It is a robust metric for model evaluation in the case of a binary confusion matrix. However, unlike M_{MCC}, it does not consider the True negative class within the confusion matrix. Therefore, it is more focused on the evaluation of model classification errors.
- **M_{Prec} -** (precision) This metric was chosen because it takes into account the false-positive (FP) classification in the calculation of the metric. Moreover, the classification represents falsely classified cyber-attacks that are critically harmful to ICS.
- **M_{FPR}** - (False positive rate (FPR)) M_{FPR} expresses positive cases that are identified as a false class. Thus, cases when normal and harmless communication in the computer network is evaluated as dangerous. False alarms are a major problem for ICS. The metric is focused on monitoring false alarms. Therefore, M_{FPR} is the most important metric.
- **Time -** Time is the criterion that expresses the time required to predict and classify the test dataset by the model. Furthermore, time is a parameter that expresses the computational complexity of each classification model.

The research can be divided into three main parts. An extensive number of experiments were conducted to select the best version of the anomaly detection system. The first part focuses on selecting the best-suited preprocessing techniques for each machine learning algorithm based on detection performance (evaluation metrics). We tested nine combinations of preprocessing methods. Three techniques for missing values problem and three techniques for data scaling problem shown in Table. 1. It is important to note that the preprocessing techniques were utilized only in the case of numerical features. Moreover, preprocessing techniques experiments were based on four cyber-attacks in dataset 1. The results were considered optimal settings and, therefore, are used in further experiments.

The second part of the research is focused on the computationally excessive demanding process of hyperparameter optimization in the machine learning field. The machine learning algorithms were examined in conjunction with three optimization algorithms for three datasets. The main goal of this section is to create an anomaly detection system that has acceptable detection capabilities and a low M_{FPR}. We examined all possible combinations of the algorithms and datasets. The execution of experiments was a highly time and computational demanding task. Therefore, the advanced infrastructure for the calculation of complex tasks had to be used. The MetaCentrum supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) was used to complete all tasks. Furthermore, the OF is calculated by the TOPSIS algorithm according to the metrics. However, there is a necessity to set up weights for all metrics. A Fuller's triangle was created for these metrics, where the dependencies between the individual metrics were calculated by pairwise comparison. The weights are shown in Table. 2.

Each hyperparameter combination of the machine learning algorithm, datasets, and optimization algorithms is calculated in the metacentrum for 300 h. Also, each combination is processed tenfold due to the stochastic nature of machine learning algorithms. The final combination of hyperparameters will be collected from the metacentrum results. Moreover, every combination of hyperparameters will be evaluated via the metrics one hundred times. The Friedman test [26] was implemented to compare all results via metrics. Milton Friedman developed this non-parametric statistical test to detect differences across a considerable number of samples. The second evaluation of the results will be focused on one metric. M_{FPR} is the most important metric for ICS. Therefore, a deep analysis of M_{FPR} is appropriate. The evaluation of the anomaly detection system is based on three cyber-attacks from dataset 2 (CA2_7, CA2_8, CA2_9) and three cyber-attacks from dataset 3 (CA3_7, CA3_8, CA3_9) are used. These cyber-attacks are completely separated from the optimization process. Thus, none of them were used in the first and second part of the research. Additionally, the Friedman test and

Table 2
Weights for metrics.

Metrics/ weights	M_{F1}	M_{MCC}	M_{Prec}	Time	M_{FPR}
Weights	0.12	0.16	0.17	0.2	0.35

M_{FPR} metric comparison are used for evaluation. The third part of the research dealing with the interpretation of results based on the importance of dataset features.

6. Results

The first part of the research focuses on selecting the most suitable preprocessing techniques for each machine learning algorithm. All the combinations of data preprocessing techniques and machine learning stochastic algorithms are tested for four cyber-attacks in dataset 1. Every combination is one hundredfold repeated due to the stochastic nature of selected machine learning algorithms. Differently, there was no need for repetition of experiments in the case of the deterministic OCSVM algorithm. The results of all experiments are shown in Table. 3. The arithmetic mean of all values represents the stochastic algorithms for each of the metrics in relation to preprocessing techniques combinations. We also take into account the aspect of machine learning algorithms and different cyber-attacks.

There are nine combinations of preprocessing techniques in the Table. 3. Data1 stand for combination (mean, <0,1>), data2 stands for combination (mean, <−1,1>), data3 stands for combination (mean, standardized), data4 stands for combination (median, <0,1>) etc. Every machine learning algorithm evaluates each combination via five metrics (M_{F1}, M_{MCC}, M_{Prec}, Time, M_{FPR}). The best preprocessing techniques combination for each machine learning algorithm is selected according to the performance of all metrics. Moreover, the stochastic algorithms (ANN, LSTM, IF) are evaluated by Friedman non-parametric statistical test via all five metrics.

The deterministic algorithm OCSVM is the first to evaluate. There is only one result for each experiment (combination) in the case of OCSVM. The mean value of the feature is the best-suited technique to handle missing values. Additionally, the scale ⟨-1,1⟩ of the feature is the optimal technique to handle scaling issues for the OCSVM algorithm. The results for the best combination are highlighted in Table. 3.

The results of the rest machine learning algorithms are evaluated by the Friedman test due to a considerable number of machine learning models for each combination (one hundred per each). Therefore, the Friedman test was calculated for each combination of three machine learning algorithms, six preprocessing techniques, and four cyber-attacks. The demonstrational results for ANN are shown in Fig. 2 in the case of CA1_1.

As shown in Fig. 2, all combinations are ranked where data3 is the best combination and data8 is the worst combination. Moreover, the dashed line represents the Nemenyi test result, where critical distance is calculated between data.

Over ten thousand machine learning models were created. Therefore, the summarization of the result was implemented. The Friedman test ranks were overall summarized of various data preprocessing techniques combinations for four cyber-attacks. Moreover, the best combination was ranked as one, and the worst combination was ranked as nine. A summary of all results over four cyber-attacks can be seen in Fig. 3.

As shown in Fig. 3, there is no best combination of data preprocessing techniques for all machine learning algorithms. Therefore, the individual approach should be adopted for every machine learning algorithm. The best combination of the ANN techniques is the mean of feature values in case of missing values. The scaling of the dataset should be in the range ⟨0,1⟩. Moreover, the same best combination of the techniques can be seen in the case of the LSTM algorithm. The possible explanation is the same origin of algorithms. Furthermore, the best preprocessing technique to address the missing values in the dataset includes the median in the case of IF. IF has the best results if the standardization technique is used to change the dataset scale. Additionally, the overall results indicate ANN as the best representative in this experiment. The best representatives of preprocessing techniques combination are also highlighted in Table. 3. Finally, these preprocessing techniques are used for further experiments.

The second part of the research is focused on a hyperparameter optimization of machine learning algorithms. The first tested machine learning algorithm is OCSVM. Due to its deterministic nature and only one hyperparameter to optimize the limited number of machine learning models can be created. The gamma parameter is set to a range of values from interval ⟨0,1⟩ with step 0.05. Thus, twenty models were created with different gamma parameters for every cyber-attack in each dataset. All results are shown in Fig. 4.

Multiple machine learning models with variable gamma parameters were created for each cyber-attack. Therefore, the box chart was used to summarize the results. The summary of the M_{FPR} results for OCSVM can be seen in Fig. 4. Moreover, the M_{FPR} is the most important metric for ICS. Therefore, the M_{FPR} should be the main criteria for the evaluation of every anomaly detection system, which is focused on the ICS environment. The presented results are insufficient in every case. Moreover, similar poor results were achieved in the case of the Time metric. Therefore, we decided to exclude the algorithm from further experiments.

The research was primarily focused on hyperparameter optimization of ANN, LSTM, and IF algorithms according to the anomaly detection performance in case of cyber-attacks. Moreover, an extensive number of experiments were executed. Each machine learning algorithm was optimized according to RS, GA, TPE for dataset 1 (four cyber-attacks), dataset 2 (six cyber-attacks), and dataset 3 (six cyber-attacks). Thus, 27 combinations were established. Moreover, each combination was executed tenfold due to the stochastic(probabilistic) nature of the machine learning algorithms. Therefore, 270 individual optimization runs had to be done. This time consuming and resources exhausting task was carried out by the MetaCentrum. More than 81,000 h of experiments were computed. Such an extensive computational task could not be realistically solved in real-time without outsourced resources.

The optimal hyperparameter combination was calculated for each machine learning algorithm in the case of different datasets. Moreover, the optimal setup of machine learning algorithms was evaluated by six cyber-attacks (dataset 2 - CA2_7, CA2_8, CA2_9, and dataset 3 - CA3_7, CA3_8, CA3_9). The cyber-attacks were completely separated from the optimization process and, therefore, were not influencing optimization results. The final results were compared according to the Friedman test. Thus, each representative was ranked based on five metrics for each individual cyber-attack. The summed results can be seen in Fig. 5.

Fig. 5 shows the summarized results for each dataset (cyber-attacks) and all optimized machine learning algorithms. The machine learning algorithm IF optimized by GA has the best overall anomaly detection performance of all representatives.

Moreover, the overall results for each optimized algorithm are shown in Table. 4. Each optimized algorithm is tested on six cyber-attacks. The performance is evaluated via five metrics. The results for the best variant (IF - GA) are highlighted in the table. We can see a highly promising performance in terms of M_{FPR}. The performance of the rest of the metrics is modest. However, despite the performance, the cyber-attacks are classified and identified.

Evaluating the most important metric (M_{FPR}) is required to develop a reliable anomaly detection system in an ICS environment. Therefore, the summarization graphs were introduced. Moreover, we use multiple box charts to present results. Every optimized machine learning algorithm created 100 models that were evaluated via six cyber-attacks. The results for dataset 2 are shown in Fig. 6, and results for dataset 3 are shown in Fig. 7.

As shown in Fig. 6, algorithm IF optimized by GA has the best results in all instances. Additionally, the margin between the best algorithm and the second-best algorithm is enormous. In the case of dataset 3 (shown in Fig. 7), the results are not unambiguous. However, we can identify the two best-suited representatives. The first is the algorithm IF optimized by GA, and the second is the algorithm IF optimized by TPE. Both of the algorithms have identical results for all cyber-attacks. Moreover, they

Table 3
Overall results of preprocessing techniques combination.

		CA1_1				CA1_2					CA1_3					CA1_4					
	Combinations	**M_{F1}**	**M_{MCC}**	**M_{Prec}**	**M_{FPR}**	**Time**	**M_{F1}**	**M_{MCC}**	**M_{Prec}**	**M_{FPR}**	**Time**	**M_{F1}**	**M_{MCC}**	**M_{Prec}**	**M_{FPR}**	**Time**	**M_{F1}**	**M_{MCC}**	**M_{Prec}**	**M_{FPR}**	**Time**
ANN	**Data1**	0.67	0.65	0.75	0.02	0.01	0.82	0.62	0.93	0.11	0.01	0.59	0.59	0.68	0.01	0.04	0.05	0.06	0.09	0.01	0.06
	Data2	0.67	0.65	0.76	0.02	0.01	0.76	0.45	0.86	0.22	0.01	0.59	0.59	0.64	0.01	0.04	0.02	0.02	0.04	0.01	0.07
	Data3	0.73	0.71	0.80	0.02	0.01	0.76	0.38	0.80	0.32	0.01	0.28	0.26	0.29	0.02	0.04	0.00	0.01	0.00	0.06	0.07
	Data4	0.72	0.70	0.80	0.02	0.01	0.78	0.50	0.88	0.19	0.01	0.55	0.55	0.64	0.01	0.04	0.10	0.11	0.17	0.00	0.07
	Data5	0.68	0.67	0.81	0.01	0.01	0.77	0.46	0.86	0.21	0.01	0.41	0.41	0.46	0.01	0.04	0.05	0.05	0.07	0.01	0.06
	Data6	0.74	0.72	0.82	0.01	0.01	0.77	0.42	0.83	0.28	0.01	0.26	0.24	0.27	0.02	0.05	0.00	0.01	0.00	0.06	0.07
	Data7	0.72	0.71	0.79	0.02	0.01	0.79	0.51	0.87	0.20	0.01	0.55	0.55	0.62	0.01	0.05	0.11	0.12	0.19	0.01	0.07
	Data8	0.47	0.44	0.56	0.03	0.01	0.77	0.48	0.88	0.19	0.01	0.52	0.51	0.56	0.01	0.04	0.02	0.02	0.04	0.01	0.07
	Data9	0.62	0.60	0.74	0.02	0.01	0.76	0.43	0.83	0.26	0.01	0.18	0.16	0.19	0.02	0.04	0.00	0.01	0.00	0.06	0.06
LSTM	**Data1**	0.89	0.88	0.91	0.01	0.01	0.96	0.90	0.98	0.03	0.02	0.72	0.72	0.83	0.00	0.35	0.28	0.28	0.34	0.00	0.07
	Data2	0.73	0.71	0.77	0.02	0.01	0.91	0.75	0.93	0.13	0.02	0.42	0.40	0.46	0.02	0.37	0.07	0.07	0.11	0.03	0.07
	Data3	0.42	0.36	0.42	0.05	0.01	0.76	0.30	0.77	0.46	0.02	0.06	0.04	0.06	0.02	0.39	0.00	0.01	0.00	0.16	0.07
	Data4	0.87	0.86	0.90	0.01	0.01	0.95	0.89	0.98	0.04	0.02	0.72	0.72	0.81	0.01	0.35	0.23	0.23	0.28	0.01	0.07
	Data5	0.71	0.69	0.74	0.03	0.01	0.92	0.77	0.93	0.13	0.02	0.49	0.48	0.54	0.02	0.37	0.05	0.04	0.06	0.04	0.07
	Data6	0.44	0.38	0.44	0.05	0.01	0.75	0.26	0.76	0.48	0.02	0.04	0.02	0.05	0.02	0.35	0.00	0.01	0.00	0.17	0.07
	Data7	0.82	0.81	0.89	0.01	0.01	0.94	0.84	0.97	0.06	0.02	0.63	0.63	0.75	0.01	0.35	0.22	0.22	0.26	0.01	0.07
	Data8	0.18	0.17	0.33	0.02	0.01	0.80	0.48	0.85	0.26	0.02	0.15	0.13	0.16	0.02	0.35	0.05	0.05	0.08	0.02	0.07
	Data9	0.42	0.37	0.43	0.05	0.01	0.73	0.21	0.74	0.52	0.02	0.08	0.06	0.08	0.02	0.35	0.00	0.01	0.00	0.11	0.07
IF	**Data1**	0.31	0.24	0.22	0.16	0.07	0.62	0.42	0.87	0.09	0.08	0.17	0.18	0.11	0.08	0.12	0.00	0.00	0.00	0.08	0.43
	Data2	0.39	0.34	0.30	0.13	0.07	0.75	0.53	0.94	0.08	0.08	0.09	0.07	0.06	0.08	0.12	0.00	0.00	0.00	0.08	0.43
	Data3	0.39	0.34	0.30	0.13	0.07	0.74	0.55	0.96	0.04	0.08	0.15	0.13	0.12	0.04	0.12	0.00	0.00	0.00	0.05	0.43
	Data4	0.32	0.26	0.24	0.16	0.07	0.56	0.36	0.83	0.09	0.08	0.18	0.19	0.12	0.08	0.12	0.00	0.00	0.00	0.08	0.43
	Data5	0.38	0.32	0.29	0.13	0.07	0.74	0.52	0.93	0.08	0.08	0.06	0.04	0.04	0.07	0.12	0.00	0.00	0.00	0.07	0.44
	Data6	0.38	0.33	0.30	0.14	0.07	0.75	0.56	0.96	0.05	0.08	0.15	0.13	0.12	0.05	0.13	0.01	0.01	0.00	0.05	0.44
	Data7	0.29	0.23	0.21	0.17	0.07	0.52	0.33	0.82	0.09	0.08	0.18	0.18	0.11	0.09	0.13	0.00	0.00	0.00	0.09	0.45
	Data8	0.41	0.36	0.31	0.14	0.07	0.77	0.56	0.94	0.08	0.08	0.17	0.16	0.11	0.07	0.13	0.00	0.01	0.00	0.08	0.45
	Data9	0.45	0.41	0.39	0.10	0.07	0.75	0.55	0.95	0.05	0.08	0.18	0.17	0.14	0.04	0.12	0.00	0.00	0.00	0.05	0.43
OCSVM	**Data1**	0.20	0.11	0.11	0.52	4.08	0.71	0.25	0.75	0.42	5.23	0.08	0.11	0.04	0.41	9.42	0.00	0.01	0.00	0.41	31.36
	Data2	0.21	0.19	0.12	0.67	4.14	0.72	0.27	0.75	0.41	5.36	0.08	0.11	0.04	0.40	9.58	0.00	0.03	0.00	0.41	32.60
	Data3	0.17	0.09	0.09	0.90	4.00	0.74	0.31	0.77	0.39	5.18	0.09	0.14	0.05	0.40	9.31	0.00	0.04	0.00	0.40	31.15
	Data4	0.18	0.08	0.10	0.58	4.10	0.71	0.25	0.75	0.42	5.82	0.08	0.11	0.04	0.41	9.54	0.00	0.01	0.00	0.41	32.53
	Data5	0.21	0.19	0.12	0.67	3.99	0.72	0.27	0.76	0.41	5.23	0.08	0.11	0.04	0.40	9.26	0.00	0.03	0.00	0.41	31.33
	Data6	0.17	0.09	0.09	0.90	4.01	0.74	0.32	0.77	0.39	5.25	0.09	0.14	0.05	0.40	9.27	0.00	0.04	0.00	0.40	31.32
	Data7	0.16	0.01	0.09	1.00	3.98	0.71	0.26	0.75	0.42	5.22	0.08	0.11	0.04	0.41	9.27	0.00	0.01	0.00	0.41	31.22
	Data8	0.01	0.00	0.01	0.09	4.09	0.72	0.27	0.75	0.41	5.35	0.08	0.11	0.04	0.40	9.66	0.00	0.02	0.00	0.41	32.08
	Data9	0.01	0.00	0.01	0.09	3.90	0.73	0.30	0.76	0.40	5.06	0.09	0.14	0.05	0.40	9.11	0.01	0.04	0.00	0.40	30.57

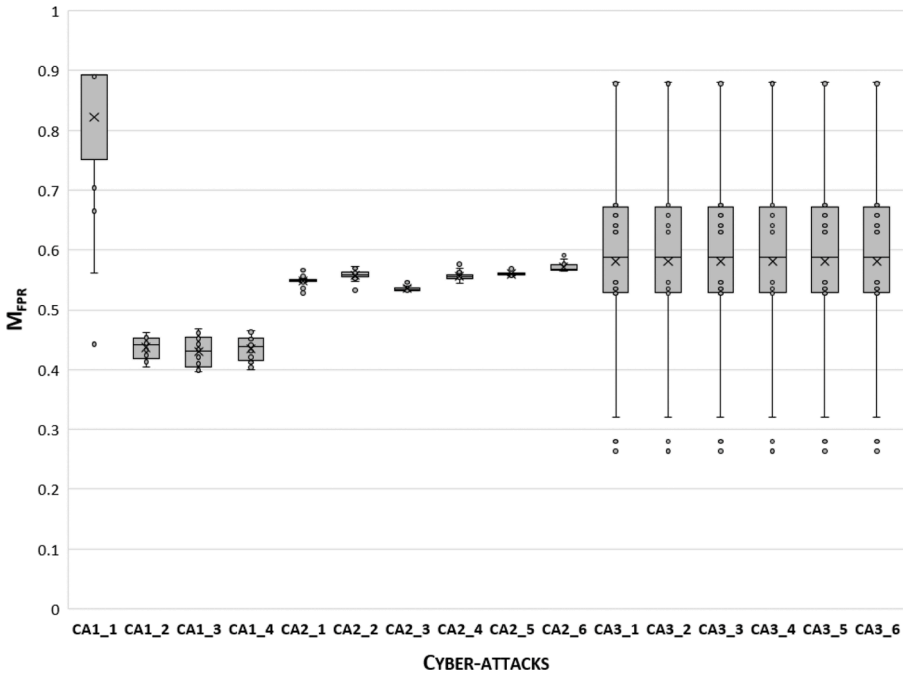

Fig. 4. Results summary OCSVM for different gamma parameters.

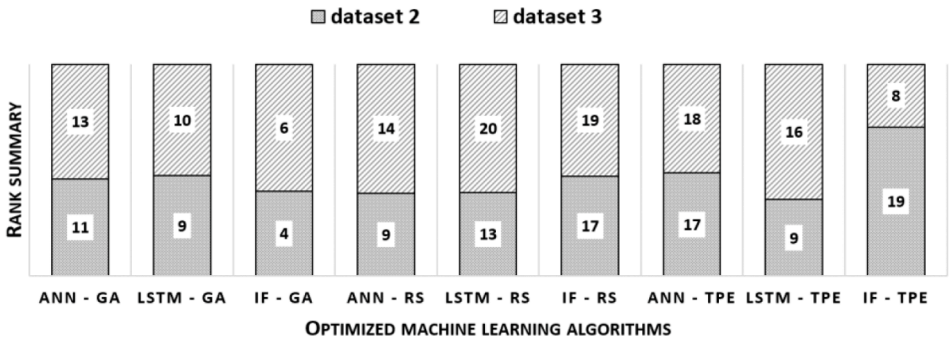

Fig. 5. Summarized rank results of the Friedman test for various combinations of machine learning hyperparameter obtained through the optimization process.

achieve the best possible results were almost all extremely low values of M_{FPR} metric.

Lastly, the possibility of the anomaly interpretation is investigated. The first step in interpreting the results is to gather information related to the model prediction. However, up to our knowledge, there are no possibilities of result interpretation in the case of IF. Therefore, a new approach was needed to implement. The second machine learning algorithm has to be implemented in parallel. We choose the Random Forest (RF) algorithm due to its exceptional interpretation abilities.

The IF algorithm performs the identification of anomalies. If there is no anomaly detected, then the prediction continues without change. Otherwise, the detection output is processed by RF to calculate the importance of different features (principal components) due to a single data point. The binary selection is used to separate important principal components from unimportant principal components (can be seen in Fig. 8). Afterward, the reverse method can be used to select the most important original attributes within the main component. Thus, the most relevant features are required for the detected anomaly. Additionally, the results are used to detect the anomaly and potentially cyber-attack to improve the reaction and mitigation measures.

7. Conclusion

The article was focused on the protection of ICS systems, which are becoming an essential part of modern society due to digitization and Industry 4.0. Therefore, there is a necessity for a reliable cyber-security solution for ICS. We introduced a comprehensive system for anomaly detection based on machine learning algorithms. Moreover, the anomaly section system was designed to solve the following issues:

- Detection of unknown cyberattacks,
- Scalability of the anomaly detection system,
- The adaptability of the anomaly detection system,
- High false alarm rate,
- High computational complexity of the anomaly detection system,
- Interpretation of cyberattacks.

The problem of detection of unknown cyber-attacks was addressed at the beginning of the anomaly detection system creation. All machine learning algorithms were implemented in semi-supervised learning mode. This approach guarantees the ability to detect unknown cyber-attacks. Moreover, the detection capabilities were proven by the detection of six evaluation cyber-attacks, which were not part of the training dataset for machine learning model creation. Hence, the

Table 4
Overall results of the evaluation process for optimized algorithms.

		GA			RS			TPE		
CA results		**ANN**	**LSTM**	**IF**	**ANN**	**LSTM**	**IF**	**ANN**	**LSTM**	**IF**
CA2_7	$\mathbf{M_{F1}}$	0.58	0.58	0.06	0.58	0.56	0.28	0.57	0.59	0.27
	$\mathbf{M_{MCC}}$	0.01	0.01	0.03	0.00	−0.02	0.00	−0.01	0.04	0.00
	$\mathbf{M_{Prec}}$	0.58	0.58	0.70	0.58	0.57	0.58	0.58	0.60	0.58
	$\mathbf{M_{FPR}}$	0.57	0.56	0.04	0.57	0.57	0.19	0.58	0.54	0.17
	Time	3.30	7.13	5.52	7.34	7.59	12.07	4.96	12.51	33.55
CA2_8	$\mathbf{M_{F1}}$	0.48	0.49	0.08	0.51	0.48	0.24	0.47	0.51	0.23
	$\mathbf{M_{MCC}}$	0.01	0.02	0.07	0.06	0.01	−0.01	−0.02	0.06	−0.01
	$\mathbf{M_{Prec}}$	0.49	0.49	0.77	0.51	0.49	0.48	0.47	0.51	0.48
	$\mathbf{M_{FPR}}$	0.46	0.47	0.05	0.45	0.47	0.16	0.48	0.44	0.15
	Time	3.51	7.81	5.53	7.32	8.19	12.11	4.99	12.47	33.38
CA2_9	$\mathbf{M_{F1}}$	0.37	0.40	0.06	0.39	0.40	0.23	0.39	0.39	0.20
	$\mathbf{M_{MCC}}$	−0.03	0.00	0.05	−0.01	0.01	−0.02	−0.01	0.00	−0.03
	$\mathbf{M_{Prec}}$	0.38	0.40	0.68	0.39	0.40	0.38	0.39	0.39	0.37
	$\mathbf{M_{FPR}}$	0.40	0.39	0.04	0.39	0.38	0.17	0.39	0.39	0.16
	Time	3.47	6.97	5.52	7.30	7.46	12.21	4.94	12.56	33.40
CA3_7	$\mathbf{M_{F1}}$	0.51	0.51	0.25	0.51	0.50	0.48	0.51	0.50	0.24
	$\mathbf{M_{MCC}}$	0.33	0.33	0.32	0.32	0.32	0.28	0.32	0.31	0.32
	$\mathbf{M_{Prec}}$	0.54	0.51	1.00	0.52	0.51	0.48	0.52	0.50	1.00
	$\mathbf{M_{FPR}}$	0.17	0.18	0.00	0.18	0.18	0.20	0.18	0.19	0.00
	Time	0.29	1.35	0.53	0.31	0.76	0.37	0.24	1.41	0.43
CA3_8	$\mathbf{M_{F1}}$	0.01	0.23	0.03	0.01	0.02	0.11	0.01	0.16	0.03
	$\mathbf{M_{MCC}}$	−0.38	−0.04	0.10	−0.33	−0.27	−0.14	−0.35	−0.10	0.09
	$\mathbf{M_{Prec}}$	0.01	0.24	1.00	0.01	0.03	0.15	0.01	0.17	1.00
	$\mathbf{M_{FPR}}$	0.39	0.25	0.00	0.32	0.24	0.20	0.34	0.23	0.00
	Time	0.11	0.16	0.56	0.12	0.10	0.39	0.10	0.16	0.47
CA3_9	$\mathbf{M_{F1}}$	0.27	0.41	0.37	0.27	0.25	0.27	0.27	0.34	0.37
	$\mathbf{M_{MCC}}$	0.07	0.19	0.42	0.06	−0.01	0.03	0.05	0.11	0.42
	$\mathbf{M_{Prec}}$	0.35	0.42	1.00	0.34	0.28	0.32	0.34	0.37	1.00
	$\mathbf{M_{FPR}}$	0.17	0.21	0.00	0.18	0.24	0.20	0.18	0.22	0.00
	Time	0.11	0.16	0.56	0.12	0.10	0.40	0.09	0.15	0.45

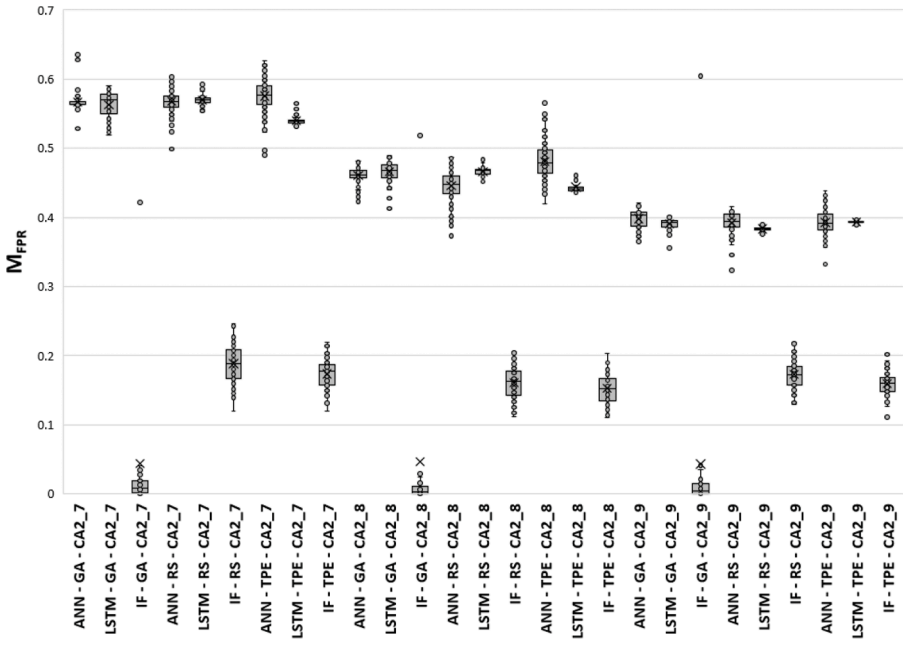

Fig. 6. Summarized M_{FPR} metric for various combinations of machine learning hyperparameter obtained through the optimization process - dataset 2.

detection of unknown cyber-attacks was proven.

The feature transformation ensured the scalability of the detection system via the PCA algorithm. This procedure enables the anomaly detection system to process theoretically very large (high dimensional) ICS datasets. The dataset can theoretically have an unlimited number of attributes, which are reduced to an acceptable level. The process of feature space reduction was described in the article. Moreover, the preprocessing techniques were investigated in terms of anomaly detection performance. Thus, the deployment of the anomaly detection system in a complex (high dimensional) ICS environment was ensured.

The optimization algorithms were implemented for the anomaly detection system to ensure the optimal performance of machine learning algorithms. The approach increases effectivity of the detection capabilities of the system via hyperparameters. Thus, each machine learning

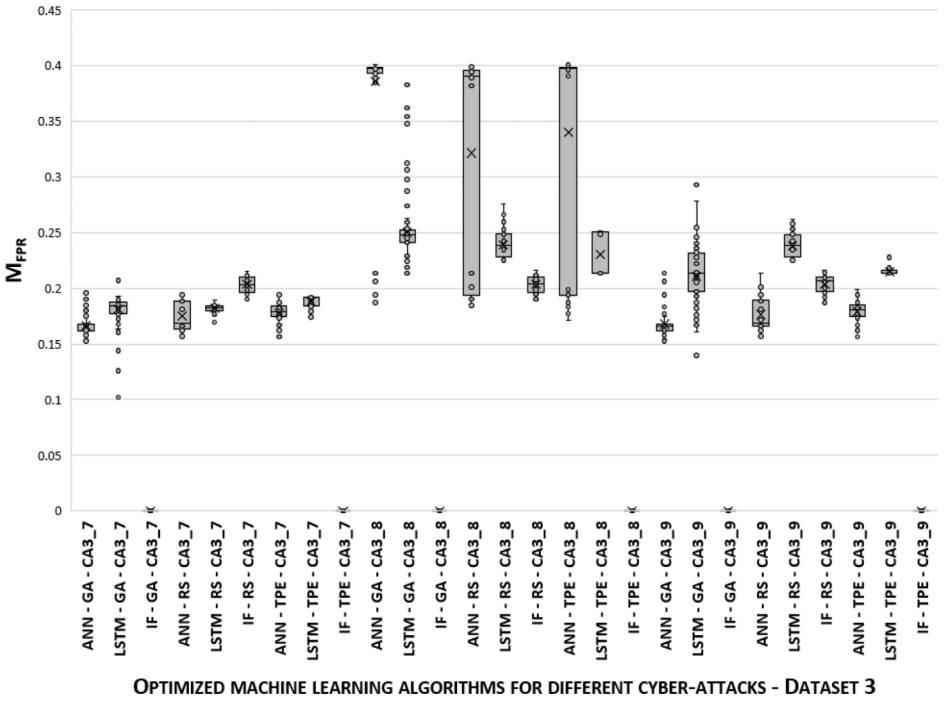

Fig. 7. Summarized M_{FPR} metric for various combinations of machine learning hyperparameter obtained through the optimization process - dataset 3.

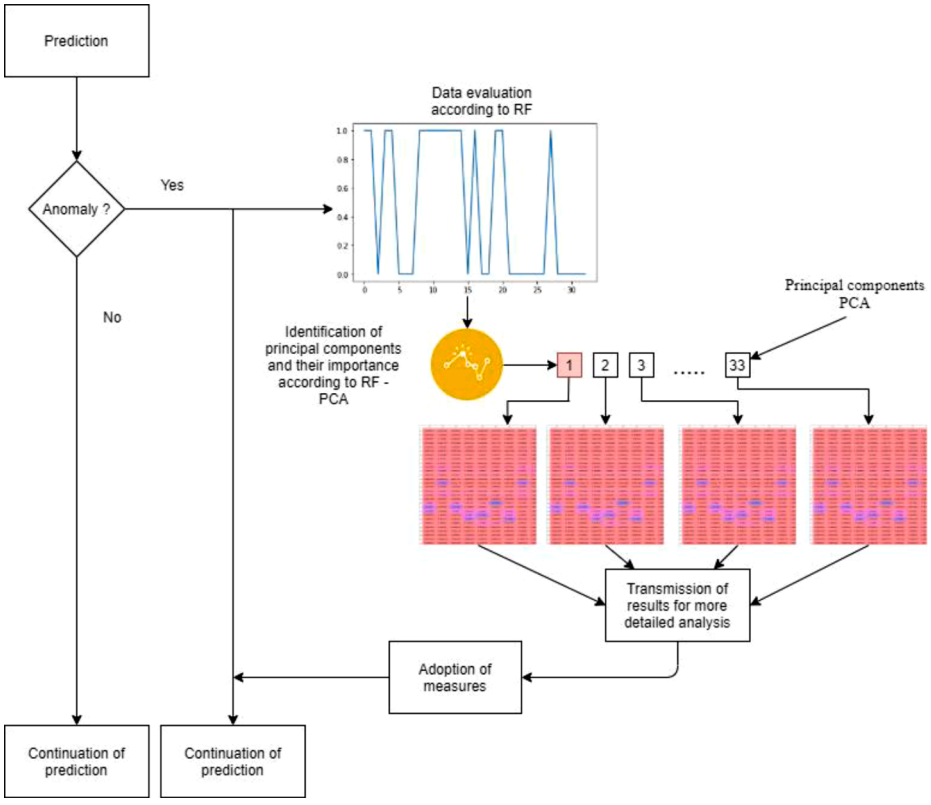

Fig. 8. Interpretation of IF algorithm results.

algorithm is adapted according to the ICS dataset. Moreover, there is no best hyperparameter setup for all ICS system. Therefore, hyperparameter tuning is necessary for each ICS system separately. Moreover, a total of 270 experiments were performed. Where each experiment lasting 300 h. In total, it was 81,000 h of machine time consumed to carry out the experiments.

Every hyperparameter combinations were evaluated via six cyber-attacks. The Friedman test was implemented to evaluate every machine learning algorithm via five metrics. Moreover, the M_{FPR} metric comparison was created to find the most suitable combination of optimization algorithm and machine learning algorithm. There is one best-suited solution in the form of machine learning algorithm IF, which is

optimized by GA. The combination has the best results in terms of all five metrics according to the Friedman test. Moreover, the solution has a considerably low value of the M_{FPR} metric. Thus, we can conclude that this combination has relatively exceptional detection capabilities of cyber-attacks, including time consumptions with a considerably low number of false alerts (M_{FPR}). Thus, it can be stated that the most important parameters of the detection system are preserved, i.e., the identification of the cyber-attack, practically zero amount of false identification, and an acceptable amount of time consumption.

The final chapter focused on the interpretation of detected anomalies. The reverse procedure was used to obtain the most significant attributes for each classified anomaly record. However, the interpretation concept does not aim at the identification of origin and type of cyber-attack. In many cases, a deep knowledge of the ICS system and its processes is necessary for effective analysis as well as the adoption of measures. Therefore, close coordination with the technical staff of the ICS system is necessary to resolve the issue.

According to the results, we identify unfit machine learning algorithms for ICS systems. This is mainly the OCSVM algorithm which has unacceptable performance in all metrics. Even though bad results, there is a considerable number of authors that recommend the implementation of OCSVM for ICS systems.

The article was focused on the dynamic area of anomaly detection related to ICS cyber-security. The anomaly detection system was developed and tested. The results confirm the applicability of the system in a real environment. Mainly due to the considerable low number of false alarms. However, other metrics (M_{F1}, M_{MCC}, M_{Prec}), in addition to metric time, show promising results.

It is important to note the weakness of the presented system of cyber-attacks detection. The strength of every anomaly detection model relies on training data. If any malicious activities corrupt the data, then it will be reflected in the final model and therefore decrease its detection capabilities. This important and actual topic is called Adversarial Machine Learning (AML). There are a considerable number of AML attack vectors and mitigation strategies that are investigated by researchers in detail. Moreover, AML can be a serious threat to machine learning solutions as a whole in years to come. However, the research was not focused on this topic. Nonetheless, there should be implemented AML mitigation strategies for the implementation of any anomaly detection system based on machine learning algorithms in a real environment.

There are multiple possible ways of future research. One possible way is the parallel application of machine learning algorithms for anomaly detection, where the results would be based on the voting of used algorithms. Another possible area of future research is the precise interpretation of cyber-attacks. Thus, the development of an analytical tool for the analysis of the results is needed.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was funded by the Ministry of the Interior of the Czech Republic under Project VI20192022151 ‘CIRFI 2019: Indication of critical infrastructure resilience failure. Moreover, this work was supported by the resources of UIUI A.I.Lab at the Faculty of Applied Informatics, Tomas Bata University in Zlin (ailab.fai.utb.cz). Furthermore, computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

References

[1] Alejandro Germán FRANK, Lucas Santos DALENOGARE, Néstor Fabián AYALA, Industry 4.0 technologies: implementation patterns in manufacturing companies, International Journal of Production Economics 210 (2019) 15–26.
[2] Tyson MACAULAY, Bryan L SINGER, Cybersecurity For Industrial Control systems: SCADA, DCS, PLC, HMI, and SIS, CRC Press, 2011.
[3] Marina KROTOFIL, Klaus KURSAWE, Dieter GOLLMANN, Securing industrial control systems. Security and Privacy Trends in the Industrial Internet of Things, Springer, Cham, 2019, pp. 3–27.
[4] Alexander N. SOKOLOV, Ilya A. PYATNITSKY, Sergei K ALABUGIN, Research of classical machine learning methods and deep learning models effectiveness in detecting anomalies of industrial control system, in: 2018 Global Smart Industry Conference (GloSIC), IEEE, 2018, pp. 1–6.
[5] Junjiao LIU, et al., A novel intrusion detection algorithm for industrial control systems based on CNN and process state transition, in: 2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC), IEEE, 2018, pp. 1–8.
[6] Moshe; KRAVCHIK, Asaf. SHABTAI, Detecting cyber attacks in industrial control systems using convolutional neural networks, in: Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy, 2018, pp. 72–83.
[7] Kevser Ovaz AKPINAR, Ibrahim OZCELIK, Analysis of machine learning methods in EtherCAT-based anomaly detection, IEEE Access 7 (2019) 184365–184374.
[8] Antoine LEMAY, José M FERNANDEZ, Providing {SCADA} network data sets for intrusion detection research. 9th Workshop on Cyber Security Experimentation and Test ({CSET} 16), 2016.
[9] Jonathan GOH, et al., A dataset to support research in the design of secure water treatment systems, in: International Conference on Critical Information Infrastructures Security, Cham, Springer, 2016, pp. 88–99.
[10] Thomas H. MORRIS, Zach THORNTON, Ian TURNIPSEED, Industrial control system simulation and data logging for intrusion detection system research, 7th annual southeastern cyber security summit (2015) 3–4.
[11] Fabian PEDREGOSA, et al., Scikit-learn: machine learning in Python. the, Journal of machine Learning research 12 (2011) 2825–2830.
[12] Ian GOODFELLOW, et al., Deep Learning, MIT press, Cambridge, 2016.
[13] Sreeraj RAJENDRAN, et al., Crowdsourced wireless spectrum anomaly detection, IEEE Transactions on Cognitive Communications and Networking (2019) 694–703, 6.2.
[14] RUFF, Lukas, et al. Deep semi-supervised anomaly detection. arXiv preprint arXiv:1906.02694, 2019.
[15] Jose CAMACHO, et al., Semi-supervised multivariate statistical network monitoring for learning security threats, IEEE Transactions on Information Forensics and Security (2019) 2179–2189, 14.8.
[16] Fabian PEDREGOSA, et al., Scikit-learn: machine learning in Python, the Journal of machine Learning research 12 (2011) 2825–2830.
[17] CHOLLET, François. et al., Keras. 2015. Available at: https://github.com/fcholl et/keras.
[18] Nicholas; ARCOLANO, Daniel. RUDOY, One-class support vector machines: methods and applications, Harvard University, Final Project Presentation (2008) 32.
[19] Fei Tony LIU, Kai Ming TING, Zhi-Hua ZHOU, Isolation forest, in: 2008 Eighth IEEE International Conference on Data Mining, IEEE, 2008, pp. 413–422.
[20] Frank. ROSENBLATT, The perceptron: a probabilistic model for information storage and organization in the brain, Psychol Rev (1958) 386, 65.6.
[21] Sepp HOCHREITER, Jürgen SCHMIDHUBER, Long short-term memory, Neural Comput (1997) 1735–1780, 9.8.
[22] LIPTON, Zachary C.; BERKOWITZ, John; ELKAN, Charles. A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019, 2015.
[23] Gwo-Hshiung TZENG, Jih-Jeng HUANG, Multiple Attribute Decision making: Methods and Applications, CRC press, 2011.
[24] John Henry HOLLAND, et al., Adaptation in Natural and Artificial systems: an Introductory Analysis With Applications to biology, control, and Artificial Intelligence, MIT press, 1992.
[25] James S. BERGSTRA, et al., Algorithms for hyper-parameter optimization, In: Advances in neural information processing systems (2011) 2546–2554.
[26] Janez DEMŠAR, Statistical comparisons of classifiers over multiple data sets, Journal of Machine learning research 7 (2006) 1–30. Jan.