A New Proposal for Detection and Location of Cyberattacks in Industrial Processes

Adrián Rodríguez-Ramos[1], Eloy Irigoyen[2], Antônio J. da Silva Neto[3], and Orestes Llanes-Santiago[1(✉)]

[1] Universidad Tecnológica de la Habana José Antonio Echeverría, CUJAE, La Habana, Cuba
adrian.rr@automatica.cujae.edu.cu, orestes@tesla.cujae.edu.cu
[2] Universidad del País Vasco, Bilbao, Spain
eloy.irigoyen@ehu.eus
[3] Instituto-Politécnico - Universidade do Estado do Rio de Janeiro, Nova Friburgo, RJ, Brazil
ajs_neto@uol.com.br

Abstract. In the Industry 4.0 paradigm, the cybersecurity is a key aim to obtain high levels of performance of the industries based on the use of the IoT technology and the Big Data analysis. To achieve this objective, the cyberphysical industrial plants must be equipped with cybersecurity systems for early detection and location of cyberattacks. This paper presents a robust approach of an industrial cybersecurity system by using non-standard Pythagorean membership grades. The proposed scheme was validated using the Two-Tanks benchmark with excellent results. The proposal was compared with other computational intelligence tools recently presented in the scientific literature, and the results showed the best performance of the proposed scheme.

Keywords: Industry 4.0 · Cybersecurity · Industrial Plants · Fuzzy algorithms · Pythagorean fuzzy sets

1 Introduction

At present, terms such as Smart Factory and Industry 4.0 are closely related to the automation of industrial plants characterized by increasingly connected physical systems and a stronger integration of digital technologies [4,8]. This higher level of integration allows higher levels of productivity, more competitive final products, and excellent compliance of the industrial safety standards. However, despite these significant commercial benefits, the safety risk of these cyberindustrial environments is also increased. Therefore, there is an urgent need for increasing the cybersecurity in industrial processes [1,2].

A major quality of the use of fuzzy sets is the insertion of membership degrees. With the aim of improving the ability of fuzzy sets to capture and model membership information, several researchers have begun to use non-standard fuzzy

Y. Hernndez Heredia et al. (Eds.): IWAIPR 2023, LNCS 14335, pp. 96–107, 2024.
https://doi.org/10.1007/978-3-031-49552-6_9

sets such as the intuitionistic [3], that allow to insert imprecision and uncertainty in the specification of the membership degrees.

The Pythagorean Fuzzy Sets (PFS) were presented in [10], where it is demonstrated that the space of Pythagorean membership degrees is larger than the space of intuitionistic membership degrees. This represents an important advantage in condition monitoring because it allows the insertion of uncertainty in the specification of membership degrees as result of noisy measurements.

The aim of this paper and its main contribution is to propose a cybersecurity scheme with a high performance in the detection and location of cyberattacks and with a robust behaviour versus noisy observations obtained from an industrial plant. The proposal is based on the modification of the Kernel Fuzzy C-Means algorithm by using the non standard Pythagorean membership grades. The modified algorithm, called Kernel Pythagorean Fuzzy C-Mean algorithm (KPyFCM), significantly reduces classification errors in the attack detection and location even in the presence of noisy observations. On the other hand, a performance comparison is developed with successful algorithms used in different applications [6,12].

2 Materials and Methods

2.1 Main Characteristics of Pythagorean Fuzzy Sets

In [10], the PFS were introduced. The Pythagorean Membership Grades (PMG) associated with them are expressed as follow:

- Two values, $r(z)$ and $d(z)$, are assigned for each $z \in Z$.
- If $r(z) \in [0,1]$, it is labeled *strength of commitment* at z
- If $d(z) \in [0,1]$, it is labeled *direction of commitment* at z.
- $\mathcal{H}_Y(z)$ is a membership grade which indicates the support for membership of z in $\mathcal{H}$.
- $\mathcal{H}_N(z)$ is a membership grade which indicates the support against membership of z in $\mathcal{H}$.
- $\mathcal{H}_Y(z)$ and $\mathcal{H}_N(z)$ are defined as

$$\mathcal{H}_Y(z) = r(z)cos(\varphi(z)) \tag{1}$$

$$\mathcal{H}_N(z) = r(z)sin(\varphi(z)) \tag{2}$$

where

$$\varphi(z) = (1 - d(z))\frac{\pi}{2} \tag{3}$$

and $\varphi(z) \in [0, \frac{\pi}{2}]$ is expressed in radians.

Lemma: *$\mathcal{H}_Y(z)$ and $\mathcal{H}_N(z)$ are Pythagorean complements with respect to r(z)*

Proof: See [11]

In general, a PMG is formalized by using a pair of values (e, f) such that $e, f \in [0,1]$ and $e^2 + f^2 \leq 1$.

Intuitionistic membership grades are also represented by a pair (e, f) which satisfies $e, f \in [0,1]$ and $e + f \leq 1$ [3].

Theorem: *The set of Pythagorean Membership Grades is greater than the set of intuitionistics membership grades*

Proof: See [11]

This result indicates the possibility of using PFS in more situations than intuitionistics fuzzy sets. For cybersecurity systems this characteristic of the PFS is very important for improving their performance.

2.2 Kernel Pythagorean Fuzzy C-Means Algorithm

Using the PFS theory, the objective function of the Pythagorean FCM (PyFCM) algorithm can be obtained in the similar form to the Intuitionistic Fuzzy C-Means algorithm (IFCM) [3] according to the equation

$$J_{PyFCM} = \sum_{i=1}^{l}\sum_{k=1}^{N} u_{ik}^{*m} d_{ik}^2 + \sum_{i=1}^{l} \pi_i^* e^{1-\pi^*} \tag{4}$$

where $m > 1$ is the fuzziness regulation factor of the partition [9], l is the quantity of classes, N is the quantity of observations.

$u_{ik}^* = u_{ik}^m + \pi_{ik}$. u_{ik}^* represents the pythagorean fuzzy membership, u_{ik} denotes the typical fuzzy membership of the kth observation in the ith class, and π_{ik} is the hesitation degree, formalized as:

$$\pi_{ik} = 1 - u_{ik}^2 - (1 - u_{ik}^{\alpha})^{2/\alpha}, \alpha > 0 \tag{5}$$

and

$$\pi_i^* = \frac{1}{N}\sum_{k=1}^{N} \pi_{ik}, k \in [1, N] \tag{6}$$

Kernel functions permit to map non-linear observation of the input space into a higher-dimensional space. This is very useful in classification tasks because allow for greater separability among classes. With this aim, the Kernel Pythagorean Fuzzy C Mean algorithm (KPyFCM) is designed. In this algorithm, the following objective function is minimized:

$$J_{KPyFCM} = \sum_{i=1}^{l}\sum_{k=1}^{N} u_{ik}^{*m} \left\|\mathbf{\Psi}(\mathbf{z_k}) - \mathbf{\Psi}(\mathbf{q_i})\right\|^2 + \sum_{i=1}^{l} \pi_i^* e^{1-\pi^*} \tag{7}$$

where, $\left\|\mathbf{\Psi}(\mathbf{z_k}) - \mathbf{\Psi}(\mathbf{q_i})\right\|^2$ denotes the square of the distance between $\mathbf{\Psi}(\mathbf{z_k})$ and $\mathbf{\Psi}(\mathbf{q_i})$. In the feature space, the distance is computed by using the kernel function as follows:

$$\left\|\mathbf{\Psi}(\mathbf{z_k}) - \mathbf{\Psi}(\mathbf{q_i})\right\|^2 = \mathbf{K}(\mathbf{z_k}, \mathbf{z_k}) - \mathbf{2K}(\mathbf{z_k}, \mathbf{q_i}) + \mathbf{K}(\mathbf{q_i}, \mathbf{q_i}) \tag{8}$$

There exist many kernel functions and the choice of the most appropriate depends on the application [5]. Nonetheless, the most used is the Gaussian Kernel function (GKF).

If the GKF is used, then $\mathbf{K}(\mathbf{z}, \mathbf{z}) = \mathbf{1}$ and $\|\mathbf{\Psi}(\mathbf{z_k}) - \mathbf{\Psi}(\mathbf{q_i})\|^2 = \mathbf{2}\,(\mathbf{1} - \mathbf{K}(\mathbf{z_k}, \mathbf{q_i}))$. So, Eq. (7) can be expressed as:

$$J_{KPyFCM} = 2\sum_{i=1}^{l}\sum_{k=1}^{N} u_{ik}^{*m}\,\|1 - \mathbf{K}(\mathbf{z_k}, \mathbf{q_i})\|^2 + \sum_{i=1}^{l} \pi_i^* e^{1-\pi^*} \tag{9}$$

where,

$$\mathbf{K}(\mathbf{z_k}, \mathbf{q_i}) = e^{-\|\mathbf{z}_k - \mathbf{q}_i\|^2/\delta^2} \tag{10}$$

where δ is the bandwidth which indicates the smoothness degree of the GKF [9]. Minimizing Eq. (9), yields:

$$u_{ik}^* = \frac{1}{\sum_{j=1}^{l}\left(\frac{1-\mathbf{K}(\mathbf{z_k},\mathbf{q_i})}{1-\mathbf{K}(\mathbf{z_k},\mathbf{q_j})}\right)^{1/(m-1)}} \tag{11}$$

$$\mathbf{q}_i = \frac{\sum_{k=1}^{N}(u_{ik}^{*m}\mathbf{K}(\mathbf{z_k}, \mathbf{q_i})\mathbf{z_k})}{\sum_{k=1}^{N} u_{ik}^{*m}\mathbf{K}(\mathbf{z_k}, \mathbf{q_i})} \tag{12}$$

KPyFCM algorithm is displayed in Algorithm 1.

Algorithm 1: KPyFCM algorithm

Data: l, $\gamma > 0$, $m > 1$, δ, Itr_{max} (maximum number of iterations)
Result: fuzzy partition **U**, class centers **Q**

```
Initialize U to random fuzzy partition;
Itr ← 1 ;
repeat
    Update the class centers Q according to (12);
    Calculate the distances according to (8);
    Update U according to (11).;
    Itr ← Itr + 1;
until ||U_t − U_{t−1}|| < γ ∧ Itr ≥ Itr_max;
```

2.3 Proposal of Scheme for Detection and Localization of Cyberattacks

The proposal of scheme for Detection and Location of Cyberattacks is shown in Fig. 1. It is formed for two phases: a training phase executed offline and a

recognition phase developed online. In the first phase, the data obtained from the process allow to train offline the Cyberattack Detection and Location (CADL) algorithm. After training, the CADL algorithm is used online to analyze each new observation taken from the process. Training is the most important stage, since the center of the different classes that represent the process operation states are determined (normal operation class and the classes that represent the different cyberattacks).

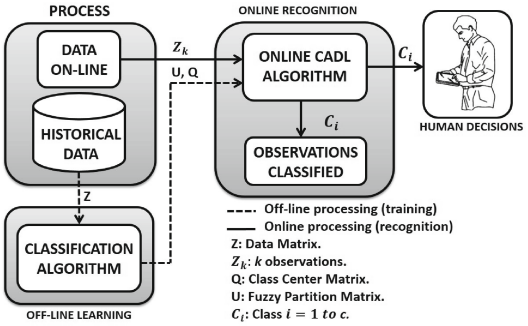


Fig. 1. Proposal of scheme for cyberattack detection and location.

Offline Training Stage. In this stage, the CADL system is trained with a set of historical data which contain the necessary information of each known operating state or class of the industrial plant (normal operation condition (NOC) and states of attack). The main aim of the training process is to determine the center of the known classes $\mathbf{Q} = \{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_c\}$ to be used in the on-line recognition stage.

On-Line Recognition Stage. In this stage, each observation (O_k) obtained from the process is assigned to a known class based on the distance between the observation and the centers of the different classes. Subsequently, the membership degree of the observation k for each class is obtained. The observation is assigned to the class with the highest membership degree such as is showed in Eq. (13). (See Algorithm 2).

$$C_i = \{i : max\{u_{ik}^*\}, \forall i, k\} \tag{13}$$

2.4 Case Study: Two-Tanks

The test system consists of two tanks (T_1, T_2) interconnected through a pipe with a valve V_b which is actuated by an ON-OFF controller (see Fig. 2) [7]. Tank T_1

Algorithm 2: Recognition stage

Data: observation z_k, class centers $\mathbf{Q}$, m, c
Result: Current State
1 Calculate the distances among the observation z_k and the class centers using Eq. (8);
2 Calculate the membership degree of the observation z_k to the c classes according to Eq. (11);
3 Determine using Eq. (13) the class to which observation z_k belongs;

is fed by the pump P_1 controlled by a proportional integral (PI) controller. Tank T_2, is equipped with the manual outlet valve V_0. The variables of the process are: Inflow to T_1 (Q_p), Water level in T_1 (h_1), Water level in T_2 (h_2), Pump control signal on T_1 (U_p, Outflow to consumers (Q_0), Outflow at T_1 (Q_{f1}) and Outflow at T_2 (Q_{f2}).

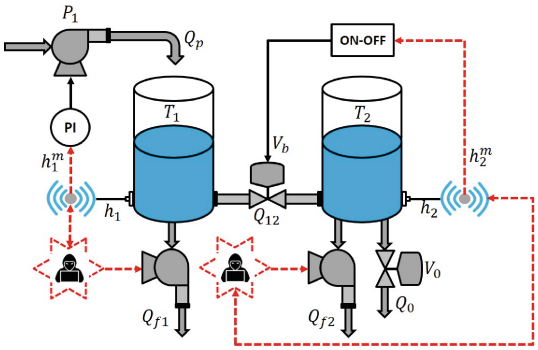


Fig. 2. Schematic diagram of the Two-Tank test system.

In dependence of the type of cyberattack, different scenarios can be obtained. In this paper, the following five scenarios were simulated:

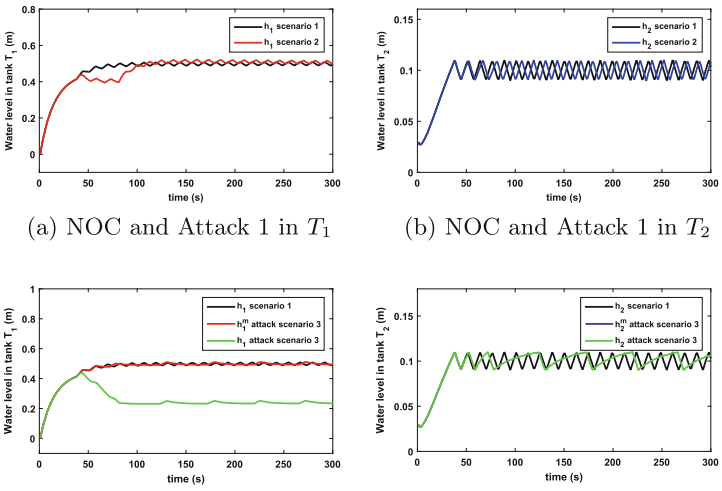
- **Scenario 1 (NOC):** Normal Operation Condition.
- **Scenario 2 (A1):** Scenario corresponding to Attack 1 due to a water leak in Tank 1 (T_1) at a constant flow $Q_{f1} = 10^{-4}$ m^3/s in the period of time from $t = 40$ s to $t = 80$ s.
- **Scenario 3 (A2):** Short-term water theft from T_1 with hidden signal added to the h_1^m level measurement (Deception attack). The attacker extracts a constant flow $Q_{f1} = 10^{-4}$ m^3/s through the pump between $t = 40$ s and 80 s. In this period of time a signal is added to the level sensor output at T_1 to hide the theft. For the PI controller the level at T_1 seems to remain constant and its output does not change.
- **Scenario 4 (A3):** This scenario corresponds to Attack 3 due to a water leak in Tank 2 (T_2) at a constant flow $Q_{f2} = 10^{-4}$ m^3/s in the time period $t = 40\,\text{s} - t = 80\,\text{s}$.

– **Scenario 5 (A4):** The attacker steals water when the system has reached the steady state. Before doing so, the attacker saves several measurements of the level sensors before the water is stolen from the tanks. In the attack phase, the attacker steals water and replacing the real data with the saved ones (Replay attack). Specifically, water is stolen in the period of time between $t = 100\,\text{s}$ and $t = 200\,\text{s}$ and the controller is fooled by using the measurements saved in the 50 s prior to the attack.

2.5 Design of Experiments

For building the training database, simulations were carried out to obtain 160 observations of each class (NOC and Attacks). For validating the behavior of the CADL system in the online stage were used another set of 40 observations of each class. Figures 3 and 4 shows a comparison between the water levels in the tanks with scenario 1 (NOC) and the different attacks scenarios (2, 3, 4 and 5). The values of the parameters used in the KPyFCM algorithm were: $\epsilon = 10^{-5}$, $m = 2$, $\sigma = 10$. The value of σ was selected after the development of 10 experiments (σ = 10,20,30,40,...,100). Three experiments were developed to evaluate the robustness of the cyberattack detection and location system against noise:

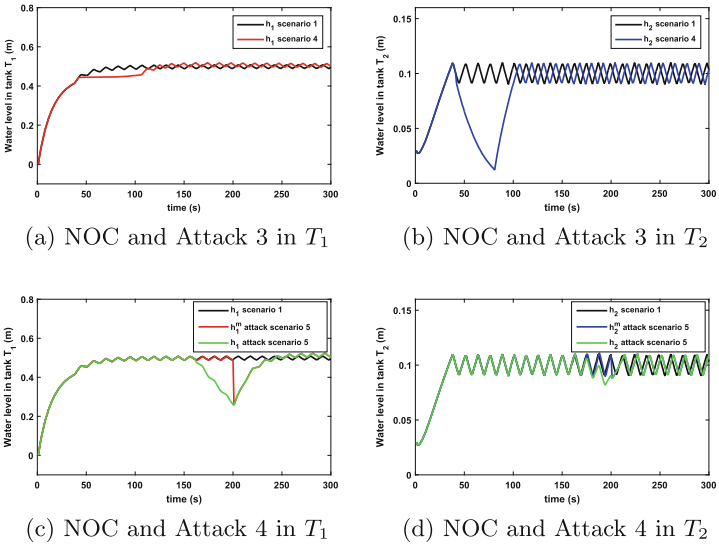
1. Without noise in the measurements.
2. Measurements with 2% of noise level.
3. Measurements With 5% of noise level.



(a) NOC and Attack 1 in T_1 (b) NOC and Attack 1 in T_2

(c) NOC and Attack 2 in T_1 (d) NOC and Attack 2 in T_2

Fig. 3. Comparison between scenarios 1 (NOC), 2 (Attack 1) and 3 (Attack 2).



(a) NOC and Attack 3 in T_1 (b) NOC and Attack 3 in T_2

(c) NOC and Attack 4 in T_1 (d) NOC and Attack 4 in T_2

Fig. 4. Comparison between scenarios 1 (NOC), 4 (Attack 3) and 5 (Attack 4).

3 Analysis and Discussion of Results

For the performance analysis of the proposed CADL system the Confusion Matrix (CM) was used. The values CM_{rs} for $r \neq s$ in the CM show the number of observations of the operation mode r that the CADL system misclassifies in the operation modes.

Table 1 shows the CM (without noise in the measurements) where NOC: Normal Operation Condition, A1: Attack 1, A2: Attack 2, A3: Attack 3 and A4: Attack 4. The main diagonal presents the number of observations successfully classified. The accuracy TA for each class was computed as $TA =$ correctly classified observations of the class/total observations of the class. The last row shows the average (AVE) of TA. Figure 5 show the classification results for the different scenarios by using the proposed CADL system for Two-Tank benchmark.

3.1 Comparison with Other Algorithms

For improving the classification process, the Density-Based Weighted FCM (DBFCM) algorithm [6], the Maximum-Entropy-Regularized Weighted FCM (EWFCM) algorithm [12], and the Kernel based EWFCM (KEWFCM) algorithm [12], all of them with excellent performance in different applications, have been presented in the scientific literature. Follow a comparison with these algorithms.

Density-Based Weighted Fuzzy C-Means Algorithm. In this algorithm, the weight of an object is decided by the density of the objects around this

Table 1. Confusion matrix: KPyFCM (NOC: 40, A1: 40, A2: 40, A3: 40, A4: 40)

	NOC	A1	A2	A3	A4	TA (%)
NOC	**38**	0	2	0	0	95.00
A1	0	**39**	1	0	0	97.50
A2	3	0	**37**	0	0	92.50
A3	0	0	0	**40**	0	100.0
A4	2	0	0	0	**38**	95.00
AVE						**96.00**

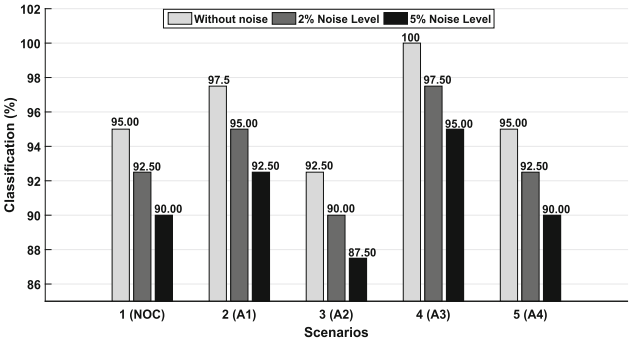


Fig. 5. Classification (%) for Two-Tank process.

object. There are two stages of the density-based weighted FCM. The first stage is designed to calculate the weights of every object, the second stage is the clustering stage.

Maximum-Entropy-Regularized Weighted Fuzzy C-Means and Kernel Maximum-Entropy-Regularized Weighted Fuzzy C-Means Algorithms. A maximum-entropy-regularized weighted fuzzy c-means (EWFCM) algorithm is proposed to extract the important features and improve the clustering. In EWFCM algorithm, the attribute-weight entropy regularization is defined in the new objective function to achieve the optimal distribution of the attribute weights. The kernel based EWFCM (KEWFCM) clustering algorithm is developed for clustering the data with non-spherical shaped clusters.

Table 2 show the results of the CM (without noise) corresponding to the algorithms used in the comparison.

Figure 6 shows the classification results by using the mentioned algorithms and the KPyFCM algorithm for the observations without noise, with 2% and 5% of noise level.

All experiments were performed on a computer with the following characteristics: Intel Core i7-6600U 2.6–2.81 GHz, memory RAM: 16 GB. The average computational time of each algorithm to perform an execution was: DBFCM

Table 2. Confusion matrix: DBWFCM, EWFCM, KEWFCM (NOC: 40, A1: 40, A2: 40, A3: 40, A4: 40)

	NOC	A1	A2	A3	A4	TA (%)
NOC	**34**	1	5	0	0	85.00
A1	0	**36**	3	1	0	90.00
A2	4	2	**34**	0	0	85.00
A3	0	4	1	**35**	0	87.50
A4	5	0	0	1	**34**	85.00
AVE						**86.50**

(a) DBWFCM

	NOC	A1	A2	A3	A4	TA (%)
NOC	**35**	1	4	0	0	87.50
A1	0	**36**	3	1	0	90.00
A2	4	2	**34**	0	0	85.00
A3	0	4	0	**36**	0	90.00
A4	5	0	0	1	**34**	85.00
AVE						**87.50**

(b) EWFCM

	NOC	A1	A2	A3	A4	TA (%)
NOC	**36**	0	4	0	0	90.00
A1	0	**38**	1	1	0	95.00
A2	4	1	**35**	0	0	87.50
A3	0	3	0	**37**	0	92.50
A4	4	0	0	0	**36**	90.00
AVE						**91.00**

(c) KEWFCM

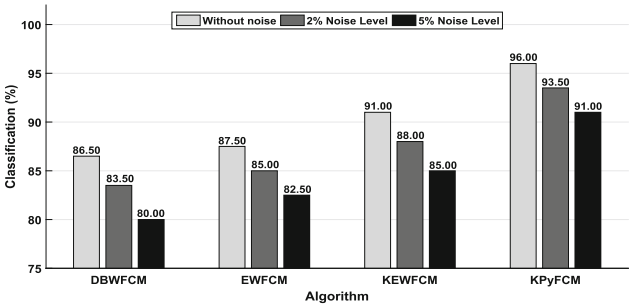


Fig. 6. Global classification (%) obtained for each algorithm.

(0.5520 s), EWFCM (0.4962 s), KEWFCM (0.6315 s), and KPyFCM (0.6035 s). When comparing these times with the time constant of the process, it can be seen that they are very small and, therefore, show the feasibility of their application to the proposed scheme.

The Friedman was applied, and its results confirmed that at least, the performance of one algorithm is significantly different to the performances of the other algorithms.

The Wilcoxon test was applied to compare the algorithms in pairs where 1: DBWFCM, 2: EWFCM, 3: KEWFCM, 4: KPyFCM. Table 3 displays the results. The first row contains the values of the sum of the positive rank (R^+) and the second row presents the sum of the negative rank (R^-) for each comparison. The

third row shows the statistical values T, and the fourth row, the critical value of T for a significance level $\alpha = 0.05$. The last row displays number the winner algorithm in each comparison. Table 4 shows the times that each algorithm was the winner. This results demonstrate the proper performance of the CADL scheme proposed.

Table 3. Results of the Wilcoxon test

	1 vs 2	1 vs 3	1 vs 4	2 vs 3	2 vs 4	3 vs 4
$\sum R^+$	5	0	0	0	0	0
$\sum R^-$	50	55	55	55	55	55
T	5	0	0	0	0	0
$T_{\alpha=0.05}$	8	8	8	8	8	8
Winner	2	3	4	3	4	4

Table 4. Final result of the comparison between algorithms

Algorithm	No.Wins	Ranking
DBWFCM	0	4
EWFCM	1	3
KEWWFCM	2	2
KPyFCM	3	1

4 Conclusions

To achieve the successful implementation of the Industry 4.0 paradigm in industrial plants, cybersecurity must be guaranteed. In this paper, an attack detection and location system with a high performance and robustness versus noisy measurements was presented. The CADL system was implemented by using a KPyFCM algorithm which significantly improves the performance in the detection and location process based on two key characteristics of the elements that make it up. The first is related with the fact that the Pythagorean membership grades permit to use a larger set of numeric values for assigning the membership degree to an observation than the standard and intuitionistic membership grades. The second is related with use of kernel functions which allow to achieve greater separability among the classes. The high performance of the proposal of CADL scheme was confirmed using the Two-Tank process benchmark. The results obtained by the proposed CADL system were compared with the results

of three algorithms of high performance recently presented in the scientific literature demonstrating the superiority of the proposal for detection and location of cyberattacks. For future research, an interesting idea will be designing a monitoring scheme that integrally addresses fault diagnosis and cyberattacks in industrial plants.

Acknowledgements. The authors acknowledge the financial support provided by FAPERJ, Fundacão Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro; CNPq, Consehlo Nacional de Desenvolvimento Científico e Tecnológico; CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, research supporting agencies from Brazil and the project PN223LH004-23 from the Science and Technology National Program in Automation, Robotic and Artificial Intelligence (ARIA) of the Ministry of Science, Technology and Environment (CITMA) of Cuba.

References

1. Alanazi, M., Mahmood, A., Morshed, M.: Scada vulnerabilities and attacks: a review of the state of the art and open issues. Comput. Secur. **125**, 1–29 (2023)
2. Alladi, T., Chamola, V., Zeadally, S.: Industrial control systems: cyberattack trends and countermeasures. Comput. Commun. **155**, 1–8 (2020)
3. Atanassov, K.: On Intuitionistic Fuzzy Sets Theory. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29127-2
4. Bashendy, M., Tantawy, A., Erradi, A.: Intrusion response systems for cyber-physical systems: a comprehensive survey. Comput. Secur. **124**, 1–27 (2023)
5. Bernal de Lázaro, J., Cruz Corona, C., Silva Neto, A., Llanes-Santiago, O.: Criteria for optimizing kernel methods in fault monitoring process: a survey. ISA Trans. **127**, 259–272 (2022)
6. Li, Y., Yang, G., He, H., Jiao, L., Shang, R.: A study of large-scale data clustering based on fuzzy clustering. Soft Comput. **20**, 3231–3242 (2016)
7. Quevedo, J., Sánchez, H., Rotondo, D., Escobet, T., Puig, V.: A two-tank benchmark for detection and isolation of cyber-attacks. IFAC Paper OnLIne **51**, 770–775 (2018)
8. Rodríguez-Ramos, A., Bernal-de Lázaro, J., Prieto-Moreno, A., Silva Neto, A., Llanes-Santiago, O.: An approach to robust fault diagnosis in mechanical systems using computational intelligence. J. Intell. Manuf. **30**, 1601–1615 (2019)
9. Rodríguez-Ramos, A., Silva-Neto, A.J., Llanes-Santiago, O.: An approach to fault diagnosis with online detection of novel faults using fuzzy clustering tools. Expert Syst. Appl. **113**, 200–212 (2018)
10. Yager, R.R.: Pythagorean membership grades in multicriteria decision making. IEEE Trans. Fuzzy Syst. **22**, 958–965 (2014)
11. Yager, R.R.: Properties and applications of Pythagorean fuzzy sets. In: Angelov, P., Sotirov, S. (eds.) Imprecision and Uncertainty in Information Representation and Processing. SFSC, vol. 332, pp. 119–136. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-26302-1_9
12. Zhou, J., Chen, L., Chen, C.P., Zhang, Y., Li, H.: Fuzzy clustering with the entropy of attribute weights. Neurocomputing **198**, 125–134 (2016)