FAULT DIAGNOSIS BASED ON NATURE-INSPIRED FEATURE SELECTION METHODS

Adel ABDELHADI[1], Ouahab KADRI[2], * and Mohamed Taki Eddine SEDDIK[1]

[1]Department of Computer Science, University of Batna 2, Algeria,
A.Abdelhadi@univ-batna2.dz
m.seddik@univ-batna2.dz
[2]Department of Statistic and Data Science, University of Batna 2, Algeria
O.kadri@univ-batna2.dz

AJME 2025, 23 (3); https://doi.org/10.5281/zenodo.17217482

***ABSTRACT**: Fault detection problems in industrial and computer fields have a major impact on the operating stability of any system. The number of inputs influences the analysis time and requires a considerable amount of memory. The use of meta-heuristics avoids exhaustive research in the solution space and represents an ideal solution to reduce the number of inputs in the fault detection process. In this article, we have proposed a set of improvements on algorithms inspired from the natural world, adapted to the problem of feature selection. Our major contribution lies in the proposal of a new algorithm based on the behavior of real ants and its application to accelerate the process of fault detection in three different systems. The obtained results demonstrate the effectiveness of the proposed algorithm.*

***KEYWORDS**: Fault diagnosis, image processing, Nature-Inspired Feature Selection Methods, industrial system, diagnosis and maintenance*

1 INTRODUCTION

Context and motivation: The objective of industrial diagnostics is to identify the operating state of one or more machines and to find the breakdown causes. There are several methods for detecting a failure using observations. We can model this phenomenon as a supervised classification problem (Dyukov et al, 2023).

Machine learning algorithms are the most used to perform supervised classification. They use structured data in the form of a table of N columns. The first (N-1) columns represent the system parameters. The different labels are recorded in the last column. The objective of these algorithms is to create a model which makes it possible to find the outputs from the inputs (Kaprielova et al, 2022).

It is very difficult to avoid the problem of Overfitting since, in the majority of cases, we do not have the operating history of the studied system. The model created, using a limited dataset and which does not contain a sufficient number of observations, does not make it possible to distinguish between several operating states (Kadri et al, 2022).Using a large dataset increases the training time. But in several cases, we can use a limited number of parameters to speed up the learning process while maintaining the same rate of accuracy (Cheeseman et al, 2022).This observation led to the definition of a pre-processing step which makes it possible to reduce the dimension of the data structure used for the creation of the model. The new table is composed of a reduced number of columns and the same number of observations of the table containing the raw data. The selection and the extraction of parameters are the two main categories of methods allowing transforming data from a high-dimensional space into a low-dimensional space by preserving the relations existing between the various data composed the table (Hashemi et al, 2022).

Feature selection uses techniques that take as input a set of parameters and return as output a reduced subset by eliminating redundant and irrelevant parameters. The principle of parameter extraction consists in calculating new parameters from the initial set. There are extraction methods that use linear transformation such as principal component analysis (PCA) and other methods that use non-linear transformation such as Kernel PCA (Haq et al, 2021).

Classification architectures that use parameter selection algorithms can be easily run on machines with limited computing power and storage capacity. A wide range of applications uses feature selection to solve various topical problems such as medical diagnosis, surveillance based on facial classification, and natural language analysis.

The selection of parameters is very used in the industrial field and more in the diagnosis of breakdowns. It reduces the number of sensors installed and increases the speed of detection of degradation or failure.

Contributions: In this paper, we present the results of recent work on feature selection methods applied to fault detection. We explain in detail the difference between industrial diagnosis and prognosis. We compare the different artificial intelligence techniques applied to the feature selection. We describe five feature selection methods and we proposed implementations of the ACO algorithm to solve the fault detection problem for different systems.

Organization of the paper: In the rest of this article, we address the following points: A presentation of recent work in the field of industrial diagnostics is provided in section 2. In section 3, we describe the most recent techniques used to carry out feature selection. Section 4 describes the use of five bio-inspired methods to perform feature selection. In section 5, the results of the application of three parameter selection methods on the data issued from a pasteurization system and a Clinkerization system. We end this article with conclusions and some perspectives.

2 PREVENTIVE MAINTENANCE

The operating behaviour of industrial systems is very complex. The use of a mathematical model is not obvious because of the enormous number of parameters that characterize this type of system. The diagnostic procedure begins with checking the operating status of the system and ends with the identification of failed components. The principle used is to start from symptoms to causes of failures. Before going deeper into the selection of parameters, we briefly review the two essential notions related to industrial diagnosis, which are pattern recognition and the representation space (Liang et al, 2022).

2.1 Pattern recognition

According to several authors, diagnosis is defined as a problem that can be solved using pattern recognition techniques. Knowing that, there are two types of techniques which are statistical and syntactic (Shu et al, 2021). In statistical pattern recognition, we use a vector to represent an operating state of the system. After acquiring the data using the sensors, we downsize the inputs using either an extraction or a selection method. In many cases, we use both methods at the same time. The last step is the evaluation of the system by classifying the data.

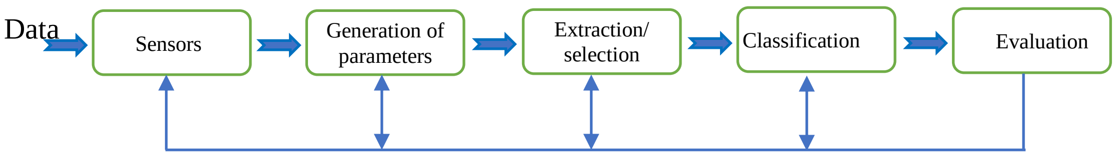

Fig. 1 The process of pattern recognition

2.2 The representation spaces

The quality of the results depends on the pre-processing applied to the inputs. The main objective of this method is to ensure a quick identification of the operating state of the systems by using a reduced number of parameters (Zhang et al, 2023). The difference between the two types of representation space reduction methods is the nature of the generated set. The extraction methods make it possible to create new parameters by applying operations on the elements of the initial set. While selection methods remove irrelevant parameters from the initial set. There are several methods for selecting parameters. However, they use the same outline to generate the final set of parameters. The selection process consists of repeating the execution of a search method which receives a set of parameters as input and generates another set as output. A function is used to determine the quality of the solution. The process ends its execution according to its initial configuration.

2.3 Industrial systems datasets

Raw data requires pre-processing to be used. Generally, the data is stored in a structure of two dimensions where the lines represent the observations at a given instant and each column contains the values relating to a specific sensor (Kadri et al, 2012).

2.3.1 Fault diagnosis using image processing

The Google scholar search engine returned 15100 results by answering a query including the keywords "fault diagnosis" and "image processing" and we have limited the results to articles published between 2019 and 2023. This proves that diagnosis using image processing is a challenging field of research. Image classification, regression and segmentation can be applied to improve the detection, diagnosis and maintenance of industrial systems. Among the most used techniques is the diagnosis based on thermal imaging. Its principle basic is the use of infrared radiation emitted by different objects to detect the defective object. One of the new methods used in monitoring is optical detection of breakdowns using the correlation of digital image. Installation of several cameras is necessary for the construction of three dimensional images. This type of image makes it possible to provide a rich vision in information of a manufacturing structure and to detect major change before the appearance of degradation (Rithani et al, 2023).

Fig. 2 Fault diagnosis based on correlation of digital image

2.3.2 Industrial signal processing

The signals used to carry out a diagnosis. It can be of the type: images, vibrations, acoustic emission, temperature, oil level, electrical or data acquisition system signals. Multiple sensors are used to collect signals from the different Industrial system components. There are several signal processing methods used to diagnose industrial systems:

A category includes methods based on statistical analysis and probability. A second category consists of Artificial Intelligence methods (AI). The main purpose of these methods is to ensure the correct pre-processing of the raw signals to extract the parameters.

A comparison with the thresholds is used to detect faults and to locate broken elements. To carry out this task, we can use classification methods based on artificial intelligence such as Extreme Learning Machine, Random Forest, etc (Infantraj & Kumaran, 2023).

According to the table below, we conclude that all the machine learning methods are able to solve the diagnosis problem.

Table 1. Description of three machine learning methods.

Method	**Domain**	**Function**	**Resolution**		**Complexity/ computational**	**Handling nonstationary signal**	**Signal sampling rate**
			Time domain	Frequency domain			
SVM	Time/ frequency	Feature selection/ diagnosis	Rely on input	Rely on input	Medium/ High	Possible	Medium/ Low
Statistical Analyses	Time/ frequency	Feature extraction	Rely on input	Rely on input	Low	Possible	Any
ANN	Time/ frequency	Feature selection/ diagnosis	Rely on input	Rely on input	Medium/ High	Possible	Medium/ Low

2.3.3 Public Datasets

The main use of public datasets is to test and evaluate new techniques. There are several servers on the net offering datasets divided according to their application areas such as industrial maintenance, classification, regression, etc. The most recent works use datasets are published on the Kaggle site. This server contains more than 234 data sets in the field of industrial maintenance. Most datasets are stored in CSV format (Quaranta et al, 2021).

Table 2 Dataset repositories for various real-world problems

Data base	Domain	Description	Path
UCI Machine Learning Repository	time-series, classification, regression, recommendation systems	This web site is created by the University of California, School of Information and Computer Science. It contains more than 600 types of machine learning datasets.	https://archive.ics.uci.edu/
Kaggle Datasets	Computer Science, Education, Classification, Computer Vision, NLP, Data Visualization, Pre-Trained Model	Kaggle contains over 15 million real-world data sets of all shapes and sizes and in many different formats. It also allows sharing of source code.	https://www.kaggle.com/
Amazon Datasets	Public Transport, Ecological Resources, Satellite Images, etc.	Datasets are stored in the Cloud to enable broad sharing	https://aws.amazon.com/
Image Sciences Institute	medical image analysis	It was created by 3D Computer Vision team. However, the activities comprised much more than volumetric computer vision.	https://www.isi.uu.nl/research/databases/

3 FEATURE SELECTION

The reduction in number of input parameters improves classification performance on several aspects. There are two very well-known techniques. Despite they have the same objectives, their principles are completely different. Each technique has a set of advantages and at the same time it suffers from several drawbacks (Jiang et al, 2021). We can use the selection of parameters in all fields since the result of the application of this method is always a subset of parameters characterizing the behaviour of the studied system. On the other hand, the extraction of parameters is not recommended in several fields since the new parameters created do not have a physical interpretation. But its application is very useful in detection systems based on image processing.

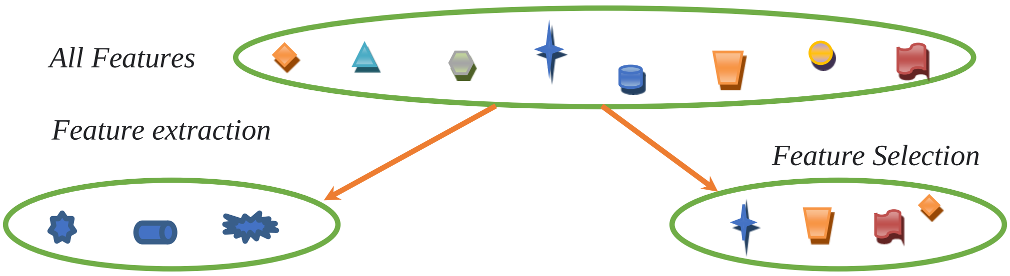

Fig. 3 Feature extraction VS feature selection.

We can obtain the selected subset by performing an individual evaluation for each parameter, for example the Fisher filter. This evaluation can be applied to a subset of parameters and in this case, we must use an objective function, for example an SVM classifier (Jiang et al, 2023). There are three types of functionality selection methods. The filter methods are based on the calculation of the relationship between the different parameters to eliminate redundancy. This technique is relatively fast. These methods can be used as heuristic factors in several algorithms like ACO, Go and Swarm.

The second type includes Wrapper methods that can be used jointly with artificial intelligence. The large number of tests makes these methods very slow (Ghosh et al, 2020). The third type includes embedded Methods. It is based on the use of artificial intelligence methods. The use of subsets evaluation function jointly with the learning operation makes this type of method very slow. Recent feature selection methods are based on metaheuristics. We can adopt one of three strategies to implement any algorithm. Esfandiari et al. proposed a multivariate filter feature selection method based on interaction-based feature clustering (IFC). Features are ranked based on the symmetric uncertainty criterion, and then clustering of features is performed by calculating their interactive weight as a measure of similarity (Esfandiari et al, 2023). Ghosh et al. proposed a wrapper-filter combination of ACO. It is based on the evaluation of subsets using a filtering method. They use a memory to keep the ants that have found the best results. The updating of pheromones is carried out

according to the dimension of the characteristics. The selection algorithm is based on a multi-objective function (Ghosh et al, 2020). Luo et al. proposed an algorithm based on the hybridization of the rough hyper cuboid approach and binary particle swarm optimization (BPSO) algorithm. A distributed algorithm is thus developed and embedded in the Apache Spark cloud computing model (Luo et al, 2022). Despite its importance in the pattern recognition process, the choice between the existing methods is sometimes impossible since the results often depend on the data used, so there are a large number of methods in the literature that share several common points. Therefore, it is preferable to carry out a comparative study between several methods to choose the method suitable for the problem treated. The dependence and the nature of the data are among the criteria most used to test and evaluate these methods.

4 METAHEURISTICS

In this section, we present the outline of some optimization methods which are Particle Swarm Optimization Algorithm (PSO), Bat algorithm (BA), Fish School Search Algorithm (FSS), Gorilla Troop Optimization algorithm (GTO), and Ant Colony Optimization algorithm (ACO).

4.1 Particle Swarm Optimization Algorithm (PSO)

Particle Swarm Optimization algorithm is inspired by the behaviour observed in groups of birds that allows self-organization. Its operating principle is based on three simple rules:

- Cohesion: the birds prefer to go to the center of the group;
- Alignment: the birds are guided by their surroundings;
- Separation: birds control the distance between them.

Initially the individuals are randomly distributed in the search space. Displacement represents possible solutions. These solutions are saved in a local memory and are communicated to the neighborhood in order to optimize each individual's solution. The solution set will allow the group to converge towards an optimal solution (Poli et al, 2007).

We can summarize the algorithm as follows:

To adapt this algorithm to the parameter selection problem, a binary string will represent the parameters.

Algorithm 1 Particle Swarm Optimization Algorithm.

Random initialization of the position and speed of N particles
Calculate the relevance of the positions of each particle

For all particles, repeat

- $\vec{g}\mathrm{best}(t+1) = \mathrm{argmin}_{\vec{p}\mathrm{best}_i} f(\vec{p}\mathrm{best}_i(t+1))\ i \leq 1 \leq N$
- $v_{ij}^{t+1} = wv_{ij}^{t} + c_1 r_{1ij}^{t}[pbest_{ij}^{t} - x_{ij}^{t}] + c_2 r_{2ij}^{t}[pbest_{j}^{t} - x_{ij}^{t}]\ , j \epsilon \{1,2, D\}$
- $v_{ij}^{t+1} = x_{ij}^{t} + v_{ij}^{t}\ , j \epsilon \{1,2, \dots D\}$

Update g best and p best using:

- $\vec{p}\mathrm{best}_i(t+1) = \begin{cases} \vec{p}\mathrm{best}_i(t) \text{ si } f(\vec{x_i}(t+1)) \geq \vec{p}\mathrm{best}_i(t) \\ \vec{x_i}(t+1), \text{sinon} \end{cases}$
- $\vec{g}\mathrm{best}(t+1) = \mathrm{argmin}_{\vec{p}\mathrm{best}_i} f(\vec{p}\mathrm{best}_i(t+1))\ i \leq 1 \leq N$

While maximum iterations or minimum error criteria is not attained

4.2 Bat algorithm (BA)

In the bat algorithm, each element has a fly velocity vi at position Xi with a fixed frequency range [fmin, fmax] velocity varying its emission rate r[0,1]and loudness A0 to search for prey, depending on the proximity of its target. The bat algorithm is based on the echolocation parameters of bats. The update of the position and the speed of each bat are defined according to the following rules (Mirjalili et al, 2014).

$$f_i = f_{min} + (f_{max} - f_{min})\beta \quad (1)$$

$$v_i^t = v_i^{t-1} + (x_i^t - x_*)f_i \quad (2)$$

$$x_i^t = x_i^{t-1} + v_i^t \quad (3)$$

A basic description of the bat algorithm is given as follows

Algorithm 2 Bat algorithm.

Initialize the group (fly velocity v_i, position x_i, frequence range f, pulse rate r and the loudness A_i)
Repeat
Generate a new solution by adjusting the frequency and update different parameters using update rules.
Generate a solution randomly
Evaluation the new solution
Update r_i and A_i
Rank the bats and find the best one

For each individual, a new solution is obtained based on the best solution of the previous iteration. A set of improvements have been introduced on the basal algorithm to allow selection of parameters. Parameters that are irrelevant or that can be deduced from other parameters are considered obstacles and removed immediately. Obviously these parameters will not be inspected in the following iterations.

4.3 Fish School Search Algorithm (FSS)

The algorithm (FSS) was proposed by Bastos Filho and Lima Neto in 2008. It was inspired by the behaviour of fish groups. The objective is to swim towards the positive gradient in order to "eat" and "gain weight". The group of group gravity changes depending on the number of fish that have caught a weight and will allow the group to change the research space from iteration to another. There are three movement operators: individual, collective-instinctive and collective-volatile (Pourpanah et al, 2023). The summary of the Binary Fish School Search algorithm for feature selection is given as follows:

Algorithm 3 Fish School Search Algorithm.

Initialize user parameters
Initialize fishes positions X_i randomly
Calculate the precision of the classification
While stopping condition is not met do
For Each fish on the school do
Update Fish position using

$$x_{i,j}\begin{cases}\overline{x_{i,j}} & \text{if } k < s_{ind}(t)\\ & \text{otherwise}\end{cases}$$

Calculate the individual movements using
$I = \frac{\sum_{i=1}^{N} \nabla x_i \nabla f_i}{\sum_{i=1}^{N} \nabla f_i}$ With I represent the weighted average of the displacements of each fish
Move each fish according to
$B(t) = \frac{\sum_{i-1}^{N} x_i(t) W_i(t)}{\sum_{i=1}^{N} W_i(t)}$ with B is the barycenter of the school
End for
End while

4.4 Gorilla Troop Optimization algorithm (GTO)

Optimization of gorilla troops Algorithm is based on five actions which are: to move to an anonymous region, another group, already visited, or follow the silver back and in competition for mature women. The aim of discovering a new area is to optimize the research space. The only condition is to keep a balance between exploitation and exploration (Tiwari et al, 2023).

$$GX(t+1) = \{(\mathrm{UB}-\mathrm{LB}) \times \mathrm{r}_1 + \mathrm{LB}, \quad \text{rand} < \mathrm{p}, (\mathrm{r}_2 - \mathrm{C}) \times \mathrm{X_r(t)} + \mathrm{L} \times \mathrm{H}, \quad \text{rand} \geq \quad 0.5\, \mathrm{X(i)} - \mathrm{L} \times (\mathrm{X(i)} - \mathrm{GX_r(i)}) + \mathrm{r}_3 \times (\mathrm{X(i)} - \mathrm{GX_r(i)}), \quad \text{rand} < 0.5\,\} \quad (4)$$

The update of each group member is calculated according to the preceding equation. The use of random behaviour by Gorillas only concerns a limited number of individuals in the group. This action is essential to discover new solutions and it does not influence the approximation of the algorithm towards the overall minimum.

$$C = F \times \left(1 - \frac{It}{MaxIt}\right) \quad (5)$$

$$F = cos\,(2 \times r_4) + 1 \quad (6)$$

$$L = C \times l \quad (7)$$

$$H = Z \times X(t) \quad (8)$$

$$L = C \times l \quad (9)$$

The number of iteration and the probability of random behaviour are initialized by the user. An update is required when the algorithm finds a solution with me with previous iterations.

$$GX\,(t+1) = X_{silverback} - (X_{silverback} \times Q - X(t) \times Q) \times A \quad (10)$$

$$Q = 2 \times r_5 - 1 \quad (11)$$

$$\mathrm{A} = \beta \times \mathrm{E} \quad (12)$$

$$E = \{N_1, rand \geq 0.5\; N_2, rand < 0.5\,\} \quad (13)$$

Q represents the impact force. At the end of this step, we compare GX (t) to X (t), if the objective value of GX (t) is less than X (t). The total number of attributes is the same number of entries in the GTO algorithm. At this stage, we calculate the retention of each attribute and its probability of appearing in the final subset of selected parameters.

$$Fitness_i = a \times (1 - C_i) + (1 - a) \times \frac{|BX_i|}{D} \quad (14)$$

There are several formulas to calculate the value of C. For the case of selection of parameters, we can use the Fisher formula. The D variable is relative to the training data interval. An increase function is used to determine the solution. Slight modifications are made at the end of each iteration to adjust all the variables. We can use two stop criteria; the easiest way is to set the maximum number of iterations. We can also use the value returned by an objective function. This value is compared by precedent values. In the event that there is no improvement, the algorithm stops.

4.5 Ant Colony Optimization algorithm (ACO)

Ants are characterized by great diversity in their means of communication. Their behaviour represented a rich source of inspiration for researchers. The majority of algorithms consider the colony as a multi-agent system and each ant as a reactive agent. A reactive agent is a non -intelligent software entity that uses its environment to communicate with other members of the colony. These simple actions result in intelligent behaviour of the colony. Ants have several means of communication. The first type is sound communication which is obtained by friction produced by an organ placed on the abdomen. The sound is generated to signal: distress, quality of food, strengthening or a threat. The second type is tactile communication that is used to transmit simple information using touching, tapping or touching. The third type is visual communication. An experience has shown that the ANTS identified the shortest path in a labyrinth with a set of drawings (round/cross, star/square, rectangle/triangle, and diamond/oval). The ants found the solution quickly without any error by memorizing the relationship between the quality of the path and each symbol represented in the way. The modification introduced to the operating principle of anti -colony algorithms resolved several problems. To apply this type of method, the system must have two district states and probabilistic transition rules. To apply ACO to the selection of parameters, we use one -dimensional vectors that affects a binary value for each scalar. The value 1 indicates the consideration of the parameter in the sub-set of selected parameters. An evaluation function uses this set to calculate the relevance of the solution (Kadri et al, 2012).

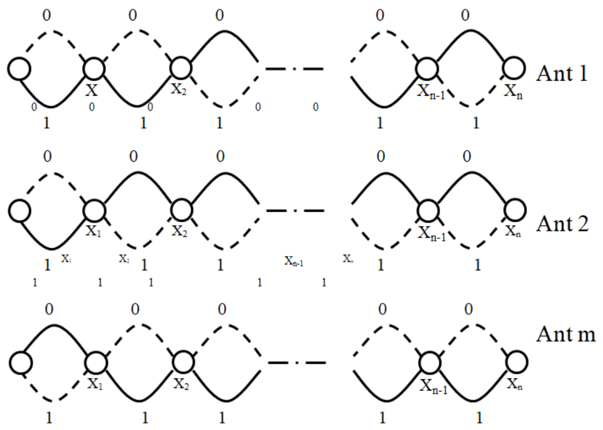

Fig. 4. Feature subset selection using Binary ACO.

Each ant selects a parameter using a probabilistic law called PS. It depends on two other variables. The first is the FP pheromone factor and the second is the FH heuristic factor which makes it possible to classify the parameters in order of importance.

$$PS_{i1} = \frac{FP_{i1} + \dfrac{FP_{i0}}{Max(FH)} FH_i}{FP_{i1} + FP_{i0}} \quad (15)$$

The pheromone update depends on the best solutions obtained during the last iteration and in all iterations. Updating the amount of pheromone placed on the path of each ant is given as follows:

$$\Delta \mathrm{FP} = \frac{1}{1 + \mathrm{F}(\mathrm{V}) - \mathrm{F}(\mathrm{V}')} \quad (16)$$

5 ACO FOR FEATURE SELECTION

In this section, we present a parameter selection algorithm in the field of anomaly detection in industrial processes and in the detection of attacks in a computer network.

5.1 Detection of network attacks using ACO

5.1.1 Dataset

The dataset represents a burst header packet flooding attack on a burst-switched optical network dataset available in the machine learning repository. The description and contents of this database are published on UCI (Seddik et al, 2021). The size of this base is relatively small. It only consists of 1075 observations distributed over 4 classes. But it is sufficient to carry out a complete study of four types of attacks which are NB-No Block, Block, No Block and NB-Wait. Each row includes 22 attributes. We can use this database to create an imputation system since several lines contain missing values.

5.1.2 Features selection

To select the best parameters, we calculated the relevance of each parameter individually and its correlation with the other parameters. The classifier is used to evaluate each result. The ACO algorithm makes it possible to speed up the search process and inspect only a small portion of the search space. ACO also makes it possible to optimize the classifier by finding the parameter values of the kernel function suitable for the database used. The values of the hyper-parameters that improve the classification accuracy are (C, γ) is (23, 2-5). Based on our previous work, we have limited the search space on the intervals [23, 211], [2-12, 22]. The following table presents the values of six parameters. Using these parameters, we can obtain results close to those found using the initial 22 parameters.

Table 3 Instances of the obtained dataset.

Packet Drop Rate	Reserved Bandwidth	10-Run-AVG-Drop-Rate	10-Run-AVG-Bandwidth	10-Run-Delay	Node Status'
1000	0	0.7	0	B	0
100	0	0.2	0	NB	0.4
900	0	0.8	0	B	0
100	0	0.3	0	NB	0.4
800	0	0.8	0	B	0
100	0	0.4	0	NB	0.2
700	0	0.8	0	B	0
100	0	0.5	0	PNB	0.1
900	0	0.8	0	B	0
100	0	0.4	0	PNB	0.1

In detail, we used a set of Python functions to change the data format from ARFF to CSV format. We have restructured the data as Data Frames. Using functions from the NumPy library, we have eliminated unnecessary information for the classification process.

Using the SVM classifier, we obtained two different results which are related to the type of kernel used. The Gaussian kernel returns classification accuracy above 91%. On the other hand, the best results obtained with the linear kernel did not exceed 70%.

This large difference in accuracy between the classifiers shows that the relationships between the data are not always direct and linear. These results also prove that to find the ideal classifier for a database, we have to carry out several tests and, in each time, we have to use a different subset of data.

5.2 Fault diagnosis of rotary kiln System

This system is used to calculate the raw materials necessary to make cement. The essential element of this system is the rotary kiln. It is made up of a shell, a thermal lining, driving rings and an internal thermal exchanger.

5.2.1 Dataset

In this study, we used 500 data records distributed over two classes representing normal operating and the malfunction of an industry system. The number of parameters of the rotary kiln system is 47 (Kadri et al, 2012).

We used a reactive multi-agent system made up of 20 entities. The majority of agents (80%) follow the same movement law and the rest of ants move randomly in the research space in order to avoid the convergence of all solutions to a local optimum.

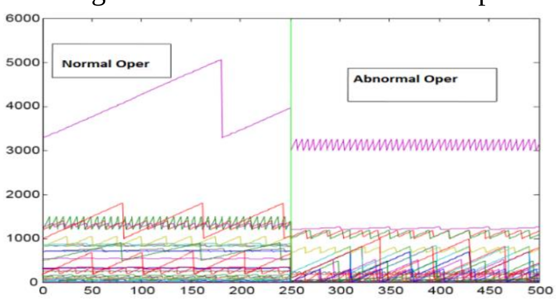

Fig. 5. RCK dataset of rotary kiln System.

5.2.2 Classification

We have calculated the accuracy of classification by using three different entries. According to the following table, we conclude that ACO has eliminated the irrelevant parameters and only keeps the most important. We also note that the elimination of a large number of parameters does not influence the classification rate and by consequence will increase the recognition speed.

Table 4 Performances of classification by using the various entries.

Features	RCK	
	accuracy	F (V)
Generated subset	89 %	0.7700
One feature	25 %	0.0325
All features	93 %	0.7875

The selected subset feature, that produces those results, is presented in table 5.

Table 5 Description of selected feature subset.

umber	Code	Description
	A54P2	Cyclone Pressure A54t.
	A50T1	Cyclone gas outlet temperature A50
	U01T1	Clinker temperature
	A54T2	Material temperature cyclone A54
	TV	Kiln shell temperature
	A53T1	Temperature gas cyclone A53
	K01T1	Secondary air temperature
	COC	Cyclone outlet CO content A50
	A54T1	Temperature gas cyclone A54
	A52P2	Cyclone Pressure A52.
	V07P1	Primary air pressure
	A53T2	Material temperature cyclone A53
	O2C	Content O2 output cyclone A50
	W01S1	Speed oven
	COP	CO content smoke box
	A53P1	Cyclone Pressure A53.
	A52T1	Cyclone gas outlet temperature A52
	V31F1	Gas flow
	V01F1	Gas flow
	W01X1	Oven time

5.3 Fault diagnosis for a milk pasteurization plant with missing data

5.3.1 Dataset description

This third study proves the effectiveness of metaheuristic methods in the diagnostic process. The database used contains monitoring observations for a milk pasteurization system. It is made up of 300 records which represent two different operating states. The different steps of process studied are summarized in the following figure:

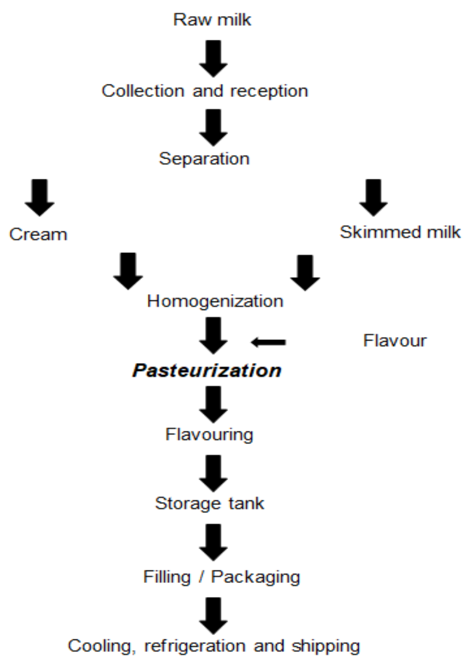

Fig. 6. Milk production scheme.

The step of pasteurization consists of four main operations which are illustrated in the following table.

Table 6 The different pasteurization operations.

Operation	Designation	Purpose	Observation
OP1	Hot water production	Heat the water which in turn will heat the milk	The steam arrives at the exchanger level will heat water
OP2	Heating the milk	Destroy pathogenic	The pasteurization temperature is about 75 °C to 90 °C
OP3	Preheating milk	The hot milk will be used for its own preheating	Heat recovery, energy consumption
OP4	Cooling milk	Cool the milk	Finished product

We have applied the previous algorithm on a pasteurization system to select the best set of parameters to ensure the diagnostic function. The description of selected parameters is cited in the following table.

Table 7 The selected parameters of pasteurization system.

N° Feature	Feature symbol	Feature Description.
P1	TP	Temperature of input product
P2	FP	Product outflow
P3	Th	Heating temperature
P4	Fw	Iced water flow
P5	Tw	Iced water temperature
P6	Tc	Cooling temperature
P7	Pc	Compressed air pressure

5.3.2 Imputation

In this third studies we have applied three imputation methods on the database containing only the attributes selected by the ACO algorithm.

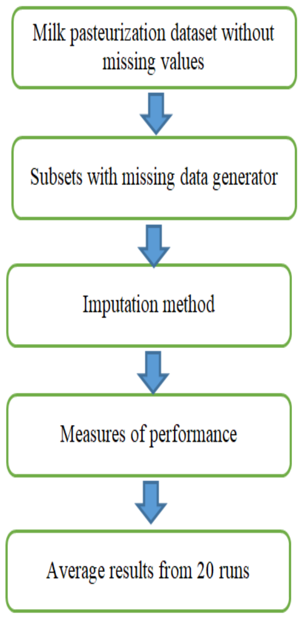

Fig. 7. Principle of missing data analysis.

After the deletions of the lines correspond to the attributed not selection by the ACO algorithm, several datasets are created with missing values varied between 5% and 50% from this data set. We have applied an evaluation criterion which consists in comparing the values predicted to real values. The results obtained are illustrated in the following figure:

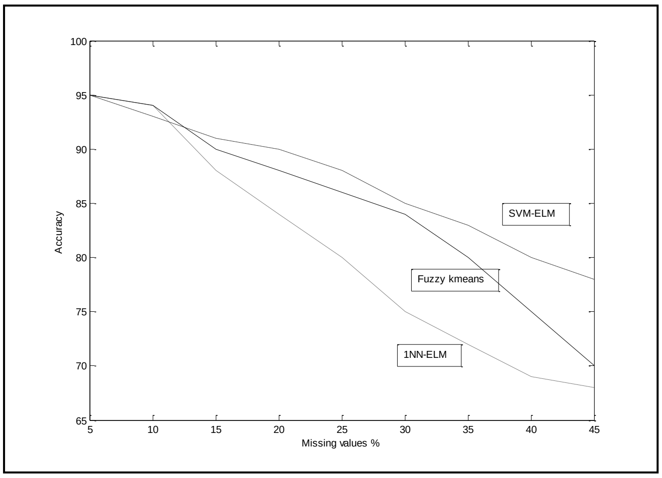

Fig. 8. Accuracy of the three algorithms over different percentages of missing data.

6 CONCLUSION

In this article, we have presented recent work in the field of parameter selection using metaheuristics. We have explained in detail the different diagnostic stages and the importance of the parameter selection phase in this process. We have exposed its importance to improve the learning and identification processes. We have also presented its importance in other applications such as the imputation of missing values. At the end of this article, we studied three different problems. The first study concerns the detection of attacks in a computer network. In the other two studies, we have analyzed pasteurization and Clinkerization systems. Thanks to the example presented, we have demonstrated the importance of the application of metaheuristics inspired by the behaviour of living beings for the selection of parameters in two industrial problems and an information security problem. For each metaheuristics presented, we have offered a version for parameter selection. For the ACO algorithm, we have proposed a new movement rule and we have analyzed the influence of random behaviour on the performance of the algorithm. It will be very important to apply the algorithms presented on Big Data datasets issued from various industrial problems. It will be a great challenge for researchers to offer hybrid methods that combine this type of algorithms with reinforcement learning.

7 ACKNOWLEDGMENTS

This work was supported by a grant from the Automation and Production Laboratory, PRFU, project no. A14N01UN050220230006.

8 REFERENCES

Dyukov, V. A., Kazantsev, A. M., Makushev, I. Y., et al. (2023). "Optimization of the Risk Functional to Control the Composition and Structure of a Heterogenous Grouping of Detection Sensors in Three-Dimensional Space." *Journal of Computer and Systems Sciences International.* https://doi.org/10.1134/S1064230723020077

Kaprielova, M. S., Leonov, V. Y., & Neichev, R. G. (2022). "Recognition of Human Movements From Video Data." *Journal of Computer and Systems Sciences International, 61*, 233–239. https://doi.org/10.1134/S1064230722020095

Kadri, O., Benyahia, A., & Abdelhadi, A. (2022). "Tifinagh Handwriting Character Recognition Using a CNN Provided as a Web Service." *International Journal of Cloud Applications and Computing (IJCAC), 12*(1), 1-17. https://doi.org/10.4018/IJCAC.297093

Cheeseman, T., Southerland, K., Park, J., et al. (2022). "Advanced Image Recognition: A Fully Automated, High-Accuracy Photo-Identification Matching System for Humpback Whales." *Mammalian Biology, 102*, 915–929. https://doi.org/10.1007/s42991-021-00180-9

Hashemi, A., Dowlatshahi, M. B., & Nezamabadi-pour, V. (2022). "Ensemble of Feature Selection Algorithms: A Multi-Criteria Decision-Making Approach." *International Journal of Machine Learning and Cybernetics, 13*, 49–69. https://doi.org/10.1007/s13042-021-01347-z

ABDELHADI, A., & KADRI, O. (2023). Transformation Of 2d Images Into 3d By The Deeplearning. *Academic Journal of Manufacturing Engineering*, 21(2), 116-123.

https://www.ajme.ro/content.php?vol=21&year=2023&issue=2&offset=10

Haq, A. U., et al. (2021). "Detection of Breast Cancer Through Clinical Data Using Supervised and Unsupervised Feature Selection Techniques." *IEEE Access, 9,* 22090-22105. https://doi.org/10.1109/ACCESS.2021.3055806

Liang, Y., Liu, Q., Xia, T., Ye, C., & Li, J. (2022). "Preventive Maintenance Strategy Optimization in Manufacturing System Considering Energy Efficiency and Quality Cost." *Energies, 15*(21), 8237. https://doi.org/10.3390/en15218237

Shu, F., & Shu, J. (2021). "An Eight-Camera Fall Detection System Using Human Fall Pattern Recognition via Machine Learning by a Low-Cost Android Box." *Scientific Reports, 11,* 2471. https://doi.org/10.1038/s41598-021-81115-9

Zhang, J., Wang, Y., Jiang, B., et al. (2023). "Realistic Fault Detection of Li-Ion Battery via Dynamical Deep Learning." *Nature Communications, 14,* 5940. https://doi.org/10.1038/s41467-023-41226-5

Kadri, O., Mouss, L. H., & Mouss, M. D. (2012). "Fault Diagnosis of Rotary Kiln Using SVM and Binary ACO." *Journal of Mechanical Science and Technology, 26,* 601–608. https://doi.org/10.1007/s12206-011-1216-z

Rithani, M., Kumar, R. P., & Doss, S. A. (2023). "Review on Big Data Based on Deep Neural Network Approaches." *Artificial Intelligence Review, 56,* 14765–14801. https://doi.org/10.1007/s10462-023-10512-5

Infantraj, A., & Kumaran, M. S. (2023). "Evaluation and Classification of Stator Turn-to-Turn Faults Using Electrical Equivalent Circuits for Surface Permanent Magnet Brushless Direct Current Motors." *Journal of Power Electronics.* https://doi.org/10.1007/s43236-023-00663-7

Quaranta, L., Calefato, F., & Lanubile, F. (2021). "KGTorrent: A Dataset of Python Jupyter Notebooks from Kaggle." *IEEE/ACM 18th International Conference on Mining Software Repositories (MSR),* Madrid, Spain, 550-554. https://doi.org/10.1109/MSR52588.2021.00072

Jiang, W., et al. (2021). "Statistical Feature Extraction and Hybrid Feature Selection for Material Removal Rate Prediction in Chemical Mechanical Planarization Process." *5th IEEE Electron Devices Technology & Manufacturing Conference (EDTM),* Chengdu, China, 1-3. https://doi.org/10.1109/EDTM50988.2021.9421002

Jiang, C., Chen, T., & Tao, G. (2023). "Gap Image Classification Based on PCA-SVM with Multiple Feature Fusion." *4th International Conference on Mechatronics Technology and Intelligent Manufacturing (ICMTIM),* Nanjing, China, 609-613. https://doi.org/10.1109/ICMTIM58873.2023.10246511

Ghosh, M., Guha, R., Sarkar, R., et al. (2020). "A Wrapper-Filter Feature Selection Technique Based on Ant Colony Optimization." *Neural Computing and Applications, 32,* 7839–7857. https://doi.org/10.1007/s00521-019-04171-3

Esfandiari, A., Farivar, F., & Khaloozadeh, H. (2023). "Fractional-Order Binary Bat Algorithm for Feature Selection on High-Dimensional Microarray Data." *Journal of Ambient Intelligence and Humanized Computing, 14,* 7453–7467. https://doi.org/10.1007/s12652-022-04450-3

Luo, C., Wang, S., Li, T., Chen, H., Lv, J., & Yi, Z. (2022). "Large-Scale Meta-Heuristic Feature Selection Based on BPSO Assisted Rough Hypercuboid Approach." *IEEE Transactions on Neural Networks and Learning Systems.* https://doi.org/10.1109/TNNLS.2022.3171614

Poli, R., Kennedy, J., & Blackwell, T. (2007). "Particle Swarm Optimization." *Swarm Intelligence, 1,* 33–57. https://doi.org/10.1007/s11721-007-0002-0

Mirjalili, S., Mirjalili, S. M., & Yang, X. S. (2014). "Binary Bat Algorithm." *Neural Computing and Applications, 25,* 663–681. https://doi.org/10.1007/s00521-013-1525-5

ABDELHADI, A.**,** MOUSS, L. H**.**, & KADRI, O**.** (2020). Hybrid Multi-Agent And Immune Algorithm Approach To Hybrid Flow Shops Scheduling With Sdst. *Academic Journal of Manufacturing Engineering,* 18(3), 120-130. https://ajme.ro/PDF_AJME_2020_3/L15.pdf

Pourpanah, F., Wang, R., Lim, C. P., et al. (2023). "A Review of Artificial Fish Swarm Algorithms: Recent Advances and Applications." *Artificial Intelligence Review, 56,* 1867–1903. https://doi.org/10.1007/s10462-022-10214-4

Tiwari, P., Raj, S., & Chhimwal, N. (2023). "Community Detection in Network Using Chronological Gorilla Troops Optimization Algorithm with Deep Learning Based Weighted Convexity." *Wireless Networks, 29,* 3809–3828. https://doi.org/10.1007/s11276-023-03430-5

Seddik, M. T., Kadri, O., Bouarouguene, C., & Brahimi, H. (2021). "Detection of Flooding Attack on OBS Network Using Ant Colony Optimization and Machine Learning." *Computación y Sistemas, 25,* 423–433. https://doi.org/10.13053/CyS-25-2-3939