ISSN 2007-9737

A Binary Walrus Optimizer for Feature Selection Problem

Seyyid Ahmed Medjahed[1,*], Fatima Boukhatem[2]

[1] University of Relizane,
Algeria

[2] University of Djilali Liabes, Sidi Belabes,
Algeria

seyyidahmed.medjahed@univ-relizane.dz, fatima.boukhatem@univ-sba.dz

Abstract. Feature selection is a very crucial step in machine learning. It plays an important role for enhancing model performance, interpretability and efficiency. The main goal of feature selection is to select the optimal subset of features which is considered as the relevant features. In this paper, we introduce a novel feature selection approach based on a novel optimization approach called Walrus Optimization Algorithm (WOA) proposed by Muxuan Han et al. in 2023 and never tested in the context of feature selection. A new binary version of Walrus Optimization is proposed and adapted to the problem of feature selection. The fitness function is composed of three important terms: classification accuracy rate, correlation and class separability measure based on Jaccard Index. To evaluate the performance of the proposed approach, five synthetic datasets was used: CorrAl, m-of-n-3-7-10, Monk1, Monk2 and Monk3. In addition, the approach was tested on real-world DNA microarray datasets: colon cancer, leukemia, breast cancer ovarian cancer, lung cancer and DLBC cancer (diffuse large B-cell lymphoma). The results demonstrate that the proposed can produce a high classification accuracy rate and a good diagnostic of cancer.

Keywords. Feature selection, Walrus optimizer, classification, gene selection, Jaccard index.

1 Introduction

Recently, machine learning has become a very important field in many applications such as, biology, environment, military, medical diagnosis, etc. In the learning phase, choosing the right information can make a big difference and can significantly improve the quality of the learning system.

That is why, feature selection can help to select the most important parts of the data [4, 18]. It provides an optimal model performance and interpretability. By selection the relevant features, the accuracy will significantly be influenced [12, 14].

Feature selection can be defined as the process of machine learning where the optimal subset of features is selected from a larger set of features. The aim is to improve the performance of a learning model and focusing only on the informative data by removing the irrelevant and redundant features.

The most advantages of using feature selection methods are enhancing the model accuracy, reducing overfitting, improving interpretability, reducing the complexity and reducing the huge phenomena. Feature selection approaches can be classed into three main types [22]. The first one is Filter approach which evaluates the optimal subset of features independently of the learning algorithm. It is based on statistical measures and rank features. The second one is Wrapper approach which uses learning algorithm to select the features by directly measuring how well a particular classifier performs with a subset of features. The last category is the Embedded approach which incorporates the process of feature selection as a part of the training process [23, 20, 1].

doi: 10.13053/CyS-29-2-5103

Many feature selection approaches have been proposed in previous studies to enhance the effectiveness and efficiency of machine learning models. In [11], the authors proposed a new feature selection approach based on multiple feature subsets extraction and result fusion for enhancing the classification output.

The proposed approach was tested on 20 well-known datasets. Zhiwen et al. [5] proposed an approach for feature selection using a new insight to construct graphs that can accommodate multi-order neighbor information.

In [10], the authors used an improved version of Salp swarm and Support vector machine for feature selection. In [17], an unsupervised feature selection is proposed based on data structure learning and feature orthogonalization.

Medjahed et al. [15] proposed a new band selection approach based on an hybrid optimization algorithm which is Salp Swarm Algorithm with threshold accepting.

In this paper, a new feature selection approach is proposed. This approach is based on a new meta-heuristic named Walrus Optimization Algorithm which has been never tested on the problem of feature selection.

A new binary version of Walrus Optimization Algorithm is proposed. Additionally, we propose a new fitness function to improve the quality of the optimal subset of selected features. The fitness function includes three terms: classification accuracy rate, correlation and Jaccard distance. The proposed approach is tested on five synthetic datasets widely used in the context of solving feature selection problem and six DNA microarray datasets which are a real-world data.

The rest of the paper is organized as follows: In section 2, we present the proposed approach and review the Walrus Optimization Algorithm. Section 3 detailed the experimental parameters and the datasets used to demonstrate the performance of the proposed approach. Results and discussion are presented in Section 3. Finally, in section 4, we draw conclusions of our study.

2 The Proposed Feature Selection Approach

The feature selection problem can be formulated as a binary optimization problem which each value represents the presence or the absence of feature in the training model. Mathematically, this can be defined as a process of finding the optimal binary vector that optimized a predefined objective function.

By using an optimization algorithm, this last explores the space of solution using the principles of the optimization algorithm. By adapting these principals to the binary feature selection problem, the optimization algorithm can find the subset of features that optimize the objective function.

Let's $D = (x_i, y_i)$ a dataset, where x_i is the instance ith and y_i is the class of the instance ith. Let's $Fe = \{fe_1, \ldots, fe_2\}$ set of features of the dataset D, we define the binary vector $B = \{b_1, .., b_N\}$ with $b_i = 1$ represents the presence of the feature ith in the training model and if $b_i = 0$ represents the absence of the feature ith from the training model. The optimization problem is to find the optimal vector B.

2.1 Walrus Optimization

The Walrus Optimization Algorithm is a new meta-heuristic inspired by the behaviors of walruses. The behaviors are: migrate, breed, roost, feed, gather and escape. These behaviors are based on a key signal received by the walruses that can be a safety signal or a danger signal.

Note that the walrus is considered as the largest mammal in the ocean. They live in group of dozen to hundred or thousands of individuals [8, 13].

The authors proposed a meta-heuristic based on the behaviors of walrus in migrating, breeding, roosting and foraging. The mathematical model can be defined as follows:

2.1.1 Initialization

Initially, the optimization algorithm starts with a random solution [8]:

$$X = LB + \mathrm{rand}(UB - LB), \tag{1}$$

where LB and UB are the lower and upper boundary of the problem and rand is a function that generates random number between $[0, 1]$.

The position of Walruses are updated in each iterations [8]:

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & ... & X_{1,d} \\ X_{2,1} & X_{2,2} & ... & X_{2,d} \\ ... & ... & ... & ... \\ X_{n,1} & X_{n,2} & ... & X_{n,d} \end{bmatrix}, \tag{2}$$

where n is the population size and d is the dimension of decision variables.

The fitness values are defined as [8]:

$$F = \begin{bmatrix} f_{1,1} & f_{1,2} & ... & f_{1,d} \\ f_{2,1} & f_{2,2} & ... & f_{2,d} \\ ... & ... & ... & ... \\ f_{n,1} & f_{n,2} & ... & f_{n,d} \end{bmatrix}, \tag{3}$$

90% of walrus polulations are adults and the remaining 10% are juveniles.

2.1.2 Danger Signal and Safety Signal

In a danger situation, a danger signal will be sent between walruses which is defined with the following equation [8]:

$$danger\_signal = A \times R, \tag{4}$$

$$A = 2 \times \alpha, \tag{5}$$

$$R = 2 \times r_1 - 1, \tag{6}$$

A and R are dangers factors

α starts with 1 and in each iteration decreases to 0 with the number of iteration t

T is the maximum number of iterations

r_1 is random number between $[0, 1]$

The safety signal is defined as follows [8]:

$$safety\_signal = r_2, \tag{7}$$

r_2 is random number between $[0, 1]$

2.1.3 Migration

The migration represents the exploration phase in the algorithm. Walruses will migrate when the risk factors are very high. For migration, the algorithm updates the position of walrules using the following equation [8]:

$$X_{i,j}^{t+1} = X_{i,j}^{t} + migration\_step, \tag{8}$$

$$migration\_step = (X_m^t - {}_n^t) \cdot \beta \cdot r_3^2, \tag{9}$$

$$\beta = 1 - \frac{1}{1 + \exp(-\frac{t - \frac{T}{2}}{T} \times 10)}, \tag{10}$$

$X_{i,j}^{t+1}$ is the new position for the ith walrus on the jth dimension

migration_step is the step size of walrus movement

β is the migration step control factor

r_3 is random number between $[0, 1]$

2.1.4 Reproduction

Roosting behavior The reproduction represents the exploitation phase of the algorithm. When the risk factors are low, the walrus herds tend to breed [8].

Note that the female, male and juvenile walruses have different behaviors. Their positions are updated in different way. The female are influenced by the male and the lead walrus [8].

The position of the male is updated using the Halton sequence. The basic idea is to split the search are into several areas and selection a random point in each area [8].

The position of the female can be described as follows:

$$Female_{i,j}^{t+1} = Female_{i,j}^{t} + \alpha \cdot (Male_{i,j}^{t} - Female_{i,j}^{t}) +$$

(1-α) $\cdot (X_{best}^{t} - Female_{i,j}^{t}),$ (11)

$Female_{i,j}^{t+1}$ is the new position of the ith female on the jth dimension

The position of juvenile is defined as follows [8]:

$$Juvenile_{i,j}^{t+1} = (O - Juvenile_{i,j}^{t}) \cdot P, \tag{12}$$

$$O = X_{best}^{t} + Juvenile_{i,j}^{t} \cdot LF. \tag{13}$$

$Juvenile_{i,j}^{t+1}$ is the new position of the ith juvenile on the jth dimension

P is the distress coefficient which is random number between $[0, 1]$

LF is vector of random number generated using Lévy distribution [8]:

$$Levy(a) = 0.05 \times \frac{x}{|y|^{\frac{1}{a}}}. \tag{14}$$

x and y are variables normally distributed, $xN(0, \sigma_x^2)$, $yN(0, \sigma_y^2)$ and σ is the standard deviations.

Foraging Behavior In the foraging behaviors we distinct two behaviors: fleeing and gathering behaviors [8].

a. Feeling Behaviors

During underwater foraging, walruses can be attacked by predators. In this case, walruses will flee from their area based on danger signals from their peers [8]. This behavior is described as follows:

$$X_{i,j}^{t+1} = X_{i,j}^{t} \times R - \left|X_{best}^{t} - X_{i,j}^{t}\right| * r_4^2, \tag{15}$$

where $\left|X_{best}^{t} - X_{i,j}^{t}\right|$ is the distance between the current walrus and the best walrus

r_4 is random number between $[0, 1]$

b. Gathering Behavior

Walruses can share their position and move according to other walruses to find the sea are with higher food [8]:

$$X_{i,j}^{t+1} = \frac{X_1 + X_2}{2}, \tag{16}$$

$$\begin{cases} X_1 = X_{best}^{t} - a_1 \times b_1 \times \left|X_{best}^{t} - X_{i,j}^{t}\right| \\ X_1 = X_{\sec ond}^{t} - a_2 \times b_2 \times \left|X_{\sec ond}^{t} - X_{i,j}^{t}\right| \end{cases}, \tag{17}$$

$$a = \beta \times r_5 - \beta, \tag{18}$$

$$b = \tan(\theta). \tag{19}$$

X_1 and X_2 are two weights affecting the gathering behavior:

a and β are the gathering coefficients

r_5 is random number between $[0, 1]$,

θ number in the range $[0, \pi]$.

The exploration phase and the exploitation phase are detemined using the danger signal. If the danger signal is greater than 1, the warluses migrate to a new area which represents the exploration phase in the algorithm. Otherwise, the warluses herd reproduce which represents the exploitation phase in the algorithm [8].

We propose a binary version of Walrus Optimizer. A binary version of a meta-heuristic algorithm is an algorithm which the decision variables are binary. In the proposed binary walrus optimization, we use the sigmoid function to convert continuous values to binary values.

The sigmoid function maps any real number to the range (0, 1), making it suitable for converting continuous values into binary values. The binary conversion typically involves thresholding the output of the sigmoid function:

$$B = \frac{1}{1 + \exp(-X)}. \tag{20}$$

2.2 Fitness Function

In feature selection technique, the fitness function is a measure used to evaluate the candidate subset of features. It has an important role in guiding the optimization algorithm to find the optimal subset of features. The proposed fitness function is composed of three terms.

2.2.1 Accuracy Term

The accuracy term is based on the Support Vector Machine classifier (SVM). It represents an important term because it helps to find the subset of features which reduces the error rate. The term is defined as follows:

$$R(X) = classification\_accuracy\_rate(X), \quad (21)$$

$R(X)$ is the classification accuracy rate obtained by SVM using all the selected features X .

2.2.2 Correlation Term

The correlation term of the fitness function is the correlation between the selected features. This term is given as follows:

$$C(X) = 1 - \left[\frac{1}{N \times (N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} C_p(X_i, X_j).\right. \quad (22)$$

$C(X)$ is the correlation between all the selected features.

$C_p(X_i, X_j)$ is the Pearson correlation coefficient between the features X_i and X_j.

We add $(1-)$ to convert the problem to a problem of maximazing $C(X)$.

2.2.3 Class Separability Term

The term of class separability is a measure that computes the ability of a feature how to well separate the classes. In this study, we propose to use Jaccard similarly which is a measure used to compute the similarity between two sets by comparing their intersection to their union. Let's consider C classes, the Jaccard index can be defined as follows [9]:

$$Jaccard\_Index(X_k^{c_i}, X_k^{c_j}) = \frac{\left|X_k^{c_i} \cap X_k^{c_j}\right|}{\left|X_k^{c_i} \cup X_k^{c_j}\right|}, \quad (23)$$

where $X_k^{c_i}$ and $X_k^{c_j}$ are respectively set of instances belonging to class c_i and class c_j,

$$Jaccard(X_k) = \frac{1}{C \times (C-1)} \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} Jaccard\_Index(X_k^{c_i}, X_k^{cj}), \quad (24)$$

$$J(X) = \sum_{i=1}^{N} Jaccard(X_i). \quad (25)$$

The final form the fitness function is defined as follows:

$$F(X) = \alpha \times R(X) + \beta \times C(X) + \gamma \times J(X), \quad (26)$$

where α, β, γ are the weight coefficients.

2.3 Binary Walrus Optimizer for Feature Selection

The proposed algorithm Binary Walrus Optimizer for Feature Selection is described in Algorithm 1.

3 Experimental Results

Within this section, we showcase the outcomes of our experimentations. The assessment of performance revolves around the classification accuracy rate obtained by the proposed approach over the test phase. We conducted two separate experiments. The first experimentation is conducted under synthetic dataset and the second experimentation is applied under microarray datasets.

3.1 Application on Synthetic Dataset

In this first experimentation, we use artificial datasets to evaluation the performance of our approach.

These datasets are artificial benchmark datasets widely used to test the performance of feature selection approach.

We consider five synthetic datasets named $CorrAL$, m-of-n-3-7-10, Monk1, Monk2 and Monk3 [19]. They can be downloaded this web site[1].

---

[1] www.sgi.com/tech/mlc/db/

Algorithm 1 Binary Walrus Optimizer for Feature Selection

```
Initialize randomly the walrus population $X(i,j)(i = 1,..walrusesagents, j = 1,..,dimension)$
Initialize the parameter $C$ of SVM and Gaussian Kernel parameter
Divide dataset into three subsets: training $Tr$, validation $V$, testing $Ts$
for $t = 1 : T$ do
    if $danger_signal >= 1$ then
        Update position of walruses using 8
    else if $safety_signal >= 0.5$ then
        for $eachmalewalrus$ do
            Update position of walruses using the Halton sequence
        end for
        for $eachfemalewalrus$ do
            Update position of walruses using 2.1.4
        end for
        for $eachjuvenilewalrus$ do
            Update position of walruses using 12
        end for
    else if $danger >= 0.5$ then
        Update position of walruses using 15
    else
        Update position of walruses using 6
    end if
    Update the walrus position
    Convert continuous values to binary values using 20
    Generate randomly training set from $Tr$ and validation set from $V$
    Calculate the fitnes function using 26
end for
Calculate the classification accuracy rate using the new features and the test set $Ts$
```

$CorrAL$ is the first dataset used in the experimentation. $CorrAL$ is composed of 6 features (a_0, a_1, b_0, b_1, Irr, $Corr$) which defined by the target concept $(a_0 \cap a_1) \cup (b_0 \cap b_1)$. The two last features (Irr, $Corr$) represent respectively the irrelevant features and the highly correlated features.

The dataset m-**of**-n-3-7-10 is the second artificial dataset considered in the experimentation.

Table 1. Classification accuracy rate (%) and the selected features obtained by the proposed approach for each artificial dataset

Datasets	Selected features	Classification accuracy rate
CorrAL	1, 2, 3, 4, 6	95.98
m-of-n-3-7-10	2, 4, 5, 6, 7, 8, 9	100
Monk1	1, 2, 5	98.66
Monk2	1, 2, 3, 6	71.12
Monk3	2, 4, 5	87.06

m-**of**-n-3-7-10 is composed of 10 features which contains 3 irrelevant features. Features 2, 3, 4, 5, 6, 7, 8 are relevant to the class label.

Monk1, **Monk**2 and **Monk**3 are three artificial dataset that contain 6 features (a_1,. . . ,a_6). The target concept of each dataset is defined as follows:

— Monk1: $(a_1 = a_2) \vee (a_5 = 1)$. Relevant features are: a_1, a_2, a_5,

— Monk2: $(a_n = 1)$ for exactly two choices of n. Relevant features are: a_1, a_2, a_3, a_4, a_5, a_6,

— Monk3: $(a_5 = 3 \wedge a_4 = 1) \vee (a_5 \neq 1 \wedge a_2 \neq 3)$. Relevant feature are: a_2, a_4, a_5.

The proposed approach's outcomes for each artificial dataset, including the selected features and corresponding classification accuracy rates, are presented in Table 1.

Table 1 illustrates the classification accuracy rate obtained by the proposed approach under the five synthetic datasets.

As indicated in Table 1, when applied to the $CorrAL$ dataset, our proposed algorithm identified a_0, a_1, b_0, b_1, $Corr$ as the optimal feature subset, achieving a classification accuracy rate of 95.98%. Notably, the m-**of**-n-3-7-10 dataset demonstrated impressive results, showcasing a notable accuracy improvement to 100%. The proposed approach selected a feature subset consisting of 7 features: $2, 4, 5, 6, 7, 8, 9$.

Turning our attention to the Monk1 dataset, our approach has chosen features 1, 2, 5, aligning with the target concept and achieving an accuracy rate of 98.66%. For the Monk2 dataset, the selected features were 1, 2, 3, 6, resulting in a 71.12% accuracy rate. However, in the case of Monk3, the outcomes were commendable, with the proposed approach identifying the features 2, 4, 5 as the optimal subset, matching the target concept with accuracy 87.06%.

3.2 Application on DNA Microarray Dataset

In this second experimentation, we propose to demonstrate the performance of our approach on a real datasets, specifically the DNA microarray datasets.

We select six distinct DNA microarray datasets commonly referenced in the literature which are associated to colon cancer [3], leukemia [6], breast cancer [21], ovarian cancer [16], lung cancer [7] and DLBC cancer (diffuse large B-cell lymphoma) [2]. The information related to the six DNA microarray datasets are described in table 2.

In the initial column, we present the dataset names and the second column indicates the number of genes. The third column is the instance quantity and finally, the last column contains the number of classes. The datasets are taken from Kent Ridge (KR) Bio-Medical Data Set Repository and can be downloaded from this web site[2].

For evaluating the effectiveness of our approach, each DNA microarray dataset is divided into three subsets: 40% of the instances for training, 30% for validation, and the remaining 30% for testing. To mitigate overfitting concerns, we introduced randomness in generating training and validation sets during each iteration of the algorithm.

We consider three key aspects in the experimentation: the classification accuracy rate, the number of selected genes and the computational time.

To determine the classification accuracy rate, we executed the proposed approach 100 times, each time randomly splitting the dataset into training, validation, and testing sets. Subsequently, we calculated the classification accuracy rate for each run, retaining the best, worst, and mean accuracy rates.

Tables 3 and 4 provide details on the classification accuracy rate, the number of selected genes, and the computational time achieved by the proposed approach for each microarray dataset, specifically focusing on the testing set.

Table 3 outlines the classification accuracy rates achieved by the proposed approach across six microarray datasets, specifically focusing on the testing set. Additionally, this table provides details on the best, mean, and worst classification accuracy rates attained by our approach for each dataset. The first column displays the dataset names, while the second column denotes the classification accuracy rate (best, mean, worst).

Table 4 shows the number of selected genes and the corresponding computational time of the proposed approach for each microarray dataset. In the first column, you'll find the dataset names, followed by the initial number of genes before any treatment in the second column. The third column represents the number of genes selected by the proposed approach, and the last column displays the computational time.

As we observe in table 3, a notable increase in the classification accuracy rate is evident. The outcomes presented in table 3 and 4 show the efficacy of the proposed approached. We clearly observe that our approach achieved a remarkable around 99% classification accuracy rate for the five DNA microarray datasets: colon (99.91%), DLBCL (99.19%), leukemia (99.96%), lung (99.87%), and ovarian (99.23%). For breast cancer, the classification accuracy rate reached 91.87%. furthermore, a significant reduction in the number of genes was observed.

The analysis of the results highlights the commendable performance for breast cancer by achieving a maximum accuracy of 91.87% with only 6385 genes out of 24481 which represents 26.08% of selected genes. For colon cancer, it achieved 99.91% of accuracy with 485 genes among the 2000 genes which represents 24.25% of genes. Similarly results for leukemia dataset, the results were noteworthy, with 99.96% of accuracy using 1630

---

[2] datam.i2r.a-star.edu.sg/datasets/krbd/

Table 2. Number of genes, instances and classes of DNA microarray datasets

Dataset name	Number of genes	Number of instances	Number of classes
Breast cancer	24481	97	2
Colon cancer	2000	62	2
DLBC	4026	47	2
Leukemia	5147	72	2
Lung cancer	12533	181	2
Ovarian cancer	15154	253	2

Table 3. Classification accuracy rate (%), obtained by the proposed approach in the first and second step

Dataset name	Classification accuracy		
	Best	Mean	Worst
Breast cancer	91.87	88.10	85.21
Colon cancer	99.91	98.63	96.14
DLBC	99.19	95.42	90.71
Leukemia	99.96	96.81	91.02
Lung cancer	99.87	98.41	96.89
Ovarian cancer	99.23	98.66	95.83

Table 4. Number of selected genes and computational time obtained by the proposed approach in the first and second step

Dataset name	Initial number of genes	Selected genes	Computational time (S)
Breast cancer	24481	6385	1024
Colon cancer	2000	485	264
DLBC	4026	1057	352
Leukemia	5147	1630	500
Lung cancer	12533	3087	865
Ovarian cancer	15154	4600	870

genes out of 5147 (31.66% of genes). In the case of DLBCL cancer, 1057 genes out of 4026 genes where selected (26.25%), achieving, 99.19% of classification accuracy rate. Analyzing the results for lung cancer, the proposed approach achieved maximum 99.87% accuracy with 3087 genes out of 12533 (24.63% of genes). The ovarian cancer dataset demonstrated high accuracy (99.23%) with 4600 genes out of 15154 (30.35%).

Based on the experimental findings, it can be inferred that our approach provides satisfactory results.

The key points encapsulating this paper include:

1. The proposed approach reduces significantly the features and select the optimal subset of features that produces a high classification accuracy.
2. The objective function is the composed of three important terms: classification accuracy, correlation between the features and the class separability measure based on Jaccard Index.
3. The proposed approach is based on a new meta-heuristic never tested on feature selection problem which is Walrus Optimizer.
4. A binary version of the new meta-heuristic was proposed.
5. The objective function is the composed of three important terms: classification accuracy, correlation between the features and the class separability measure based on Jaccard Index.

4 Conclusion

This paper introduces a new feature selection approach based on a new metaheuristic called Walrus Optimizer. A binary version of Walrus

doi: 10.13053/CyS-29-2-5103

Optimizer was proposed and adapted to the problem of feature selection. Additionally, we propose a new fitness function based on three terms: the classification accuracy rate obtained by C-SVM, the correlation between the features and a class separability measure based on Jaccard Index. The proposed approach was tested in two different type of datasets. The first one is synthetic datasets which the relevant subset of features are predefined. We used five synthetic datasets. The second datasets are a real dataset which are the DNA microarray datasets. We evaluated the performance of our approach across six microarray datasets representing various cancers: breast cancer, colon cancer, ovarian cancer, DLBCL cancer, leukemia, and lung cancer.

Recognizing the significance of gene selection in medical research, our approach enhances the quality and complexity of the classifier model, consequently improving the classification accuracy rate. For future works, the proposed approach can be tested and adapted to other type of problems such as the hyperspectral images classification by selecting the relevant bands. Also, an alternative objective function could be designed, focusing on factors.

References

1. **Aghaee, R., Momeni, M., Moallem, P. (2023).** A fusion-based approach to improve hyperspectral images classification using metaheuristic band selection. Applied Soft Computing, Vol. 148, pp. 110753. DOI: https://doi.org/10.1016/j.asoc.2023.110753.
2. **Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J. J., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., Staudt, L. M. (2000).** Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling. Nature, Vol. 403, No. 6769, pp. 503–511.
3. **Alon, U., Barkai, N., Notterman, D., Gish, K., Ybarra, S., Mack, D., Levine, A. (1999).** Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proceedings of the National Academy of Sciences, Vol. 96, No. 12, pp. 6745–6750.
4. **Barraza, J. F., Droguett, E. L., Ramos Martins, M. (2024).** Fs-scf network: Neural network interpretability based on counterfactual generation and feature selection for fault diagnosis. Expert Systems with Applications, Vol. 237, pp. 121670. DOI: https://doi.org/10.1016/j.eswa.2023.121670.
5. **Cao, Z., Xie, X. (2024).** Multi-view unsupervised complementary feature selection with multi-order similarity learning. Knowledge-Based Systems, Vol. 283, pp. 111172. DOI: https://doi.org/10.1016/j.knosys.2023.111172.
6. **Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D. (1999).** Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science, Vol. 286, pp. 531–537.
7. **Gordon, G. J. G., Jensen, R. V. R., Hsiao, L.-L. L., Gullans, S. R. S., Blumenstock, J. E. J., Ramaswamy, S. S., Richards, W. G. W., Sugarbaker, D. J. D., Bueno, R. R. (2002).** Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. Cancer Research, Vol. 62, No. 17, pp. 4963–4967.
8. **Han, M., Du, Z., Yuen, K. F., Zhu, H., Li, Y., Yuan, Q. (2024).** Walrus optimizer: A novel nature-inspired metaheuristic algorithm. Expert Systems with Applications, Vol. 239, pp. 122413. DOI: https://doi.org/10.1016/j.eswa.2023.122413.

9. **Jaccard, P. (1901).** Etude comparative de la distribution florale dans une portion des alpes et des jura. Bulletin de la SociÃ©te Vaudoise des Sciences Naturelles, Vol. 37, pp. 547–579.
10. **Lin, R., Dong, C., Zhou, P., Ma, P., Ma, S., Chen, X., Liu, H. (2024).** Motor imagery eeg task recognition using a nonlinear granger causality feature extraction and an improved salp swarm feature selection. Biomedical Signal Processing and Control, Vol. 88, pp. 105626. DOI: https://doi.org/10.1016/j.bspc.2023.105626.
11. **Liu, J., Li, D., Shan, W., Liu, S. (2024).** A feature selection method based on multiple feature subsets extraction and result fusion for improving classification performance. Applied Soft Computing, Vol. 150, pp. 111018. DOI: https://doi.org/10.1016/j.asoc.2023.111018.
12. **Medjahed, S. A., Fatima, B. (2024).** Applying support vector machines with different kernel to breast cancer diagnosis. Computacion y Sistemas, Vol. 28, No. 2, pp. 659–667.
13. **Medjahed, S. A., Fatima, B. (2024).** Narwhal optimizer: A novel nature-inspired metaheuristic algorithm. International Arab Journal of Information Technology, Vol. 21, No. 3, pp. 418–426.
14. **Medjahed, S. A., Fatima, B. (2024).** On the performance assessment and comparison of features selection approaches. Computacion y Sistemas, Vol. 28, No. 2, pp. 607–622.
15. **Medjahed, S. A., Ouali, M. (2020).** A new hybrid ssa-ta: Salp swarm algorithm with threshold accepting for band selection in hyperspectral images. Applied Soft Computing, Vol. 95, pp. 106534. DOI: https://doi.org/10.1016/j.asoc.2020.106534.
16. **Petricoin, E. F., Ardekani, A. M., Hitt, B. A., Peter J Levine, V. A. F., JD, Steinberg, S. M., Mills, G. B., Simone, C., Fishman, D. A., Kohn, E. C., Liotta, L. A. (2002).** Use of proteomic patterns in serum to identify ovarian cancer. Mechanisms of Disease, Vol. 259, No. 9306, pp. 572–577.
17. **Samareh-Jahani, M., Saberi-Movahed, F., Eftekhari, M., Aghamollaei, G., Tiwari, P. (2024).** Low-redundant unsupervised feature selection based on data structure learning and feature orthogonalization. Expert Systems with Applications, Vol. 240, pp. 122556. DOI: https://doi.org/10.1016/j.eswa.2023.122556.
18. **Sewwandi, M. A. N. D., Li, Y., Zhang, J. (2023).** Granule-specific feature selection for continuous data classification using neighbourhood rough sets. Expert Systems with Applications, pp. 121765. DOI: https://doi.org/10.1016/j.eswa.2023.121765.
19. **Soufan, O. M. (2012).** An Empirical Study of Wrappers for Feature Subset Selection based on a Parallel Genetic Algorithm: The Multi-Wrapper Model. Master's thesis, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia.
20. **Sun, Y., Pei, J. (2023).** Band selection based on hyperspectral piling fisher graphs (hspfigs) analysis. Infrared Physics and Technology, Vol. 133, pp. 104754. DOI: https://doi.org/10.1016/j.infrared.2023.104754.
21. **van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., Friend, S. H. (2002).** Gene expression profiling predicts clinical outcome of breast cancer. Nature, Vol. 415, No. 6871, pp. 530–536.
22. **Wang, J., Xu, P., Ji, X., Li, M., Lu, W. (2023).** Mic-shap: An ensemble feature selection method for materials machine learning. Materials Today Communications, Vol. 37, pp. 106910. DOI: https://doi.org/10.1016/j.mtcomm.2023.106910.
23. **Xie, F., Li, F., Lei, C., Yang, J., Zhang, Y. (2019).** Unsupervised band selection based on artificial bee colony algorithm for hyperspectral image classification. Applied Soft Computing, Vol. 75, pp. 428–440. DOI: https://doi.org/10.1016/j.asoc.2018.11.014.

Article received on 29/07/2024; accepted on 17/01/2025.
**Corresponding author is Seyyid Ahmed Medjahed.*