2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Weight Binary Fish School Search Algorithm for Feature Selection

Fabiana Souza da Fonte Alexandria
School of Engineering
University of Pernambuco - UPE
Recife, Brazil
fsfa@ecomp.poli.br

Fernando Buarque de Lima Neto
School of Engineering
University of Pernambuco - UPE
Recife, Brazil
fbln@ecomp.poli.br

***Abstract*— This study proposes a multimodal approach to the Improved version of Binary Fish School Search (IBFSS) algorithm by incorporating aspects of the Weight Based Fish School Search (wFSS) algorithm to address the attribute selection problem. The proposed model, named Weight Binary Fish School Search (wBFSS), was evaluated on three benchmark datasets, consistently delivering the best solutions in most of the runs. Additionally, two variations of the new wBFSS model were tested to understand the impact on the algorithm's performance by adding a fitness function evaluation before executing the Collective Instinctive Movement.**

Keywords*— *Fish School Search, Swarm Intelligence, Feature Selection, Multimodal Problems.

I. INTRODUCTION

Optimization problems can be classified as unimodal or multimodal. Unimodal problems have only one global optimal solution, while multimodal problems are characterized by the existence of more than one optimal solution [1].

The Feature Selection problem can be considered a multimodal optimization problem since different feature subsets can exhibit similar classification abilities [2]. However, most algorithms treat feature selection as a unimodal problem, aiming to find only one subset of features as the optimal solution.

Currently, population-based metaheuristic algorithms have been widely used as an alternative to traditional methods to solve feature selection problems, mainly due to their superior global search capability and fast convergence speed [2]. Consequently, numerous swarm intelligence algorithms have been proposed to optimize the feature selection problem, including Particle Swarm Optimization (PSO) [3], Ant Colony Optimization (ACO) [4], Fish School Search (FSS) [5] [6], among others.

The Fish School Search (FSS) [5], originally proposed by Bastos Filho and Lima Neto in 2008, is a unimodal optimization algorithm inspired by the collective behavior of schools of fish. The FSS incorporates mechanisms of feeding and coordinated movements, forming the basis of its search mechanism. The main idea is to guide the fish to swim in the direction of the positive gradient to find food and gain weight. Collectively, heavier fish have more influence on the search process, causing the barycenter of the fish school to move to better locations in the search space over the iterations.

In 2014, Lima Neto and Lacerda [1] proposed a multimodal version of the FSS, known as the Weight Based Fish School Search (wFSS), which introduces a mechanism for forming virtual links between fish based solely on their weights. This mechanism leads lighter fish to follow heavier ones, and the collective movements ensure that the influence of the heavier fish on the lighter ones becomes stronger as the weight difference between them increases. More details about wFSS will be presented in Section II.

Also, in 2014, Sargo et al. [7] implemented an adaptation of the FSS algorithm, resulting in the binary version of the model named BFSS (Binary Fish School Search). The BFSS was applied to address the feature selection problem in a unimodal manner. In this work, the representation of the fish position vector was modified, enabling the algorithm to handle binary variables. In 2016, Carneiro and Bastos-Filho [8] introduced several improvements to the BFSS, leading to the Improved version of Binary Fish School Search (IBFSS) algorithm. In the IBFSS, the initialization of the School now selects approximately 25% of the total number of resources instead of the 50% in the original BFSS. Additionally, a change was implemented regarding the use of the Individual Step in the Individual Movement; instead of using it as a threshold as in the original BFSS, the Individual Step was employed to invert selected features. Furthermore, improvements to the Collective Instinctive and Collective Volitional movements were added to maximize algorithm performance. Section III provides a more detailed presentation of the IBFSS.

This study proposes a multimodal approach to enhance the Improved version of Binary Fish School Search (IBFSS) algorithm by incorporating aspects of the Weight-Based Fish School Search (wFSS) algorithm to address the feature selection problem. The proposed model, called Weight Binary Fish School Search (wBFSS), was evaluated on three benchmark datasets, consistently delivering the best solutions in most of the runs.

The remainder of the paper is organized as follows: Section II and Section III provide detailed descriptions of the wFSS and IBFSS Algorithms, respectively. Section IV presents the proposed wBFSS algorithm. Section V details the experiments whose results are presented in Section VI. Finally, the conclusions are presented in Section VII.

II. WEIGHT-BASED FISH SCHOOL SEARCH (WFSS)

The wFSS algorithm is the niche version of the FSS algorithm developed by Lima Neto and Lacerda [1], designed to address multimodal problems with low computational cost. In this version, the operator for forming virtual links between fish based solely on their weights is added to the original algorithm. This approach does not use topological information, providing a significant advantage compared to other swarm intelligence techniques [1]. Along with creating the link formation operator, modifications were made to some of the existing operators in the original FSS. Lima Neto and Lacerda [1] described the wFSS operators as follows.

A. *Individual Movement*

The Individual Movement remains unchanged from the original FSS [5]. In this movement, fish move randomly and independently towards the positive gradient. Consequently, each fish randomly selects a new position in the search space and evaluates it using the objective function. The fish only moves to the new position if there is an improvement in its fitness. This movement is described by (1), and the new step size is calculated using (2).

$$x_{ij}(t+1) = x_{ij}(t) + r\, step_{ind}(t), \tag{1}$$

$$step_{ind}(t+1) = step_{ind}(t) - \frac{step_{ind_{init}} - step_{ind_{final}}}{iterations}. \tag{2}$$

Where $x_{ij}(t+1)$ is the new value of the dimension j in the position vector of the individual i, $x_{ij}(t)$ is the old value, r is a random value between 0 and 1 and $step_{ind}(t)$ is the step size on time t [1].

B. *Feeding Operator*

The Feeding Operator also remains unaltered from the original FSS [5]. This operator is responsible for updating fish weights when a fish improves its fitness during the Individual Movement. The weight update is calculated as shown in (3).

$$W_i(t+1) = W_i(t) + \frac{\Delta f_i}{\max(\Delta f)}. \tag{3}$$

Where Δf_i, is the fitness variation after the Individual Movement of the fish i, and $max(\Delta f)$ is the maximum fitness variation in the whole population [1].

C. *Link Formation*

This operator was the major innovation introduced by Lima Neto and Lacerda [1]. The Link Formation is responsible for establishing links (leader-follower relationships) between fish. In this context, if a fish links to another fish, it will be influenced by the behavior of its leader during collective movements. The weight difference between them determines the degree of leadership of the heavier fish. Heavier fish are more likely to become leaders of lighter fish.

The rule for link formation between two fishes is defined as follows. For each fish “a”, it randomly chooses another fish “b” from the school. If “b” is heavier than “a”, then “a” establishes a link with “b” and starts following “b” (i.e., “b” becomes the leader of “a”). Otherwise, nothing happens [1].

However, if “a” already has a leader “c”, and the sum of the weights of “a's” followers is greater than the weight of “b”, then “a” stops following “c” and starts following “b”. Otherwise, “a” will continue following “c”. In the subsequent evaluations (in each interaction), if “a” becomes heavier than its leader, the link will be broken. Each fish can have only one leader at most [1].

D. *Collective Instinctive Movement*

In the original FSS [5], the Collective Instinctive Movement is calculated based on the resultant vector with the aim of guiding all fish in the direction indicated by successful individual movements. In contrast, in wFSS, this operator is performed considering only the leader fish (if any) and the fish itself. Additionally, Lima Neto and Lacerda [1] employed a multiplying factor alpha to adjust the operator's influence in the algorithm over iterations, leading to a linear increment over time. The Collective Instinctive Movement is described by (4), and the multiplying factor is calculated by (5).

$$x_{ij}(t+1) = x_{ij}(t) + \alpha\left(\frac{\Delta x_{ij}\,\Delta f(\overline{x_i}) + L\,\Delta x_{lj}\,\Delta f(\overline{x_l})}{\Delta f(\overline{x_i}) + L\,\Delta f(\overline{x_l})}\right) \tag{4}$$

$$\alpha = \frac{current\_iteration}{maximum\_iterations}, \tag{5}$$

Where L is a variable that assumes the value 0 if the fish i has no leader l and 1 otherwise [1].

E. *Collective Volitive Movement*

Lima Neto and Lacerda [1] retained the basic logic of the Collective Volitional Movement from the original FSS [5]. This operator is responsible for the expansion and contraction of the school of fish, according to the total weight of the school increases or decreases, respectively.

Thus, like the Instinctive Movement, the Volitive Movement requires each fish to calculate its own reference point to serve as the barycenter. These individual reference points are calculated solely based on the leader of each fish (if it exists) and the fish itself [1]. The barycenter calculation is described by (6).

$$B_i(t) = \frac{x_{ij} W_i(t) + L x_{lj} W_l(t)}{W_i(t) + L W_l(t)}. \tag{6}$$

Where the variable L takes the value 0 if the fish i has no leader l and 1 otherwise [1].

III. IMPROVED VERSION OF BINARY FISH SCHOOL SEARCH (IBFSS)

The IBFSS algorithm is applied to solve feature selection problems in a unimodal manner. The IBFSS algorithm presented in this section is the binary version of the FSS, developed by Carneiro and Bastos-Filho [8], which includes improvements to the original BFSS [7].

A. *Initialization of the School*

For each fish i, the initial position was randomly initialized according to (7). Here, N represents the total number of fish, u is a random number in the interval [0,1], and F is the total number of dimensions of the problem [8].

$$x_{ij} = \begin{cases} 1 & if\ u < 0.25 \\ 0 & otherwise \end{cases},\ i = 1, \dots, N, \quad j = 1, \dots, F, \tag{7}$$

B. *Individual Movement*

Like the FSS, the Individual Movement is only executed if there is an improvement in the fish's fitness. In this movement, the individual step, $S_{ind}(t)$, is used to flip the selected feature. For instance, if $S_{ind}(t)$ equals 4, the fish will flip 4 feature variables when executing the Individual Movement [8].

The individual step $S_{ind}(t)$ is a percentage of the search space and is updated (decreased) throughout the iterations, gradually reaching a value of zero. To prevent the fish from not moving individually, if the $S_{ind}(t)$ value reaches zero, it is set to 1 if a random variable v is greater than a certain threshold. This ensures that the fish will flip at least one random feature [8].

C. Collective Instinctive Movement

In this movement, each fish is compared to the resultant vector considering all the fish in the school [7]. For the improved version of the BFSS, Carneiro and Bastos-Filho [8] proposed a new way of determining the number of resources that will be inverted in each iteration. The Collective Instinctive Movement is executed as shown in Algorithm 1.

Algorithm 1 pseudocode of Improved Collective-Instinctive Movement.

$j \leftarrow 1;$
$numberOfFlips \leftarrow \%\ of\ Instinctive\ Step;$
for all $fish\ i\ in\ Fish\ School$ **do**
  for $j \leq numberOfFlips$ **do**
    $r \leftarrow Random(1, ..., M);$ ▷ //r never repeats
    $x_{ir} \leftarrow \overline{x_{ir}};$
    $j \leftarrow j + 1;$

Where numberOfFlips represents a percentage of the number of features that are marked to flip, *r* is a random number in the interval [0, 1], x_{ir} is the r-th dimension of the position of fish i, and M is the number of features marked to be flipped. The random number r is always unique and does not repeat [8].

D. Collective Volitive Movement

Carneiro and Bastos-Filho [8] also added some changes in this movement that is responsible for approximating or expanding the fish of the school. For the improved version of the BFSS, the Volitive Step was considered to determinate how far the fish should move closer or farther away (step size) to the barycenter.

In this movement, the position of each fish is compared to a resultant vector or its opposite, depending on an increase or decrease in the total weight of the school of fish, respectively. All the dimensions that are different to the resultant vector are marked to be flipped.

The Collective Volitional Movement is executed as shown in Algorithm 2.

Algorithm 2 pseudocode of Improved Collective-Volitive Movement.

$j \leftarrow 1;$
$numberOfFlips \leftarrow \%\ of\ Vollitive\ Step;$
for all $fish\ i\ in\ Fish\ School$ **do**
  for $j \leq numberOfFlips$ **do**
    $r \leftarrow Random(1, ..., K);$ ▷ //r never repeats
    $x_{ir} \leftarrow \overline{x_{ir}};$
    $j \leftarrow j + 1;$

Where numberOfFlips represents a volitive step percentage of fish *i*, *r* is a random number in the interval [1,K], K is the number of features marked to be flipped, and x_{ir} is the r-th dimension of the fish's position [8].

E. Feeding Operator and Upgrade Steps

The feeding operator remains unchanged from the Original FSS [5][6] and is used to increase or decrease the weight of the fish and update the individual and volitive steps of the fish [8].

IV. Weight Binary Fish School Search (wBFSS)

Considering that subsets of different features may exhibit similar classification ability, this study proposes the wBFSS algorithm, a multimodal version of the IBFSS algorithm, to address the feature selection problem. To achieve this, aspects of the wFSS algorithm were incorporated into the IBFSS, as described below.

Additionally, two variations of the wBFSS, Model 1 and Model 2, were implemented to assess the impact on the algorithm's performance when adding a fitness function evaluation before executing the Collective Instinctive Movement.

It is important to note that the representation of the search space, the initialization of the school and the Individual Movement remain unchanged compared to the IBFSS. While the Formation of Links has inherited the behavior from the wFSS Model; Therefore, there is no need to present these mechanisms again.

A. Feeding Operator

The feeding operator has been adjusted to prevent fish with a high initial fitness value from failing to attract followers. This issue arose because, at each iteration, fish with a high fitness value were unable to perform the Individual Movement, resulting in a fitness variation that was consistently zero. This lack of fitness variation meant that the fish did not feed and remained "skinny," which hindered their ability to attract followers. On the contrary, these fish ended up following other fish with a lower fitness value but a higher variation value.

To address this, the fitness variation used in the feeding operator was replaced by the fitness value, as described in (8).

$$W_i(t+1) = W_i(t) + \frac{f_i}{\max(f)}. \tag{8}$$

Where *fi* is the fitness after the Individual Movement of fish *i*, and *max(f)* is the maximum fitness in the entire population.

B. Collective Instinctive Movement

This movement is only executed if the fish has a leader. Here, each fish is no longer compared to the resultant vector, as in previous binary versions, but rather to its leader.

For wBFSS Model 1, an additional check was used, which did not exist in previous versions of this movement. This additional verification ensures that, like the Individual Movement, the Collective Instinctive Movement is only executed if there is an improvement in the fitness of the new position. Model 2 does not have this additional verification; the Collective Instinctive Movement is performed normally, without the fitness function verification. The mechanism of this movement is described in Fig. 3.

For each fish ***do***
  If *the fish has a leader* ***then***

Get number of positions with different values between the fish and its leader;

Get the number of positions that will have their values flipped according to the individual step;

Among the positions with different values, randomly choose the positions that will be flipped, according to a percentage calculated in the previous step;

Change the values;

Check the fitness of the new position (only for Model 1);

If *the new fitness is better* ***then*** *(only for Model 1)*
Move the fish to the new position

Fig. 1. Steps of the Collective Instinctive Movement.

The number of positions that will have their values flipped (*num_flips*) is described by (9), where *num_dif_values* is the number of positions with different values between the fish and its leader.

num_flips = individual_step * num_dif_values. (9)

C. Collective Volitive Movement

This movement promotes the contraction or expansion of the school of fish, causing the fish to become either closer or more distant from their leader, in the event of an increase or decrease, respectively, in the total weight of the school after the execution of the Instinctive Movement. For that, each fish is compared to its leader or its leader's opposite and all the dimensions that are different are marked to be flipped.

This movement only occurs in Model 2, as there is no decrease in the total weight of the school in Model 1. The mechanism of this movement is described in Fig. 2.

For each fish ***do***
  If *the fish has a leader* ***then***

Get number of positions with the same value between the fish and its leader;

Get the number of positions that will have their values flipped according to the volitive step;

Among the positions with the same value, randomly choose the positions that will be flipped, according to a percentage calculated in the previous step;

Change the values;

Move the fish to the new position

Fig. 2. Steps of the Collective Volitive Movement

Like the previous movement, the number of positions that will have their values flipped (*num_flips*) is described by (10), where *num_dif_values* is the number of positions with different values between the fish and its leader.

num_flips = volitive_step * num_dif_values. (10)

D. Individual Step Update

The individual step is updated at each iteration, just as described in the original FSS. The equation for updating the individual step is described in (2).

V. Experiments Description

The objective of the experiments is to find the best solution sets for each analyzed dataset and understand the impact on the algorithm's performance when using an an additional check of the fitness function in the Instinctive Movement. For this purpose, two models were implemented, and their performance was compared using the datasets and parameters described in the following sections. Each algorithm was executed for 10 independent runs.

Similarly to what is applied in [1], four metrics were used to evaluate each algorithm: (i) the percentage of simulations in which the algorithm found 4 or more of the best existing solutions; (ii) the percentage of simulations in which the algorithm found 3 of the best existing solutions; (iii) the percentage of simulations in which the algorithm found 2 of the best existing solutions; (iv) the percentage of simulations in which the algorithm found only 1 or none of the best existing solutions.

To evaluate the fish's fitness, the non-parametric K Nearest Neighbors (KNN) classifier was chosen, with k = 5. The objective function used in this work is described in (11).

$$Fitness(x) = \beta * precision(x) + \alpha * \left(1 - \frac{\sum_{i=1}^{N} x_i}{N}\right). \quad (11)$$

Where x represents the fish being evaluated, β represents the weight assigned to model precision, and α represents the weight assigned to the complement of selected attributes. N represents the total number of dataset attributes (independent variables), and x_i represents the selected attribute [9].

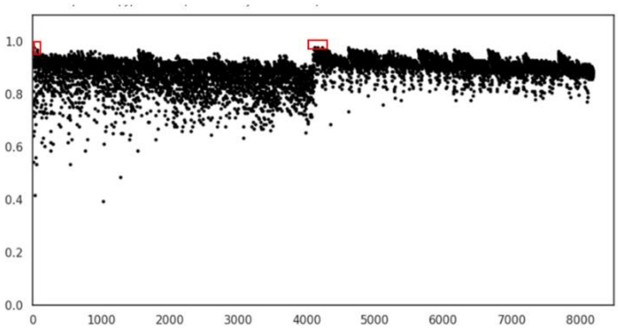
A. Datasets

Table I displays the benchmark datasets used to obtain the initial evaluations of the proposed algorithm's performance. Only low-dimensional datasets were included, enabling their optimal solutions to be found through enumeration. These datasets are from the UCI database [10].

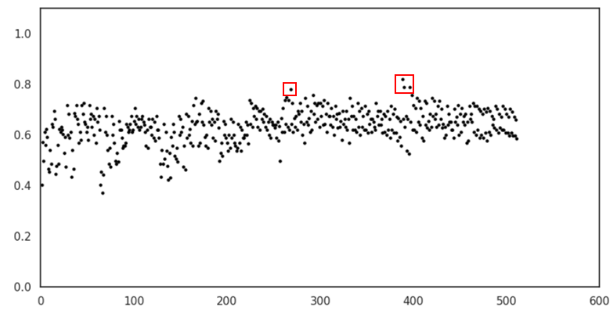
TABLE I. DATASETS USED IN THE EXPERIMENTS

Datasets	Nº of Features	Nº of Categories	Nº of Samples	Normalized Dimensions Ranges	Number of Possible Solutions
WINE	13	3	178	[0, 1]	2^{13} = 8191
GLASS	9	6	214	[0, 1]	2^{9} = 511
VOWEL	10	11	528	[-1, 1]	2^{10} = 1023

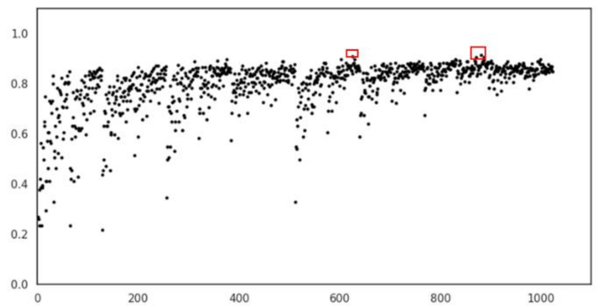
The dataset enumeration results are illustrated in Fig. 3, which uses the binary string decimal representation of the solution on the x-axis, while the y-axis shows the corresponding fitness. Solutions with the highest fitness value are indicated by red rectangles.



a) Wine dataset (6 best solutions, all with values of 0.98 in the fitness function)



b) Glass dataset (4 best solutions, all with values above 0.78 in the fitness function)



c) Vowel dataset (4 best solutions, all with values above 0.90 in the fitness function)

Fig. 3. dataset enumeration results: a) Wine, b) Glass and c) Vowel

B. Parameters

The parameter values used in the algorithms are described in Table II.

TABLE II. PARAMETERS USED IN THE EXPERIMENTS

Parameters	Value
Nº of fishes	15
Nº of iterations	30
$Step_{initial}$	0,5
$Step_{final}$	0,0000001
$Step_{institive}$	0,3
$Step_{volitive}$	0,3
variable v *	0,5
α (function fitness)	0,9
β (function fitness)	0,1

* Used in the individual movement of the IBFSS

VI. RESULTS

Table III presents the comparative results from 10 independently conducted runs. Two variations of the multimodal model were employed to address the feature selection problem, with the objective of analyzing the impact of adding a fitness assessment to the Collective Instinctive Movement. Model 1 incorporates this additional evaluation, whereas Model 2 is run without it.

TABLE III. RESULTS OF THE EXPERIMENTS.

Dataset	Metric (i) 0 or 1 optimal solution found		Metric (ii) 2 optimal solutions found		Metric (iii) 3 optimal solutions found		Metric (iv) 4 or more optimal solutions found	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Wine	10%	60%	30%	30%	**30%**	10%	**30%**	0%
Glass	30%	0%	40%	**60%**	30%	**40%**	0%	0%
Vowel	0%	60%	30%	20%	**30%**	0%	**40%**	20%

Analyzing the presented results, it is possible to observe that Model 1 outperformed Model 2 in the Wine and Vowel datasets for metrics (iii) and (iv). This indicates that the algorithm with a fitness check before executing the Collective Instinctive Movement can more consistently provide a greater number of better solutions. However, Model 2 surpassed Model 1 only on the Glass dataset.

VII. CONCLUSIONS AND FUTURE WORK

In this work, a new version of the Fish School Search algorithm, the Weight Binary Fish School Search (wBFSS), was introduced, which incorporated features from the Weight-Based Fish School Search (wFSS) and the Improved Version of Binary Fish School Search (IBFSS) to address the feature selection problem using a multimodal approach. Two variations, Model 1 and Model 2, of the wBFSS were analyzed to understand the impact on the algorithm's performance when evaluating the fitness of the fish in the Collective Instinctive Movement. In Model 1, this movement is executed only if there is an improvement in the fitness function, while Model 2 does not include this evaluation.

The results showed that Model 1 obtained better results in two (wine and vowel) of the three datasets used in the experiments.

As future work, wBFSS will be tested on datasets with a larger number of attributes and instances to evaluate its scalability.

REFERENCES

[1] F. Buarque De Lima Neto and M. Gomes Pereira de Lacerda, "Weight based fish school search," 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), San Diego, CA, USA, 2014, pp. 270-277, doi: 10.1109/SMC.2014.6973919.

[2] Multimodal Particle Swarm Optimization for Feature Selection

[3] J. Kennedy and R. Eberhart, "Particle swarm optimization," Neural Networks, 1995. Proceedings., IEEE International Conference on, vol. 4, pp. 1942 – 1948, November 1995.

[4] M. Dorigo and L. M. Gambardella, "Ant colony system: A cooperative learning approach to the traveling salesman problem," IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, vol. 1, no. 1, pp. 53 – 66, April 1997.

[5] C. J. A. Bastos Filho, F. B. de Lima Neto, A. J. C. C. Lins, A. I. S. Nascimento, and M. P. Lima, "A novel search algorithm based on fish school behavior," IEEE International Conference on Systems, Man and Cybernetics, 2008.

[6] C. J. A. Bastos Filho, F. B. de Lima Neto, M. F. C. Sousa, M. R. Pontes, and S. S. Madeiro, "On the influence of the swimming operators in the fish school search algorithm," Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics, October 2009.

[7] J. A. G. Sargo, S. M. Vieira, J. M. C. Sousa, and C. J. A. Bastos Filho, "Binary fish school search applied to feature selection: Application to icu readmissions," IEEE International Conference on Fuzzy Systems (FUZZIEEE), 2014.

[8] Carneiro, Raphael F., and Carmelo JA Bastos-Filho. "Improving the binary fish school search algorithm for feature selection." 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI). IEEE, 2016.

[9] Rojas, Matías Gabriel, et al. "A memetic cellular genetic algorithm for cancer data microarray feature selection." IEEE Latin America Transactions 18.11 (2020): 1874-1883.

[10] P. M. Murphy and D. W. Aha, "UCI repository of machine learning databases," Dept. Inf. Comput. Sci., University of California, Irvine, CA, USA Tech. Rep. [Online]. Available: http://www.ics.uci.edu/mlearn/ MLRepository.html