

# Dual Heuristic Feature Selection Based on Genetic Algorithm and Binary Particle Swarm Optimization

**Ali Hakem Jabor**

*University of Al-Qadisiyah  
College of Computer Sciences and IT  
[alihakemalsaeedi@gmail.com](mailto:alihakemalsaeedi@gmail.com)*

**Ali Hussein Ali**

*Al-Qadisiyah Education Directorate  
[alihusseinalialjanabi@gmail.com](mailto:alihusseinalialjanabi@gmail.com)*

---

## ARTICLE INFO

Submission date: 29/7/2018

Acceptance date: 4/10/2018

Publication date: 10/3/2019

---

## Abstract

The features selection is one of the data mining tools that used to select the most important features of a given dataset. It contributes to save time and memory during the handling a given dataset. According to these principles, we have proposed features selection method based on mixing two metaheuristic algorithms Binary Particle Swarm Optimization and Genetic Algorithm work individually. The K-Nearest Neighbour (K-NN) is used as an objective function to evaluate the proposed features selection algorithm. The Dual Heuristic Feature Selection based on Genetic Algorithm and Binary Particle Swarm Optimization (DHFS) test, and compared with 26 well-known datasets of UCI machine learning. The numeric experiments result imply that the DHFS better performance compared with full features and that selected by the mentioned algorithms (Genetic Algorithm and Binary Particle Swarm Optimization).

**Keyword:** Data Mining, Features Selection, Genetic Algorithm, Binary Particle Swarm Optimization, Metaheuristic Optimization.

## I. Introduction

Over the last decades the devices, sensors, and users are on increasing, therefore, the dimensions of the datasets are increasing. Logically, the size of data is directly proportional to the execution time. As a result, reducing the dimensions of the data becomes necessary to decrease execution time or processing. Using Data mining technique in many fields such as Artificial intelligence [1], Databases [2], Image and video processing [3], and others, make it in interesting topic for researchers. It is one of the important techniques used for filtering data. The data mining searches of the data (features or instances) that related to the objective of the dataset and removing garbage data from the dataset [4]. The techniques that omitted unimportant features from the dataset called features selection (FS). There are many algorithms used for FS such as Filter [5], Wrapper [6], and Embedded [7] Methods. The metaheuristic algorithms use in optimizing FS of verity styles. The advantages of the stochastic search are fast, flexible, and succeed to solve many hard optimization problems, but disadvantages no grantee to

find a global solution and may be suffering from stagnation at local optimum [8] [9]. The Stagnation phenomena is a problem happens when the algorithm gave the same solution (local optimum) during several search steps. To reduce stagnation need to decrease convergence between candidate solutions. Therefore, went to increase the diversity of new solutions. The increases randomness or number of mutation genes leads to make the population more diversified.

[10] Have proposed a wrapper approach with Harmony Search algorithm (HS) that adaptive for features selection. [11] The Binary Approach of Artificial Bee Colony Optimization (BABC) for features selection. The binary vector generated by BABC represents the FS, which were the features that the corresponding one have been selected. [12] Modify Gravitation Search Algorithm (GSA) in order to find a subset of features that maximize the Optimum-Path Forest (*OPF*) accuracy over a validation set.

The previous works [10] [11] [12] modify the stochastic search algorithms for FS without trying to reduce the stagnation phenomena in these algorithms. The Binary Particle Swarm Optimization (BPSO) may be suffering from stagnation problem [8] [9] in some points of search (as in Evolutionary Algorithms EA). The hybrid algorithm is more efficient than the algorithms it has built, because it combines the good features of them. Therefore, in order to reduce stagnation in BPSO, we combine it with GA sequentially, and uses both of them for features selection. The sequential of calling for BPSO and GA individually makes the proposed algorithm more robustness with keeping on original formatted of the mentioned algorithm by calling them sequentially. Increasing the number of mutation genes and decreasing the crossover operations during search progress help the proposed algorithm to be more robust to deal with stagnation problem. The wrapper method uses a heuristic to rank the features [10] [11] [12], which are used for FS in the proposed algorithm. The experiments have been performed in 26 well-known datasets of *UCI* machine learning to compare the proposed algorithm (DHFS), full set classification, and FS by GA and BPSO. The numeric experiments results imply that proposed DHFS is better performance compared to mentioned algorithms.

The remainder of the paper is organized as follows: In Section II we explain the Particle Swarm Optimization (PSO). Section III presents the Binary Particle Swarm Optimization BPSO (BPSO), Genetic Algorithm (GA) is described in Section IV. Section V illustrates the Features selection (FS). Dual Heuristic Feature Selection (DHFS) present in Section VI. Section IX discusses Validate and test algorithm. Finally, the conclusion and future works stated in Section X.

## II. Particle Swarm Optimization (PSO)

The PSO is popular metaheuristic algorithm inspired by the behaviour of social animals. The robustness, stability, and simplicity enough to be it quite use for enhancing the different fields [13] such as Data mining [14] medical apply [15] image processing [3] speech recognition [16]. There are many similar features between PSO and other Evolutionary Algorithms (EA) [17]. All EA start with a random population and calculate the fitness of each participant (candidate solution) to evaluate the performance of the population (all candidate solution). It uses the random mathematical model to update the

population to search for a local optimum solution. The main difference between the PSO techniques with other EA [14]: (i) The PSO uses the velocity of the particle to update population (particles) rather than uses genetic operation such as parent selection, crossover, and mutation. (ii) The easiness of implementation and does not use complex operations make PSO more popular compared with other EA. (iii), The computational time of PSO is small compared to other EA because it has had a limit number of parameters and calculation formal. (iv) The PSO is simple in both theory and mathematically implementation.

The principal search of PSO based on a group of particles (candidate solutions), each particle has own velocity (random initialization) and current position (solution), meanwhile, the particles stored the best solution figure out by itself, this solution called local best optimum ( $p_{best}$ ). The PSO select the best solution among whole candidate solutions and update it dynamically during the search process this solution called global best ( $g_{best}$ ). The velocity of particles updates in each step of a search process according to formula 2.1 [18].

$$V_i^d(t+1) = w(t)V_i^d(t) + c_1r_1(pbest_i^d - x_i^d(t)) + c_2r_2(gbest^d - x_i^d(t))$$

(2.1) where:  $r_1$  and  $r_2$  are random variables in the range [0, 1].  $c_1$  and  $c_2$  are positive constants.  $w$  is the inertia weight.  $v_i^d(t), x_i^d(t)$  are indicate the velocity and position of  $i^{th}$  particle at iteration  $t$  in  $d^{th}$  dimension, respectively. The PSO uses 2.2 to update the value of the particles (candidate solution) [18].

$$x_i^{t+1} = x_i^t + v_i^{t+1}$$

(2.2)

Where:  $x_i^t$  is old particle value,  $x_i^{t+1}$  is new particle value.

### III. Binary Particle Swarm Optimization BPSO (BPSO)

Most versions of metaheuristic algorithms are suitable only for continuous search space, so that, it needed to modify to be suitable for binary search space. The transfer function [19] one of the ways that adaptive continuous metaheuristic algorithm for binary search space. In the proposed method, Formal 3.1 [19] used to modify PSO to be suitable for binary search.

$$V(v_i^d(t)) = \left| \frac{2}{\pi} \arctan\left(\frac{2}{\pi} v_i^d(t)\right) \right|$$

(3.1)

After calculating the velocity ( $V(v_i^d(t))$ ), the equation 3.2 uses to define the possibility of flipping the value of particles:

$$X_i^k(t+1) = \begin{cases} X_i^k(t)^{-1} & \text{if } round < V(v_i^k(t)) \\ X_i^k(t) & \text{if } round \geq V(v_i^k(t)) \end{cases} \quad (3.2)$$

Where:  $X_i^k(t+1)$  is new particle value,  $X_i^k(t)$  is old particle value,  $X_i^k(t)^{-1}$  the flipping value of particle if the particle is  $0 \rightarrow 1$ , or  $1 \rightarrow 0$ .

#### IV. Genetic Algorithm (GA)

The genetic algorithm is a discrete population metaheuristic algorithm inspired by the genetic behaviour of natural life according to Charles Darwin's theory of natural evolution [18]. It follows the rule of natural selection where the good individuals are contributing produce offspring. The mechanism search of GA essentially based on three genetic operations: a parent selection, crossover, and mutation. The parent selection is a process of selecting two or more parents from the crossover pool to produce new offspring (new candidate solution). The good parents have more chance to be selected for reproduction according to Darwin's theory [20]. There are other methods to select parents of GA such as a roulette wheel, tournament [21]. The crossover is mix genes of parents that elected for crossover operation. There are many crossover techniques such as one point crossover, two-point crossover [9], Arithmetic Crossover[22], heuristic crossover [23]... etc. The mutation is tweak change in genes of offspring. The parent selection and crossover are not enough to solve the stagnation in local optimum [3]. Therefore, the mutation process in GA is important to make diversity and reduce stagnation effect on the search processing [17].

#### V. Features selection (FS)

During the last decades, many datasets have huge information and high dimensional with hundreds or ten thousand features. Some of these features may be not important to the main object of the dataset [4]. The selection of important features that relate to a dataset goal called features selection. It employs the specific technique to remove garbage features from the dataset. The features selection technique is important tools to save train time and enhance the accuracy ratio of machine learning algorithms [24]. As a result makes the features selection a hot topic area for researchers. There are many features of selection methods: Filter, Wrapper, and Embedded [25]. Filter Methods depends on the relationship between the features and the target of the dataset to select the importance of features [5]. Embedded method for feature selection, which achieves by the insights using in some Machine Learning models such as LASSO Linear Regression and Tree-based models [7]. Wrapper Methods generate models with a subset of feature and gauge their model performances [6]. The stochastic search for important features can be select subset features of the wrapper model [10]. The time complexity of running an algorithm depends on data domination.

#### VI. Dual Heuristic Feature Selection (DHFS)

In this section, we focus on features selected based on a combination of two population metaheuristic techniques: Genetic algorithm (GA) and Binary Particle Swarm Optimization (BPSO). The hybrid algorithm has the advantages of multiple algorithms when deploy to solve optimization problems [26]. The search process of GA depends on three main operations: parent selection, crossover, and mutation. The parent selection and crossover are not enough to solve stagnation in local optimum [3]. Therefore, the mutation operation of GA tries to reduce stagnation at local optimum by increasing the chance to product diversity solutions [17]. The BPSO also suffering from stagnation problem [8] [9] [19]. The stagnation in BPSO happens when the local best solution  $P_{best}$

and global best solution  $G_{best}$  has no change during several steps of the search process. The proposed algorithm reduces the stagnation in BPSO by calling GA to decrease convergence in newly candidate solutions (population). The proposed method (DHFS) as shown in bellowing:

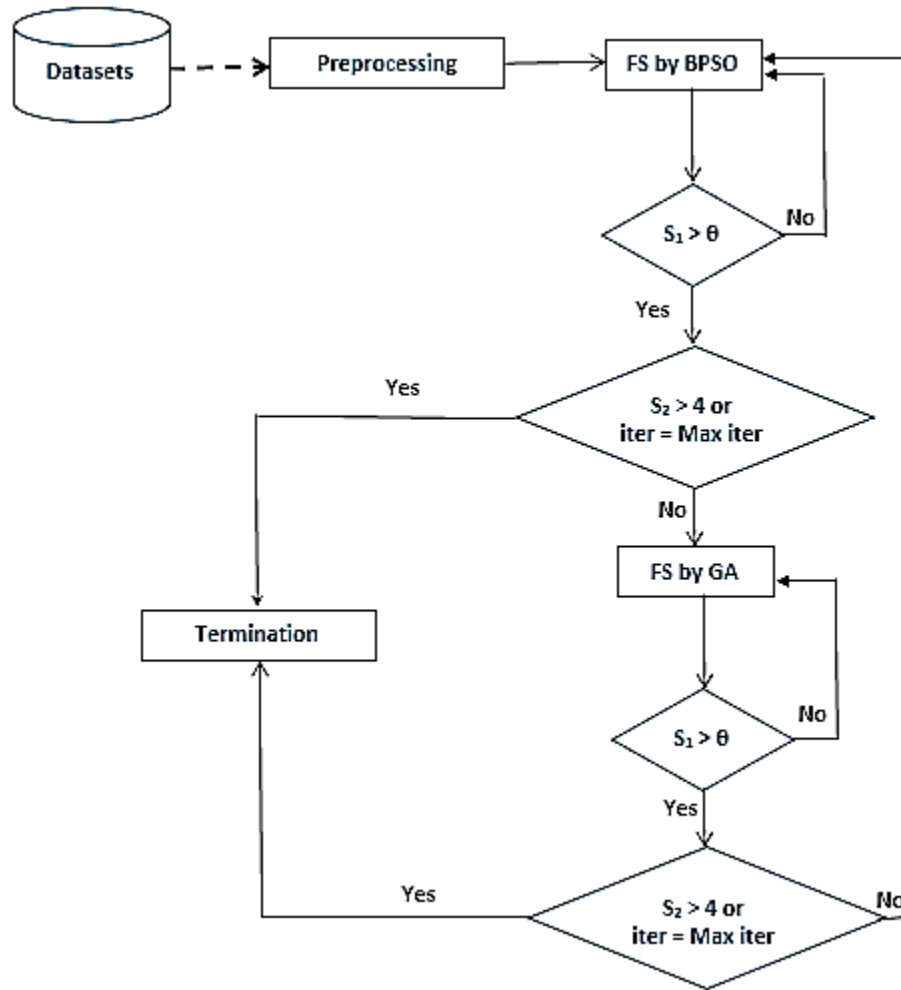
```

1:  $p \leftarrow$  initial random binary population
2:  $best_{solution} = -\infty$ 
3:  $S_1 \leftarrow 0, S_2 \leftarrow 0$ 
4: While  $S_2 < 4$ 
5:   reset velocity and  $S_1$ 
6:   for each iteration
7:     for each  $p$ 
8:       select  $P_{best}$  and  $G_{best}$ 
9:       update velocity (eq 2.1) and position (eq 3.2)
10:      if  $best_{solution} < G_{best}$ 
11:         $best_{solution} = G_{best}$ 
12:        reset  $S_1$ 
13:        reset  $S_2$ 
14:      Else
15:         $S_1 \leftarrow +1$ 
16:         $S_2 \leftarrow +1$ 
17:      if  $S_1 > \theta$  break
18:    if  $S_2 > 4$  break
19:    reset  $S_1$ 
19:    for each iteration (contain with BPSO)
20:      for each  $p$ 
21:        parent  $\leftarrow$  roulette ( $p$ )
22:        upadate ( $p$ )  $\leftarrow$  crossover (eq. 6.1) and mutation (eq. 6.2)
23:        upadte best
24:      if  $best_{solution} < best$ 
25:         $best_{solution} = best$ 
26:        reset  $S_1$ 
27:        reset  $S_2$ 
28:      Else
29:         $S_1 \leftarrow +1$ 
30:         $S_2 \leftarrow +1$ 
31:      if  $S_1 > \theta$  break
32:    if  $S_2 > 4$  break
33: Return  $best_{solution}$ 

```

$S_1$  is a number of steps where no enhance in results when applied an algorithm (BPSO or GA). It reset after calling another algorithm.  $S_2$  is a number of calls an algorithm (BPSO or GA) but without enhancing in results. The  $S_2$  reset when any algorithms (BPSO, BGA) find a new best solution. The ( $\theta$ ) represents the number of valid steps without a change in results for both algorithms (BPSO or GA). The proposed method consists of four parts: binary metaheuristics optimization algorithms (BPSO and

GA), feature selection, switching between algorithms, and termination. Figure 1 illustrate the relation between the main parts of the DHFS.



**Figure1: Main steps of DHFS**

The algorithm starts with BPSO to grantee find first local optimum early [26]. The PSO when going through the stagnation as progress search in several steps ( $\theta$ ) the proposed algorithm called GA to makes diversity in population (candidate solutions), then recall the BPSO when GA failed to make diversity in candidate solutions. The Arithmetic Crossover (AC) used in GA to produce new offspring has different genes of its parents rather than using segment exchange [22]. AC products offspring by using 6.1

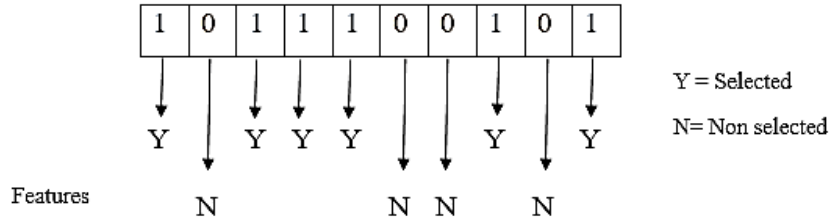
$$Offspring = \alpha.parent1 + (1 - \alpha).parent2 \quad (6.1)$$

Where:  $\alpha$  either [0, 1].

Generally, the probability of falling into stagnation increases based on the search progress, it must increase the number of genes for a mutation in GA as search progress. Formal 6.2 calculates the number of genes ( $m$ ) for the mutation to each participant.

$$m = round(m_{min} + \left(\frac{current\ iteration}{maximum\ iteration}\right)(m_{max} - m_{min})) \quad (6.2)$$

Where:  $m_{min}$  is the minimum number of genes for mutation,  $m_{max}$  is the maximum number of genes for mutation. The features that corresponding 1 in the vector that been generating by BPSO or GA as shown in figure 2:



**Figure 2 : Features selection strategies (10-dimensional problem. Here  $i^{th}$  particle  $x_i = [1\ 0\ 1\ 1\ 1\ 0\ 0\ 1\ 0\ 1]$  indicate that the 1<sup>st</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, 5<sup>th</sup>, 8<sup>th</sup>, 10<sup>th</sup> features are selected.)**

The DHFS stops when satisfying one of the stop criteria: the algorithm reaches to the optimal solution, No change in result during several iterations, or the algorithm gets the maximum iterations. Each metaheuristic algorithm must have the cost function (object function or benchmark function) to evaluate the performance of the search processing on an algorithm. The cost function that uses with the proposed method is the classification result of the dataset by K-NN. In addition, check the validated the proposed method by CEC'15 benchmark functions.

## VII. Algorithm Parameters

**Table 1 illustrates the parameters of the algorithm**

Table1: parameters of GA, BPSO, and DHFS

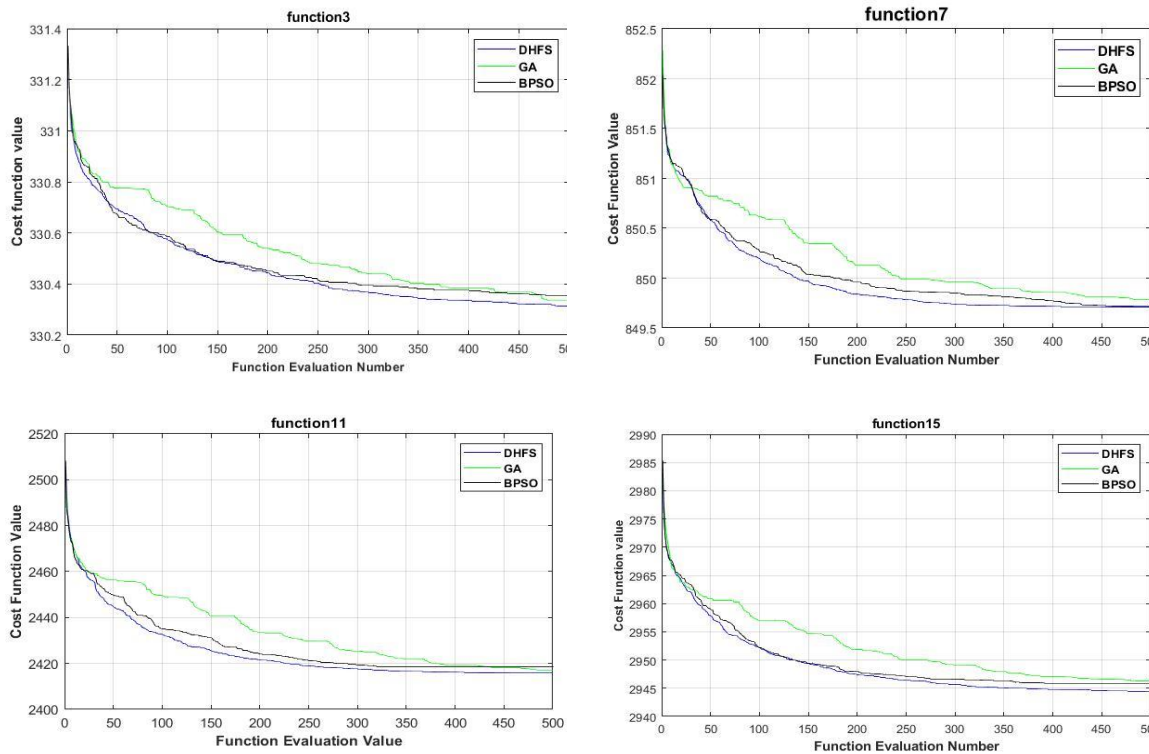
GA	Crossover rate 0.8, $\alpha$ random [0 or 1], mutation rate 0.8, $m_{min}=1$ , $m_{max} = 60\%$ of given vector.
BPSO [29]	$c1=c2=2$ , maximum velocity =6 ,minimum velocity =0.4, inertia weight ( $W_{min} = 0.4, W_{max} = 0.9$ )
DHFS	$\theta = 20$

## VIII. Discusses Validate and test algorithm

### A- Check Validate by CEC'15

After, combining two of the metaheuristic algorithms for features selection, we have to test the proposed method whether successful or not. The CEC2015 is Congress Evolutionary Computation function uses to test any given search algorithm [27] [28]. The CEC2015 has fifteen functions divided on four-group unimodal, Simple multimodal, hybrid multimodal, and composition multimodal. Four functions of CEC'15 use to check the validate of the DHFS algorithm compared to mention algorithms. Figures 3 shows the

DHFS is best in Function3, Function11, and Function15. It overcomes on most stagnation stages in search processing. The proposed algorithm failed to record better performance in Function 7, mainwheel the BPSO get the best result.



**Figure 3 : Compare the minimum value found by DHFS , BPSO , and GA over 4 functions of CEC2015, 500 iteration, 20 x100 population, and average 30 run times**

Table2 illustrates the enhancing in the standard deviation (STD) of the BPSO when combined with GA in the proposed method (DHFS). The same above function used in comparative and same parameters of CEC'15 that set (500 iterations, 20 x100 population, and average 30 runtimes.)

**Table2: (STD) of BPSO comparing with DHFS**

Function NO	STD of DHFS	STD of BPSO
Function 1	<b>104094969.6</b>	93759278.47
Function 7	0.321599518	<b>0.348831963</b>
Function 11	<b>12.26498303</b>	11.57955552
Function 15	<b>5.543502808</b>	4.880154904

### **B- Test by K-NN classification**

The proposed method test over 26 datasets from UCI machine learning [<https://archive.ics.uci.edu>]. The dataset that choice from UCI machine learning has deferent domination (Features, sample, and classes). Table 3 shows the descriptions of datasets that use in the comparative study.



**Table3: Datasets descriptions.**

SQ	Dataset name	Symbol	Size		
			classes	features	samples
1	Heart	HA	2	13	270
2	Ozone level detection	OZD	2	72	2536
3	Parkinson	PA	2	22	195
4	Hepatitis	HE	2	19	155
5	Segmentation	SE	7	19	210
6	Sonar	SO	2	60	208
7	Spectf	SP	2	44	267
8	Wine	WI	3	13	178
9	Wisc. pronostic breast cancer	WPBC	2	33	198
10	Adenocarcinoma	AD	2	9868	76
11	Brain	BN	5	5597	42
12	Brain_Tumor1	BNT1	5	5921	90
13	Brain_Tumor2	BNT2	4	10368	50
14	breast2	BR2	2	4869	77
15	breast3	BR3	3	4869	95
16	Colon	CO	2	2000	62
17	DLBCL	DL	2	5470	77
18	Leukemia	LE	3	4026	62
19	Leukemia1	LE1	3	5328	72
20	Leukemia2	LE2	3	11226	72
21	Lung_Cancer	LC	5	12601	203
22	Lymphoma	LY	2	3051	38
23	Nci	NC	8	5244	61
24	Prostate	PR	2	6033	102
25	Prostate_Tumor	PT	2	10510	102
26	SRBCT	SR	4	2309	83

Our approach test is experimenting by KNN classification techniques and table 4 shows the comparison result of the classification datasets with full features and features selection by BPSO, GA, and proposed DHFS. The numeric experiments result imply that the proposed method (DHFS) performance better compared to full features classification and with feature selection by the mentioned algorithm. The DHFS recorded the best result in the most dataset. The BPSO records the best result in the dataset name (Heart, Brain). The results of the SRBCT dataset The GA has a better result. All algorithms get the same results in the dataset (Leukemia, Lymphoma). The rest of the dataset record the best results for the proposed DHFS. Finally, by reviewing all previous results we could say that our algorithm succeeded in 80.7% of mention datasets.

**Table 4: K-NN classification results of the dataset with full features and features selection by BPSO, GA, and proposed DHFS (The K in K-NN is 5)**

SQ	Dataset Symbol	full Set	FS BPSO	FS GA	FS DHFS
1	HA	82.41	<b><u>86.42</u></b>	80.25	81.37
2	OZD	97.23	97.23	97.63	<b><u>100.00</u></b>
3	PA	91.38	92.10	89.66	<b><u>93.82</u></b>
4	HE	77.78	86.67	88.89	<b><u>94.25</u></b>
5	SE	73.02	90.48	88.89	<b><u>91.10</u></b>
6	SO	69.35	80.65	84.55	<b><u>90.32</u></b>
7	SP	73.42	86.08	82.00	<b><u>87.34</u></b>
8	WI	65.38	95.08	93.15	<b><u>98.08</u></b>
9	WPBC	64.41	74.58	77.97	<b><u>81.38</u></b>
10	AD	86.36	95.45	90.91	<b><u>96.47</u></b>
11	BN	75.00	<b><u>80.41</u></b>	75.00	77.00
12	BNT1	84.62	92.31	90.15	<b><u>96.15</u></b>
13	BNT2	78.57	90.86	85.71	<b><u>92.86</u></b>
14	BR2	64.55	78.18	81.82	<b><u>84.23</u></b>
15	BR3	57.41	62.96	66.67	<b><u>68.67</u></b>
16	CO	78.33	88.89	83.33	<b><u>91.33</u></b>
17	DL	85.91	96.00	85.91	<b><u>100.00</u></b>
18	LE	98.82	<b><u>100.00</u></b>	<b><u>100.00</u></b>	<b><u>100.00</u></b>
19	LE1	87.50	90.00	87.00	<b><u>95.00</u></b>
20	LE2	93.33	<b><u>95.87</u></b>	93.48	94
21	LC	93.05	96.61	94.31	<b><u>98.31</u></b>
22	LY	97.27	<b><u>100.00</u></b>	<b><u>100.00</u></b>	<b><u>100.00</u></b>
23	NC	70.71	85.71	92.86	<b><u>94.52</u></b>
24	PR	85.00	90.21	86.67	<b><u>94.83</u></b>
25	PT	78.67	83.33	80.00	<b><u>90.00</u></b>
26	SR	93.04	94.57	<b><u>95.33</u></b>	94.57

## I.Conclusion and Future works

The stagnation phenomena increase as search progress due to new candidate solutions are convergence. The mutation operation in GA reduce these phenomena by produce diversity in the new population. The BPSO also suffering some time from stagnation at a local optimum. The Stagnation problem increasing as search progress, therefore, the metaheuristic search algorithms need to make diversity in the population for reducing the effects of stagnation on search processing. The DHSF save on the original format of both algorithms (GA and BPSO) by calling them sequentially and

share to find the optimum solution. The future work we suggest to use other metaheuristic algorithms and adaptive it to work in the parallel model.

## CONFLICT OF INTERESTS.

There are non-conflicts of interest .

## II. Reference

- [1] Mann, C.J.H. "Handbook of Approximation: Algorithms and Metaheuristics" Kybernetes 37.2 (2008).
- [2] Siarry, Patrick, and Zbigniew Michalewicz, eds. *Advances in metaheuristics for hard optimization*. Springer Science & Business Media, 2007.
- [3] Maitra, Madhubanti, and Amitava Chatterjee. "A hybrid cooperative–comprehensive learning based PSO algorithm for image segmentation using multilevel thresholding." *Expert Systems with Applications* 34.2 (2008): 1341-1350.
- [4] Han, Jiawei, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.
- [5] Weston, Jason, et al. "Feature selection for SVMs." *Advances in neural information processing systems*. 2010.
- [6] Talavera Luis. "An evaluation of filter and wrapper methods for feature selection in categorical clustering." *International Symposium on Intelligent Data Analysis*. Springer, Berlin, Heidelberg, 2005.
- [7] Chandrashekar, Girish, and Ferat Sahin. "A survey on feature selection methods." *Computers & Electrical Engineering* 40.1 (2014): 16-28.
- [8] Zubietta, Francisco Javier Orellana. *Metaheuristics in requirements engineering: refining the next release planning problem (meta-heurísticas en ingeniería de requisitos: refinación del problema de planificación de la siguiente versión de software)*. Diss. Universidad de Almería, 2015.
- [9] Luke, Sean. *Essentials of metaheuristics*. Vol. 113. Raleigh: Lulu, 2009.
- [10] C. Ramos, A. Souza, G. Chiachia, A. Falcão, and J. Papa, "A novel algorithm for feature selection using harmony search and its application for non-technical losses detection," *Computers & Electrical Engineering*, vol. 37, no. 6, pp. 886–894, 2011.
- [11] Schiezero, Mauricio, and Helio Pedrini. "Data feature selection based on Artificial Bee Colony algorithm." *EURASIP Journal on Image and Video Processing* 2013.1(2013): 47.
- [12] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "GSA: A gravitational search algorithm," *Information Sciences*, vol. 179, no. 13, pp. 2232–2248, 2009.
- [13] Chen, Wei, Mahdi Panahi, and Hamid Reza Pourghasemi. "Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling." *Catena* 157 (2017): 310-324.
- [14] Holden, Nicholas, and Alex A. Freitas. "A hybrid PSO/ACO algorithm for discovering classification rules in data mining." *Journal of Artificial Evolution and Applications* 2008.

- [15] Chandra, Satish, Rajesh Bhat, and Harinder Singh. "A PSO based method for detection of brain tumors from MRI." *Nature & Biologically Inspired Computing*, 2009. NaBIC 2009. World Congress on. IEEE, 2009.
- [16] Najkar, Negin, Farbod Razzazi, and Hossein Sameti. "A novel approach to HMM-based speech recognition systems using particle swarm optimization." *Mathematical and Computer Modelling* 52.11-12 (2010): 1910-1920.
- [17] Deb, Kalyanmoy, et al. "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE transactions on evolutionary computation* 6.2 (2002): 182-197.
- [18] Eberhart, Russell, and James Kennedy. "A new optimizer using particle swarm theory." *Micro Machine and Human Science*, 1995. MHS'95., Proceedings of the Sixth International Symposium on. IEEE, 1995.
- [19] Clerc, Maurice. "Stagnation analysis in particle swarm optimisation or what happens when nothing happens." ,2006.
- [20] Dréo, Johann, et al. *Metaheuristics for hard optimization: methods and case studies*. Springer Science & Business Media, 2006.
- [21] Mitchell, Melanie. *An introduction to genetic algorithms*. MIT press, 1998.
- [22] Köksoy, Onur, and Tankut Yalcinoz. "Robust design using Pareto type optimization: a genetic algorithm with arithmetic crossover." *Computers & Industrial Engineering* 55.1 (2008): 208-218.
- [23] Kumar, Rakesh, Girdhar Gopal, and Rajesh Kumar. "Novel crossover operator for genetic algorithm for permutation problems." *International Journal of Soft Computing and Engineering (IJSCE)* 3.2 (2013): 252-258.
- [24] Jamuna A. "Efficient Flow based Network Traffic Classification using Machine Learning." (2013).
- [25] Tang, Jiliang, Salem Alelyani, and Huan Liu. "Feature selection for classification: A review." *Data classification: Algorithms and applications* (2014): 37.
- [26] Mirjalili, Seyedali, Seyed Mohammad Mirjalili, and Xin-She Yang. "Binary bat algorithm." *Neural Computing and Applications* 25.3-4 (2014): 663-681.
- [27] Banitalebi, Akbar, Mohd Ismail Abd Aziz, and Zainal Abdul Aziz. "A self-adaptive binary differential evolution algorithm for large-scale binary optimization problems." *Information Sciences* 367 (2016): 487-511.
- [28] Liang, J. J., et al. "Problem definitions and evaluation criteria for the CEC 2015 competition on learning-based real-parameter single objective optimization." *Technical Report201411A*, Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou China and Technical Report, Nanyang Technological University, Singapore (2014). [29] Saremi, Shahrzad, Seyedali Mirjalili, and Andrew Lewis. "How important is a transfer function in discrete heuristic algorithms." *Neural Computing and Applications* 26.3 (2015): 625-640.

## الخلاصة

اختيار الصفات هو أحد ادوات تنقيب البيانات الذي يستخدم لاختيار الصفات المهمة للبيانات المعطاة. ان الفائدة من اختيار صفات البيانات هو توفير الوقت وتقليل الذاكرة المستخدمة في معالجة البيانات. حسب تلك المبادئ صممنا خوارزمية اختيار الصفات على اساس دمج خوارزميتين من خوارزميات البحث العشوائي هما خوارزمية الأسراب الثنائية والخوارزمية الجينية لتعملاً معاً بشكل منفصل. استخدم التصنيف على اساس الجيران كدالة لتقييم عمل الخوارزمية المقترحة. فحصت وقورنت مع بيانات مصنفة بدون اختيار الصفات المهمة وبيانات مصنفة باختيار الصفات على اساس خوارزمية الأسراب الثنائية والخوارزمية الجينية. استخدمت في عملية التصنيف ٢٦ مجموعة من البيانات التابعة لـ UCI , نتائج التجارب الرقمية بينت ان الخوارزمية المقترحة أفضل مقارنة مع البيانات بدون اختيار الصفات او باختيار الصفات للخوارزميات المشار اليها سابقاً.

**الكلمات الدالة:** تنقيب البيانات - اختيار الصفات - الخوارزمية الجينية - خوارزمية الأسراب الثنائية - طرق البحث العشوائي.