An Excited Binary Grey Wolf Optimizer for Feature Selection in Highly Dimensional Datasets

Davies Segera, Mwangi Mbuthia and Abraham Nyete
Department of Electrical and Information Engineering, University of Nairobi, Harry Thuku Road, Nairobi, Kenya

Keywords: Meta-Heuristic, Feature Selection, Evolutionary Algorithm, Binary Grey Wolf Optimizer.

Abstract: Currently, feature selection is an important but challenging task in both data mining and machine learning, especially when handling highly dimensioned datasets with noisy, redundant and irrelevant attributes. These datasets are characterized by many attributes with limited sample-sizes, making classification models over-fit. Thus, there is a dire need to develop efficient feature selection techniques to aid in deriving an optimal informative subset of features from these datasets prior to classification. Although grey wolf optimizer (GWO) has been widely utilized in feature selection with promising results, it is normally trapped in the local optimum resulting into semi-optimal solutions. This is because its position-updated equation is good at exploitation but poor at exploration. In this paper, we propose an improved algorithm called excited binary grey wolf optimizer (EBGWO). In order to improve on exploration, a new position-updating criterion is adopted by utilizing the fitness values of vectors $\vec{X}_1, \vec{X}_2$ and $\vec{X}_3$ to determine new candidate individuals. Moreover, in order to make full use of and balance the exploration and exploitation of the existing BGWO, a novel nonlinear control parameter strategy is introduced, i.e. the control parameter of $\vec{a}$ is innovatively decreased via the concept of the complete current response of a direct current (DC) excited resistor-capacitor (RC) circuit. The experimental results on seven standard gene expression datasets demonstrate the appropriateness and efficiency of the fitness value based position-updating criterion and the novel nonlinear control strategy in feature selection. Moreover, EBGWO achieved a more compact set of features along with the highest accuracy among all the contenders considered in this paper.

1 INTRODUCTION

The major challenge in analysing big data is the vast number of features. Out of the many available features, only a few of them will be useful in distinguishing samples that belong to different classes while majority of the of the features will be irrelevant, noise, or redundant. Foremost, these irrelevant features lead to noise generation in big data analysis. In addition, they result in increased dataset dimensions and a further computational complexity in both clustering and classification operations. All these consequently decreases the rate of classification accuracy. Thus, superior approaches are needed to identify diverse features, compute the relationship between the features and optimally select informative attributes from these highly dimensioned datasets (Almugren & Alshamlan, 2019).

For a dataset containing N number of features, there exists 2^N number of candidate subsets. The main objective of designing different feature selection techniques has always been to determine a compressed and optimal subset with the highest precision among the possible candidate subsets.

Since the scope of possible solutions is wide and the size of this set of responses is on the rise due to the ever-increasing number of features, determining the best subset of N features is extremely difficult and costly(Pirgazi et al., 2019), (Liang et al., 2018).

Feature selection techniques can be broadly categorized into two i.e. filters and wrappers. Filter approaches utilizes the dependency, mutual information, distance, and information theory in carrying out feature selection(Shunmugapriya & Kanmani, 2017). Unlike filters, wrappers utilize classifiers as the learning algorithm in optimizing the classification performance by selecting the informative features. Commonly, filter techniques are often faster compared to wrappers, which is largely attributed to their reduced computational complexity (Sun et al., 2018). Nevertheless, wrapper techniques can usually offer better performances compared to filters(Zorarpacı & Özel, 2016). Wrappers apply metaheuristic optimization approaches, such as

Segera, D., Mbuthia, M. and Nyete, A.
An Excited Binary Grey Wolf Optimizer for Feature Selection in Highly Dimensional Datasets.
DOI: 10.5220/0009805101250133
In *Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2020)*, pages 125-133
ISBN: 978-989-758-442-8

binary genetic algorithm (BGA)(De Stefano et al., 2014), binary grey wolf optimization (BGWO)(Emary et al., 2016), binary ant colony optimization (BACO) (Aghdam et al., 2009), binary particle swarm optimization (BPSO) (He et al., 2009), to select the optimal informative feature subsets.

BGWO is a recent feature selection approach, which usually offers better performance compared to other conventional methods(Emary et al., 2016) . However, the wolves' new positions are based on the experience of leaders i.e. alpha, beta and delta, which normally leads to premature convergence. In addition, a proper balance between the diversification (global search) and intensification (local search) is still the challenging issue in BGWO (Too et al., 2018).

In this paper, we propose a new excited binary grey wolf optimizer (EBGWO) that aims to improve the performance of BGWO(Emary et al., 2016) in selecting informative features in highly dimensioned microarray datasets. Foremost, to overcome the insufficiency of the existing BGWO regarding its position-updated equation, which is good at exploitation but poor at exploration, a new position-updated equation utilizing the fitness values of vectors $\vec{X}_1, \vec{X}_2$ and $\vec{X}_3$ is proposed to determine new candidate individuals. Moreover, inspired by the concept of the complete current response of a direct current (DC) excited resistor-capacitor (RC) circuit, a new nonlinear control parameter strategy for parameter $\vec{a}$ is introduced in order to make full use of and balance the diversification and intensification of the existing BGWO algorithm.

The performance of EBGWO is tested using seven standard gene expression datasets. To evaluate the effectiveness of proposed method, EBGWO is compared with five existing binary metaheuristic algorithms i.e. BGWO1, BGWO2, BPSO, BDE and BGA. The experimental results indicate EBGWO has a very efficient computational complexity, while keeping a comparative performance in feature selection.

The rest of paper is organized as follows. Preliminaries for the work, GWO is presented in section 2. The proposed excited grey wolf optimizer (EGWO) is presented in section 3. The binary version of EGWO i.e. Excited binary grey wolf optimizer (EBGWO) is presented in section 4. Section 5 reports the experimental setting and a discussion of the obtained results. Finally, a conclusion and future works are given in section 6.

2 GREY WOLF OPTIMIZER

Grey wolf optimizer (GWO) is a recently proposed metaheuristic optimization technique(Mirjalili et al., 2014). In nature, grey wolves live in groups ranging between 5 to 12. GWO mimics the behaviour portrayed by these grey wolves while hunting and searching of a prey. In GWO, to simulate the leadership hierarchy in a pack, the population is divided into four types of wolves i.e. Alpha (α), beta (β), delta (δ), and omega (ω). The alpha wolf is the overall leader responsible for decision-making. The beta wolf is the second in command is aids the alpha in making the decision or other activities. Delta wolf is referred as the third leader in the group, which dominates the omega wolves. The three leaders i.e. α, β and δ guide the hunting (optimization) while the remaining omega wolves (ω) follow them(El-Gaafary et al., 2015).

Equation 1 depicts the encircling behaviour of the pack while hunting a prey.

$$X(t+1) = X_p(t) - A.D \tag{1}$$

Where X_p is the position of prey, A is the coefficient vector, and D is defined by equation 2.

$$D = \left|C.X_p(t) - X(t)\right| \tag{2}$$

Where C is the coefficient vector, X is the position of grey wolf, and t is the number of iterations.

The coefficient vectors, A and C, are determined by equations 3 and 4 respectively.

$$A = 2.a.r_1 - a \tag{3}$$

$$C = 2.r_2 \tag{4}$$

Where r_1 and r_2 are two independent random numbers uniformly distributed between $[0,1]$, and a is the encircling coefficient that is used to balance the trade-off between exploration and exploitation.
In GWO, a is linearly decreased, from 2 to 0, according to Equation (5).

$$a = 2 - 2.(\frac{t}{T}) \tag{5}$$

Where t and T represent the number of iterations and maximum number of iterations respectively.

In GWO, the three leaders α, β and δ leaders are deemed knowledgeable of the potential position of the prey. Thus, they guide the remaining omega wolves to move toward the optimal position.

Mathematically, the new position of wolf is updated as per Equation (6).

$$X(t+1) = \frac{(X_1+X_2+X_3)}{3} \quad (6)$$

Where X_1, X_2 and X_3 are determined according to Equations (7)-(9)

$$X_1 = |X_\alpha - A_1.D_\alpha| \quad (7)$$

$$X_2 = |X_\beta - A_2.D_\beta| \quad (8)$$

$$X_3 = |X_\delta - A_3.D_\delta| \quad (9)$$

Where X_α, X_β and X_δ are the position of α, β and δ respectively during iteration t.

A_1, A_2 and A_3 are determined using Equation 3; and D_α, D_β and D_δ are defined in Equations (10)-(12) respectively.

$$D_\alpha = |C_1.X_\alpha - X| \quad (10)$$

$$D_\beta = |C_2.X_\beta - X| \quad (11)$$

$$D_\delta = |C_3.X_\delta - X| \quad (12)$$

Where C_1, C_2 and C_3 are determined using Equation (4).

3 EXCITED GREY WOLF OPTIMIZER (EGWO)

3.1 Nonlinearly Controlling Parameter a via the Complete Current Response of the DC Excited RC Circuit

It is a well-established fact that for population-based metaheuristics, both exploration (diversification) and exploitation (intensification) are conducted concurrently.

Exploration is termed as the ability of a population-based metaheuristic to examine new areas within the defined search space with the aim of determining the global optima. On the hand, exploitation is the ability to utilize the information of already identified individuals in deriving better individuals(Long et al., 2018; Luo et al., 2013).

In every population-based metaheuristic, both exploration (diversification) and exploitation (intensification) abilities are attained by applying specific operators.

In the conventional GWO algorithm, the control parameter a plays a critical role in balancing between diversification and intensification of an individual candidate search (Long et al., 2018). A larger value of a enhances global exploration, while its smaller value promotes local exploitation. Thus, selection of a suitable control strategy for parameter a is critical in attaining an effective balance between local exploitation and global exploration. From literature, one proved way to achieve the required balance is critically studying the control of parameter a. To date, various approaches have been proposed to control the conventional GWO's parameter a (Long, 2016; Long et al., 2018; Mittal et al., 2016).

However, in the conventional GWO, a linearly decreases from 2 to 0 using Equation (5). Since GWO incorporates a highly complicated nonlinear search process, the utilized linear control of parameter a doesn't reflect the actual search process (Long et al., 2018). Moreover, (Mittal et al., 2016) suggested that the performance of GWO would improve if parameter a is nonlinearly controlled.

Motivated by both the above consideration and the complete current response of a direct current (DC) excited resistor-capacitor (RC) circuit(Alexander & Sadiku, 2016), a novel nonlinear adaptation of parameter a is proposed in this paper.

The complete current response of the RC circuit to a sudden application of a dc voltage source, with the assumption that the capacitor is initially not charged is given in Equation 13.

$$i(t) = \frac{V_s}{R}((\frac{1}{e^t})^\tau) \quad (13)$$

Where $\tau = R \times C$ is the time constant that expresses the rapidity with which the value if i decreases from the initial value $\frac{V_s}{R}$ to zero over time . V_s is value of a constant DC voltage while R and C are the resistor and capacitor values of the circuit.

We adopt this concept i.e. the exponential decay of i over time to develop a new nonlinear control strategy of parameter a as presented in Equation 14.

$$a_{i,t} = a_{initial} \times (\frac{MaxIter - t}{MaxIter})^{\tau_{i,t}} \quad (14)$$

Where $a_{i,t}$ is the value of the control parameter a assigned to grey wolf i during iteration t. $MaxIter$ indicates the total number of iterations (generations) and $a_{initial}$ is the initial value of the control parameter a. $\tau_{i,t}$ is a nonlinear modulation index assigned to the grey wolf i during iteration t.

To ensure that the value of the control parameter $a_{i,t}$ is proportional to the fitness value of grey wolf i during iteration t, a new formulation of the value of the nonlinear modulation index $\tau_{i,t}$ is given in Equation (15).

$$\tau_{i,t} = \left| \frac{(\frac{F\alpha_t + F\beta_t + F\delta_t}{3}) - FX_t}{(\frac{F\alpha_t + F\beta_t + F\delta_t}{3}) - Fw_t} \right| \tag{15}$$

Where $F\alpha_t$, $F\beta_t$ and $F\delta_t$ are the fitness values of α, β and δ wolves (the 3 best wolves) respectively during the current iteration t. FX_t is the fitness value of grey wolf i during iteration t and finally Fw_t is the worst fitness value among the omega (ω) wolves during iteration t .

Consequently, A_1, A_2 and A_3 are determined using Equation (16) which is a variant of Equation 3.

$$A = 2.a_{i,t}.r_1 - a_{i,t} \tag{16}$$

From the literature of conventional GWO (Mirjalili et al., 2014), when $A < 1$ the grey wolves are compelled to attack the prey (exploitation) and when $A > 1$ the grey wolves are compelled to move away from the current prey with the hope of finding another fitter prey. This implies that smaller values of the control parameter a promotes local exploitation while larger values facilitates global exploration.

According to Equation (5) of the conventional GWO algorithm, it is evident that half of the iterations are committed to exploration and the remaining half to exploitation. This strategy fails to consider the impact of effective balancing between these two conflicting milestones to guarantee accurate approximation of the global optimum.

The nonlinear control strategy of parameter a proposed in Equation (14), tries to overcome this challenge by adopting a variant of decay function to strike a proper balance between exploration and exploitation. Since this strategy promotes allocates a large proportion of the iterations to global exploration compared to local exploitation, the convergence speed of the proposed EGWO algorithm is enhanced while minimizing the local minima trapping effect.

Moreover, since the proposed scheme is proportional to the fitness values of the individual grey wolves in the search space and the current iteration (generation), diversity and the quality of the solutions is enhanced.

3.2 Socially Strengthened Hierarchy via a Fitness-value based Position-updating Criterion

In the conventional GWO, social hierarchy is the cornerstone in both the internal management and the hunting patterns of the pack(Tu et al., 2019) . All the wolves in the pack conduct hunting under the guidance of the α, β and δ wolves. An assumption that these three dominant wolves have a better knowledge of the prey's location. Consequently, the omega (ω) wolves update their positions with the aid of these three leaders during the hunting process. This implies that the conditions of the α, β and δ wolves are key in updating the whole pack. Meanwhile, the higher the rank a wolf attains during the search, the closer it gets to the global optimum.

In addition, all the wolves including the three leaders utilize Equation (6) to update their positions. That is to say the α wolf will utilize the lowly ranked β and δ wolves to update its position. Likewise, β wolf will utilize the lowly ranked δ wolf to update itself. Since the conditions of the β and δ wolves are worse compared to that of the α wolf, there are higher chances that the two wolves will compel the α wolf to move away from the global optimum. Likewise, β wolf may also be misled by the δ wolf. Ultimately, the accumulative error will have an adverse effect on updating the positions of the all the wolves in the pack and the convergence efficiency of the GWO will drastically reduce(Tu et al., 2019).

On the other hand, since all the omega (ω) wolves are attracted towards the α, β and δ wolves, they may prematurely converge due to limited exploration within the search space. Thus, the conventional GWO is good at exploitation but poor at exploration.

Thus, to overcome the GWO's premature converge and still maintain the social hierarchy of the pack, a different scheme for updating both the dominant (α, β and δ) and the omega (ω) wolves is needed. To attain this, a new position-updated equation utilizing the fitness values of vectors $\vec{X}_1$, $\vec{X}_2$ and $\vec{X}_3$ is utilized in determining new candidate individuals.

Foremost, for each wolf in the pack, vectors X_{vec1}, X_{vec2} and X_{vec3} are computed using Equations (17)-(19).

$$X_{vec1} = \bigcup_{j=1}^{d} X_1(j) \tag{17}$$

$$X_{vec2} = \bigcup_{j=1}^{d} X_2(j) \tag{18}$$

$$X_{vec3} = \bigcup_{j=1}^{d} X_3(j) \tag{19}$$

Where d is the dimension of the search space and $X_1(j), X_2(j)$ and $X_3(j)$ are determined using Equations (7)-(9) respectively.

Next, the fitness values FX_{vec1}, FX_{vec2} and FX_{vec3} for vectors X_{vec1} , X_{vec2} and X_{vec3} respectively are determined and the one with the best fitness forms the new position as depicted by Equations (20)-(21).

$$[fittest, Pos] = min\left(\bigcup_{a=1}^{3} FX_{vec(a)}\right) \tag{20}$$

$$X(t+1) = \left(\bigcup_{a=1}^{3} X_{vec(a)}\right)_{Pos} \tag{21}$$

4 EXCITED BINARY GREY WOLF OPTIMIZER (EBGWO)

Feature selection (FS) is a significant problem in pattern recognition and machine learning areas. The aim of FS is to select the most informative feature subset guided by a given evaluation criterion(Salesi & Cosma, 2017; Tu et al., 2019).

In essence, FS is a broad-based optimization problem that is characterized by huge computations.

Since the FS problem utilizes a binary search space, it is important to convert the proposed continuous EGWO to binary i.e. EBGWO. One of the commonly adopted approach for this transformation is the utilization of transfer functions(Salesi & Cosma, 2017; Tu et al., 2019).

In our experiments, the transfer function we utilized in converting the real values of each solution to binary is depicted by Equation (22).

$$X^j(t+1) = \begin{cases} 1, if\ S\left(X^j(t+1)\right) > \rho, \\ 0, \quad otherwise \end{cases} \tag{22}$$

Where $\rho \in [0,1]$ depict a random threshold and S is the considered sigmoid function as expressed by Equation (23).

$$S(x) = \frac{1}{1 + \exp(-10(x - 0.5))} \tag{23}$$

$X^j(t+1)$=1 imply that the j^{th} element of $X(t+1)$ is selected as an informative attribute while $X^j(t+1)$=0 imply that the corresponding j^{th} element is ignored.

For instance, if $X(t+1) = [0.55, 0.21, 0.35, 0.8]$ and $\rho = 0.5$, the output of Equation (21) becomes $X(t+1) = [1, 0, 0, 1]$ which imply that the 1^{st} and 4^{th} features be selected while the 2^{nd} and 3^{rd} features will be ignored.

By doing so, the number of features is reduced without affecting the classification performance.

Since the FS task aims at attaining better classification accuracy with the utilization of fewer attributes, the objective function Fit utilized in this paper is given by Equation (24)(Salesi & Cosma, 2017).

$$Fit = \varepsilon * \frac{|S|}{|N|} - ((1 - \varepsilon) * Acc) \tag{24}$$

Where Acc is indicates the accuracy of a given classifier, $|S|$ is the size of the selected feature subset and $|N|$ is the number of the total features of a dataset. Thus, FS is turned into a problem of determining the minimum value of Equation (24).

Herein, ε and $(1 - \varepsilon)$ are weights corresponding to the feature subset size and average accuracy respectively. The parameter of ε in Equation (24) is set 0.2 (Salesi & Cosma, 2017).

The pseudocode of the proposed excited binary grey wolf optimizer (EBGWO) algorithm is presented in Algorithm 1.

Algorithm 1: Pseudo-code for the EBGWO.

Input: labelled gene dataset *D*, Total number of iterations $MaxIter$, Population size *N*, Initial value of the control parameter $a_{initial}$

Output: Optimal Individual's position X_α , Best fitness value *Fit* (X_α)

1. Randomly initialize *N* individuals' positions to establish a population
2. Using Equation (23), evaluate the fitness of all wolves, *Fit* (X)
3. *[~, Index] =Sort* ($Fit\ (X), 'Ascend'$)
4. $F\alpha$ =$Fit\ (X)_{Index(1)}$
5. $F\beta$ =$Fit\ (X)_{Index(2)}$
6. $F\delta$ =$Fit\ (X)_{Index(3)}$
7. Fw =$Fit\ (X)_{Index(N)}$
8. X_α=$X(Index(1))$
9. X_β=$X(Index(2))$
10. X_δ=$X(Index(3))$
11. ***For*** *t=1* ***To*** $MaxIter$
12. ***For*** *i=1* ***To*** *N*
13. *Determine* $a_{i,t}$ *using Equation (14)*
14. *Compute* X_{vec1}, X_{vec2} *and* X_{vec3} *using Equations (17)-(19)*
15. *Generate* $X_{vec1}{}^{new}$, $X_{vec2}{}^{new}$ *and*

$X_{vec3}{}^{new}$ using Equation (21)
16. *Evaluate the fitness values FX_{vec1}, FX_{vec2} and FX_{vec3} of the binary vectors $X_{vec1}{}^{new}$ $X_{vec2}{}^{new}$ and $X_{vec3}{}^{new}$ respectively using Equation (24)*
17. *Determine the minimum value(fittest) of the three evaluated fitness values and its Index using Equations (20)*
18. ***If** (fittest<$Fit\ (X)_{Index(i)}$) **Then***
19. *$Fit\ (X)_{Index(i)}$= fittest*
20. *Update $X_{Index(i)}$ using Equation (21)*
End If
21. ***Next** i*
22. *Repeat steps 3 to 10*
23. ***Next** t*

5 EXPERIMENTAL RESULTS AND DISCUSSION

All the experiments were conducted in Windows Windows 10 Home Single Language 64-bit operating system; processor Intel(R) Core (TM) i7-3770CPU , processor speed of 3.4GHZ; 12GB of RAM. All the considered algorithms were implemented using MATLAB 2017 environment.

5.1 Dataset Description

In order to evaluate the performance of the proposed algorithm, seven gene expression datasets were utilized. The datasets were selected to have a variety of instances (sample-size), genes and classes as a representative of various issues. Table 1 outlines the detailed distribution of instances, genes and classes for each considered dataset.

Table 1: Microarray datasets used in the experiments.

Dataset	No. of Instances	No. of Genes	No. of Classes
Brain_Tumour1	90	5920	5
Brain_Tumour2	50	10367	4
CNS	60	7129	2
DLBCL	77	5469	2
Leukemia	72	7129	2
Colon	62	2000	2
Lung Cancer	203	12600	4

5.2 Parameter Setting

The proposed EBGWO is benchmarked with two novel binary grey wolf optimizations (i.e. BGWO1 and BGWO2) (Emary et al 2016), binary particle swarm optimization (BPSO), binary differential evolution (BDE) and binary genetic algorithm(Too et al., 2019). The optimizer- specific settings of the considered algorithms are presented in Table 2.

Table 2: Parameter settings for each considered algorithm.

Algorithm	Year	Parameter settings
EBGWO		*N=10*, $MaxIter = 100$, $a_{initial} = 2$
BGWO1	2016	*N=10*, $MaxIter = 100$, $a_{initial} = 2$
BGWO2	2016	*N=10*, $MaxIter = 100$, $a_{initial} = 2$
BPSO	2019	*N=10*, $MaxIter = 100$, $C_1 = C_2 = 2$, $V_{max} = 6$, $W_{max} = 0.9$, , $W_{min} = 0.4$
BDE	2019	*N=10* , $MaxIter = 100$, $CR = 0.9$
BGA	2019	*N=10*, $MaxIter = 100$, $CR = 0.8$, *MR=0.01*

Additionally, all the considered algorithms are repeated over 10 independent runs to ensure both stability and statistical significance of the obtained results. Furthermore, the commonly utilized 10-fold cross validation scheme is used to divide the considered microarray datasets into training and testing (Arlot & Celisse, 2010).

A wrapper approach based on the K-Nearest Neighbour (K-NN) classifier(Emary et al., 2016; Pirgazi et al., 2019) is used for feature selection in this paper. The K-NN classifier (where k=5) is utilized to obtain the classification accuracy of the solutions.

Table 3 presents the performance of all the algorithms considered for the feature selection task using the gene expression datasets whose details are presented in Table 1.

The following information is presented in each column of Table 3:

i) Algorithm: Provides the abbreviations of the considered algorithms i.e. Excited Binary Grey Wolf Optimizer (EBGWO), Binary Grey Wolf Optimizer 1(BGWO1), and Binary Grey Wolf Optimizer 2 (BGWO2)
ii) *Max_Acc* Maximum Accuracy value obtained when a given algorithm is repeated for 10 independent runs.

Table 3: Experimental Results.

Algorithm	**Accuracy**			**Number of Genes**			**Dataset**
	Max_Acc	Min_Acc	Avg_Acc	Max_Nfeat	Min_Nfeat	Avg_Nfeat	
EBGWO (Ours)	**0.933**	**0.911**	**0.919**	**673**	**440**	**501.9**	Brain_Tumour1
BGWO1	0.889	0.856	0.871	*3831*	*2952*	*3356.9*	
BGWO2	0.911	0.878	0.894	1656	1094	1343.3	
BPSO	*0.854*	*0.823*	*0.843*	2972	2763	2863.9	
BDE	0.864	0.834	0.854	3017	2737	2937.6	
BGA	0.869	0.844	0.859	2950	2840	2889.4	
EBGWO (Ours)	**0.920**	**0.84**	**0.884**	**2811**	**712**	**1151.5**	Brain_Tumour2
BGWO1	0.840	0.820	0.838	*7415*	*6103*	*6813.4*	
BGWO2	0.880	0.820	0.846	4019	2528	3083.8	
BPSO	0.800	0.780	0.798	5126	5090	5122.4	
BDE	*0.728*	*0.713*	*0.714*	5198	5076	5172.3	
BGA	0.767	0.753	0.752	5139	5039	5089.5	
EBGWO (Ours)	**0.85**	**0.8**	**0.827**	**1020**	**564**	**710.3**	CNS
BGWO1	0.783	0.750	0.760	*4942*	*4217*	*4606.4*	
BGWO2	0.800	0.750	0.780	2502	1842	2175.8	
BPSO	0.767	0.733	0.737	3502	3486	3487.6	
BDE	*0.693*	*0.663*	*0.683*	3530	3478	3521.9	
BGA	0.727	0.707	0.717	3528	3428	3501.7	
EBGWO (Ours)	**1.000**	**0.987**	**0.997**	**534**	**333**	**426.7**	DLBCL
BGWO1	0.987	0.961	0.971	*3706*	*2826*	*3343.4*	
BGWO2	**1.000**	0.948	0.986	1700	1002	1408.3	
BPSO	0.919	0.891	0.901	2703	2672	2675.1	
BDE	*0.885*	*0.869*	*0.882*	2732	2687	2721.4	
BGA	0.906	0.883	0.896	2709	2699	2685.1	
EBGWO (Ours)	**0.931**	**0.889**	**0.903**	**913**	**524**	**649.8**	Leukemia
BGWO1	0.861	0.833	0.849	*5065*	*3897*	*4428.5*	
BGWO2	0.889	0.847	0.874	2141	1618	1805.5	
BPSO	0.828	0.809	0.814	3516	3505	3514.9	
BDE	*0.782*	*0.751*	*0.784*	3537	3527	3531.2	
BGA	0.801	0.782	0.792	3501	3461	3481.8	
EBGWO (Ours)	**0.935**	**0.903**	**0.919**	**220**	**103**	**143.4**	Colon
BGWO1	0.887	0.855	0.865	*1316*	*1096*	*1189.4*	
BGWO2	0.919	0.871	0.900	622	351	455.2	
BPSO	0.849	0.829	0.839	986	931	936.5	
BDE	*0.810*	*0.780*	*0.794*	995	955	965.3	
BGA	0.881	0.875	0.878	990	984	987.3	
EBGWO (Ours)	**0.985**	**0.970**	**0.977**	**1148**	**781**	**1005.5**	Lung Cancer
BGWO1	0.966	0.941	0.951	*7598*	*6621*	*7211*	
BGWO2	0.975	0.956	0.966	2672	2167	2413.2	
BPSO	0.936	0.931	0.935	6196	6179	6180.7	
BDE	*0.931*	*0.921*	*0.924*	6256	6218	6226.8	
BGA	0.952	0.939	0.945	6235	6214	6218.2	

Values in bold represent the best result and values in italic denote the worst in each column, respectively.

iii) Min_Acc Minimum Accuracy value obtained when a given algorithm is repeated for 10 independent runs.
iv) Avg_Acc: Is the average of all the accuracy values obtained when a given algorithm is repeated for 10 independent runs.
v) Max_Nfeat: Is the Maximum number of features reported by a given algorithm during the 10 independent runs.
vi) Min_Nfeat: Is the Minimum number of features reported by a given algorithm during the 10 independent runs.
vii) Avg_Nfeat Is the average of all the number of features reported by a given algorithm during the 10 independent runs.
viii) Dataset: Captures the datasets utilized for experimentation as articulated in Table 1.

It is important to point out that the best result achieved in each column for all the considered gene expression datasets is highlighted in bold while the worst is italicized.

Concerning the classification accuracy, as presented in Table 3, the proposed EBGWO algorithm outperformed all the competing when the fitness function (refer to Equation 24) was adopted.

Concerning the selection of the informative subset of genes, again the proposed EBGWO identified subsets with the least number of features to achieve the highest classification accuracy for all the seven highly dimensional microarray datasets considered in this paper.

6 CONCLUSION AND FUTURE WORKS

An excited grey wolf optimizer (EGWO) is proposed in this paper. In the proposed algorithm, the concept of the complete current response of a direct current (DC) excited resistor capacitor (RC) circuit are innovatively utilized to make the non-linear control strategy of parameter a of the GWO adaptive. Since this scheme allocates a large proportion of the number of iterations to global exploration compared to local exploitation, the convergence speed of the proposed EGWO algorithm is enhanced while minimizing the local minima trapping effect. Moreover, since the proposed scheme assigns each wolf a value of parameter a that is proportional its fitness values in both the search space and the current iteration (generation), diversity and the quality of the solutions is improved as well.

To overcome premature converge (a limitation of existing versions of GWO algorithms) and still maintain the social hierarchy of the pack, a new position-updated equation utilizing the fitness values of vectors $\vec{X}_1, \vec{X}_2$ and $\vec{X}_3$ is proposed in determining new candidate individuals.

As a feature selector, EBGWO is compared with five metaheuristic algorithms i.e. BGWO1, BGWO2, BPSO, BDE and BGA that are in existence. The obtained experimental results revealed that EBGWO yielded the best performance and overtook the other algorithms. EBGWO not only attained the highest classification accuracy, but also selected subsets with the least number of informative features (genes). In conclusion, the proposed EBGWO is successful, and more appropriate to be used as a feature selector in highly dimensional datasets. For further works, a chaotic map can be adopted to fine-tune the parameters of the EBGWO. Utilizing EBGWO as a hybrid filter-wrapper for feature selection seeking to evaluate the generality of the selected attributes will be another valuable contribution. Moreover, EGWO will be applied to other optimization areas, such as training neural network, knapsack, and numerical problems.

REFERENCES

Aghdam, M. H., Ghasem-Aghaee, N., & Basiri, M. E. (2009). Text feature selection using ant colony optimization. Expert Systems with Applications, 36(3, Part 2), 6843–6853. https://doi.org/10.1016/j.eswa.2008.08.022

Alexander, C., & Sadiku, M. (2016). Fundamentals of Electric Circuits. https://www.mheducation.com/highered/product/fundamentals-electric-circuits-alexander-sadiku/M9780078028229.html

Almugren, N., & Alshamlan, H. (2019). A Survey on Hybrid Feature Selection Methods in Microarray Gene Expression Data for Cancer Classification. *IEEE* Access, 7, 78533–78548. https://doi.org/10.1109/ACCESS.2019.2922987

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*, 40–79. https://doi.org/10.1214/09-SS054

De Stefano, C., Fontanella, F., Marrocco, C., & Scotto di Freca, A. (2014). A GA-based feature selection approach with an application to handwritten character recognition. *Pattern Recognition Letters, 35, 130–141.* https://doi.org/10.1016/j.patrec.2013.01.026

El-Gaafary, A. A. M., Mohamed, Y. S., Hemeida, A. M., & Mohamed, A.-A. A. (2015). Grey Wolf Optimization for Multi Input Multi Output System. *Universal Journal of Communications and Network, 3(1), 1–6.* https://doi.org/10.13189/ujcn.2015.030101

Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary grey wolf optimization approaches for feature selection. *Neurocomputing, 172, 371–381*. https://doi.org/10.1016/j.neucom.2015.06.083

Long, W. (2016, December). Grey Wolf Optimizer based on Nonlinear Adjustment Control Parameter. *2016 4th International Conference on Sensors, Mechatronics and Automation (ICSMA 2016)*. https://doi.org/10.2991/icsma-16.2016.111

Long, W., Jiao, J., Liang, X., & Tang, M. (2018). An exploration-enhanced grey wolf optimizer to solve high-dimensional numerical optimization. *Engineering Applications of Artificial Intelligence, 68, 63–80*. https://doi.org/10.1016/j.engappai.2017.10.024

Luo, J., Wang, Q., & Xiao, X. (2013). A modified artificial bee colony algorithm based on converge-onlookers approach for global optimization. *Applied Mathematics and Computation, 219(20)*, 10253–10262. https://doi.org/10.1016/j.amc.2013.04.001

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. *Advances in Engineering Software, 69, 46–61*. https://doi.org/10.1016/j.advengsoft.2013.12.007

Mittal, N., Singh, U., & Sohi, B. S. (2016). Modified Grey Wolf Optimizer for Global Engineering Optimization [Research Article]. *Applied Computational Intelligence and Soft Computing*. https://doi.org/10.1155/2016/7950348

Pirgazi, J., Alimoradi, M., Esmaeili Abharian, T., & Olyaee, M. H. (2019). An Efficient hybrid filter-wrapper metaheuristic-based gene selection method for high dimensional datasets. *Scientific Reports, 9(1)*, 18580. https://doi.org/10.1038/s41598-019-54987-1

Salesi, S., & Cosma, G. (2017). A novel extended binary cuckoo search algorithm for feature selection. *2017 2nd International Conference on Knowledge Engineering and Applications (ICKEA), 6–12*. https://doi.org/10.1109/ICKEA.2017.8169893

Shunmugapriya, P., & Kanmani, S. (2017). A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid). *Swarm Evol. Comput., 36*. https://doi.org/10.1016/j.swevo.2017.04.002

Sun, Y., Lu, C., & Li, X. (2018). The Cross-Entropy Based Multi-Filter Ensemble Method for Gene Selection. *Genes, 9(5)*. https://doi.org/10.3390/genes9050258

Too, J., Abdullah, A. R., Mohd Saad, N., Mohd Ali, N., & Tee, W. (2018). A New Competitive Binary Grey Wolf Optimizer to Solve the Feature Selection Problem in *EMG Signals Classification. Computers, 7(4), 58*. https://doi.org/10.3390/computers7040058

Too, J., Abdullah, A. R., Mohd Saad, N., & Tee, W. (2019). EMG Feature Selection and Classification Using a Pbest-Guide Binary Particle Swarm Optimization. *Computation, 7(1), 12*. https://doi.org/10.3390/computation7010012

Tu, Q., Chen, X., & Liu, X. (2019). Hierarchy Strengthened Grey Wolf Optimizer for Numerical Optimization and Feature Selection. *IEEE Access, 7, 78012–78028.* https://doi.org/10.1109/ACCESS.2019.2921793.

Zorarpacı, E., & Özel, S. A. (2016). A hybrid approach of differential evolution and artificial bee colony for feature selection. *Expert Systems with Applications, 62, 91–103*. https://doi.org/10.1016/j.eswa.2016.06.004.