Opposition-Based Enhanced Grey Wolf Optimization Algorithm for Feature Selection in Breast Density Classification

Rahul Hans and Harjot Kaur

Abstract—**Breast Cancer is the most widespread cancer amongst women in the world. An increase in the density of breast may lead to an increase in the risk of breast cancer. For the accurate diagnosis of breast cancer, these days Machine learning-based computer-aided diagnosis systems are widely used to assist the radiologists. However, there is still a scope of improvement in the computer-aided diagnosis system for the feature selection problem, which is considered as an optimization problem. Optimization techniques are the methods where the finest solution to a problem is found by using random search mechanism. Grey wolf optimization is one of the most recent optimization algorithms. This paper presents an enhanced version of Grey Wolf Optimization algorithm and tests the performance of the proposed version of the algorithm on thirteen different benchmark functions and the binary version of the proposed algorithm is used for selecting optimal number of features for breast density classification problem.**

Index Terms—**Breast density classification, feature selection grey wolf optimization, opposition based learning.**

I. INTRODUCTION

In recent studies, it has been found that the increase in the density of breast may lead to an increase in the probability of having breast cancer which is the major cause of death among women globally [1]. One of the most successful screening tools for breast cancer detection these days is Mammography. However, Radiologists estimate the breast density by visual inspection of the mammogram images. These days Machine learning-based Computer-Aided diagnosis Systems (CAD) are widely used, which assists the radiologist in the diagnosis of breast cancer, however for the exact diagnosis by CADs researchers still find room for improvement in the feature selection techniques. A feature can be defined as discrete quantifiable importance of the method being experiential.

The search process of finding an optimal feature subset is a complex task, as the prime aim is to find the best feasible solution in an acceptable time, which can be defined as an optimization problem [2]. In recent times Metaheuristic

Manuscript received September 24, 2019; revised January 11, 2020.
Rahul Hans is with the Department of Computer Science and Engineering, DAV University, Jalandhar and also with Guru Nanak Dev University, Amritsar, Punjab, India (e-mail: rahulhans@gmail.com).
Harjot Kaur is with the Department of Computer Science and Engineering, Guru Nanak Dev University, Regional Campus, Gurdaspur, Punjab India (e-mail: harjotkaursohal@rediffmail.com).

algorithms, have gained the attention of various researchers due to their much better capability to solve complex problems. This paper puts forward an enhanced version of the Grey Wolf Optimization algorithm (GWO) for solving feature selection problem in the breast density classification. The rest of this paper is structured as follows: Section II presents the related work. Section III throws light upon the concept of GWO algorithm, Section IV proposes Opposition based Enhanced version of GWO (OEGWO), Section V, presents materials and methods. Section VI presents the results and discussions. Finally, in Section VII, conclusions and future works are presented.

II. RELATED WORKS

As stated in [3] it can be perceived that not all the metaheuristic algorithms are suitable for getting to the bottom of every type of optimization problem. Which is the reason, the researchers keep on trying to enhance the capability of optimization algorithms; In [4] authors proposed the use of moth-flame optimization algorithm for feature selection and evaluated the performance on eighteen different datasets. Authors in [5] proposed binary variants of multi-verse optimization algorithm, by using the concept of transfer functions to get into the bottom of feature selection problem. The results indicated that the binary versions of the proposed algorithms outperform the algorithms considered for performance comparison. In [6] authors used simulated annealing as a local search mechanism with whale optimization for feature selection. Binary versions of grey wolf optimization algorithm were proposed in [7], the authors transformed the continuous to its corresponding binary version; using transfer functions. The results show the outperformance of the proposed algorithms on various datasets. In [8] authors proposed chaotic Multi-verse optimization algorithm by incorporating five different chaotic maps for solving the feature selection problem. The algorithms were evaluated on five different datasets. The results indicated that the chaotic maps increased the performance of multi-verse optimization algorithm. In [9] authors improved the GWO algorithm by adding the concept of random walk and evaluated the performance of the algorithms on various benchmark functions. Quasi-opposition based strategy was incorporated in ant-lion optimization in [10] along with chaotic maps. In [11] authors proposed the use of Grasshopper Optimization with Evolutionary Population Dynamics based strategy for feature

selection. In [12] authors improved particle swarm optimization algorithm by proposing an autonomous group's based strategy. In [13] authors proposed an improved GWO algorithm in which a new position-updating equation is proposed by applying a random individual in the population. Binary versions of particle swarm optimization were proposed in [14] using the concept of S-shaped and V-Shaped transfer functions. In [15] authors enhanced the sine-cosine algorithm by adding an opposition-based strategy. In [16] authors proposed an improved version of the SCA, a modified position-updating equation and a new nonlinear conversion parameter decreasing strategy is used.

III. GREY WOLF OPTIMIZATION (GWO)

In recent years, Nature-inspired metaheuristic algorithms have gained popularity because of their capabilities of solving NP-hard and NP-complete problems. There working is totally inspired by the phenomena of nature. Metaheuristic algorithms majorly depend upon the intensification and diversification or exploration and exploitation [2]. The grey wolf optimization algorithm [17] is a metaheuristic algorithm inspired by the hunting behavior of Grey Wolfs. Grey wolfs live in a pack and the leader of the pack called the alpha Wolf. Alpha wolf is followed by the entire pack and then there come the Beta Wolves, they assist the Alpha Wolves in decision-making. At the third level Delta wolves work as a safeguard to protect the group also warned the pack in case of danger. Omegas are the last wolves following the instructions of senior wolves so alpha; beta and Delta are three best solutions and according to their position rest of all change their position. The algorithm can be mathematically modeled as: Firstly, grey wolves encircle the prey, shown in (1) and (2)

$$\vec{D} = |\vec{C} . \vec{X}_p(t) - \vec{X}(t)| \tag{1}$$

$$\vec{X}(t+1) = |\vec{X}_p(t) - \vec{A} . \vec{D}| \tag{2}$$

$$\vec{A} = 2 . \vec{a} . \vec{r}_1 - \vec{a} \tag{3}$$

$$\vec{C} = 2 . \vec{r}_2 \tag{4}$$

where t indicates the current iteration $\vec{A}$ and $\vec{C}$ are the coefficient vectors $\vec{X}_p$ is the prey position and $\vec{X}$ is the grey wolf position. The vector $\vec{A}$ and $\vec{C}$ can be calculated as shown in (3) and (4). Here the $\vec{a}$ is linearly decreased from 2 to 0 over the course of iteration and r_1 and r_2 are the random vectors in [0, 1]. Where $\vec{X}_\alpha$, $\vec{X}_\beta$ and $\vec{X}_\delta$ are the first three best solutions at given iteration t where A is defined in equation (3) and D can be calculated as follows

$$\vec{D}_\alpha = |\vec{C}_1 . \vec{X}_\alpha(t) - \vec{X}| \tag{5}$$

$$\vec{D}_\beta = |\vec{C}_2 . \vec{X}_\beta(t) - \vec{X}| \tag{6}$$

$$\vec{D}_\delta = |\vec{C}_3 . \vec{X}_\delta(t) - \vec{X}| \tag{7}$$

$$\vec{X}_1 = |\vec{X}_\alpha - \vec{A}_1 . \vec{D}_\alpha| \tag{8}$$

$$\vec{X}_2 = |\vec{X}_\beta - \vec{A}_2 . \vec{D}_\beta| \tag{9}$$

$$\vec{X}_3 = |\vec{X}_\delta - \vec{A}_3 . \vec{D}_\delta| \tag{10}$$

The Position updating equation can be seen in (11)

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{11}$$

$$\vec{a} = 2 - t \frac{2}{MaxIteration} \tag{12}$$

In every iteration $\vec{a}$ is updated from 2 to 0 according to (12).

1. Initialize random solutions(wolfs) in the Population X_i
2. Initialize the maximum numbers of iterations Max_iter, A,C and a;
3. Evaluate the fitness of each search agent X_i in the population and save the fitness of best three solutions(as alpha,beta and delta)
4. **While** end criterion is not met (t< Max_iter) do
   - Update the value of a (a decreases linearly from 2 to 0)
   - Generate randomly new values for r_2, r_3.
     For each search agent X_i in the population **do**
   Update the A and C for all the three wolfs.
   Compute the value of X1, X2 and X3 using eq. (8),(9),(10).
     • Update X_i using equation (11).
   End
   - Calculate the fitness of each search agent
   - Update the value of alpha, beta and delta.
   - Increment the iteration counter t=t+1
5. **End**

Algorithm I: Pseudo code of Grey wolf Optimization.

IV. OPPOSITION BASED ENHANCED GREY WOLF OPTIMIZATION (GWO)

This section proposes an enhanced version of GWO; the following strategies have been incorporated to enhance the performance of the GWO algorithm:

A. Opposition Based Learning

The first step optimization algorithm is to initialize a random set of solutions which may not be good enough to

Reach optima. To deal with this problem, in [15], [18], [19] authors proposed the use of both the randomly produced solutions and their opposite solutions. Considering the updated population and opposite of the updated population and considering best among them in every iteration, aids to get better the convergence rate. In one-dimensional space suppose 'a' is a real number between interval [lb, ub]. The opposite of a number Op_a is defined by (13)

$$Op\_a = (ub + lb) - a \tag{13}$$

$$Op\_a_i = (ub_i + lb_i) - a_i \tag{14}$$

The opposite of the point of 'a' is considered in (13) and

for n-dimensional space, the opposite of a point is shown in (14).

B. Modification of Exploration and Exploitation Controlling Parameter (a)

Parameter '*a*' in (4), controls the exploration and exploitation process. The basic parameter *'a'* is modified to as shown in (15) to get a better convergence.

$$a = 2 \times \exp onential(-(\frac{CurrentIteration}{MaxIteration})) \qquad (15)$$

C. Adding Importance Factor to Position Updating Equation

As seen in (11) the position is updated by taking the mean of $\vec{X}_1$, $\vec{X}_2$ and $\vec{X}_3$. This research modifies the position updating equation (11) by adding the importance factors IF_1, IF_2 and IF_3 based on the fitness as shown in (16).

$$\vec{X}(t+1) = IF_1 \times \vec{X}_1 + IF_2 \times \vec{X}_2 + IF_3 \times \vec{X}_3 \qquad (16)$$

The fittest Wolf is given the highest importance and so on. In this algorithm the value of IF_1 is considered as 0.7, the value of IF_2 considered as 0.2 and the value of IF_3 is taken 0.1.

D. Binary Version of OEGWO

To deal with 0's and 1's in a binary space. A way is proposed by using (16) for transforming the position of the agent from 0 to 1 or from 1 to 0 using transfer function [14]. In OEGWO, the position is updated is shown in (17) and (18)

$$Z(t) = IF_1 \times \vec{X}_1 + IF_2 \times \vec{X}_2 + IF_3 \times \vec{X}_3 \qquad (17)$$

$$\vec{X}(t+1) = sigmoid\ (z) = \frac{1}{1+e^{-Z(t)}} \qquad (18)$$

---

1. Initialize random solutions(wolfs) in the Population X_i
2. Initialize the maximum numbers of iterations Max_iter, A,C and *a*;
3. Evaluate the fitness of each search agent X_i in the population and save the fitness of best three solutions(as alpha,beta and delta)
4. **While** end criterion is not met (*t*< Max_iter) do
   - Select top best fit Search agents from X_i and Opposite of Xi(using (13) and (14))
   - Update the value of **a** using (15)
   - Generate randomly new values for r_2, r_3 .
   For each search agent X_i in the population **do**
   Update the A and C for all the three wolfs.
   Compute the value of $X1$, $X2$ and $X3$ using eq. (8),(9),(10).
     - Update X_i using equation (16).
     - For Binary conversion Update the value of X_i using equations (17) and (18)
       End
   - Increment the iteration counter *t=t+1*
5. **End**

Algorithm II: Pseudo code of OEGWO.

V. EXPERIMENTAL SETUP

For the experimental setup, a Core 2 Duo 2.00 GHz processor and 3.00 GB RAM are considered. For the analysis

of algorithms on various benchmark functions as shown in Table I and Table II, the population size is taken as 30 and number of iterations 300 and mean is taken from different 30 runs. For feature selection the results have been evaluated using 10 search agents, 100 iterations and 10 different runs.

A. Fitness Function

The fitness function is based on the K-NN classifier [20] used to evaluate the candidate solutions as shown in (16). $£rr(Dim)$ is the rate of error calculated from the k-NN classifier.

$$Fitness = IC_1 \frac{|FSub|}{N} + IC_2 * £rr(Dim) \qquad (19)$$

The selected feature subset size is represented by $|FSub|$ whereas the original size is $|N|$, IC_1 and IC_2 constants represent the importance of classification and feature subset size, where $IC_1 \in [0, 1]$ and $IC_2 = (1 - IC_1)$ adapted from [6].

To quantify the performance of proposed algorithms on various benchmark functions [17] two measures have been considered viz. Average fitness value(fitness value averaged on all the runs) and Standard Deviation value; measures the divergence of the solutions obtained in different runs .

B. Feature Extraction

For evaluation of feature selection in breast density classification problem, the authors have considered 322 images from mini-mias datasets[1, 21] and extracted forty-five texture features from Region of interests(ROI) extracted from 322 images as shown in Fig. 1.
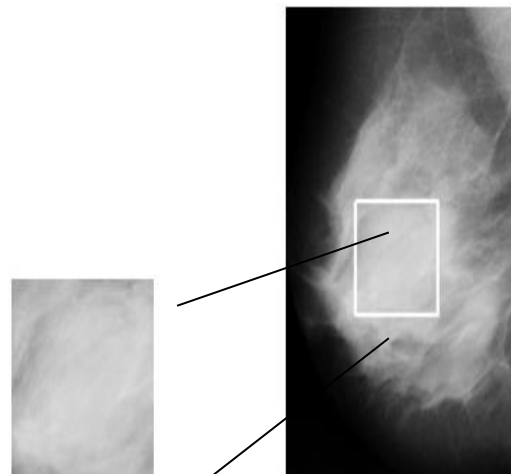

Fig. 1. Region of interest (ROI) extraction.

Texture features viz.

1) *Spatial Gray Level Co-occurence Matrix (SGLCM):* It Includes features[22] like Second Moment, Contrast, Correlation, Sum of Squares, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, Information Measure of Correlation1, Information Measure of Correlation2, Maximal Correlation Coefficient).

2) *Level Difference Statistics (GLDS)* [23]: it includes features like Homogeneity, Contrast, Mean, Energy, Entropy.

3) First-order Statistics (FoS) [24]: include Mean, Variance, Skewness, Kurtosis.

4) *Statistical Feature Matrix (SFM)* [25]: It includes Features like Coarseness, Contrast, Periodicity, Roughness.

5) *Law's Texture Energy Measures (TEM)* [26]: It includes Edge Level, Spot Level, Wave Level, Ripple Level, Spot Edge, Wave Edge, Ripple Edge, Wave Spot, Ripple Spot, Ripple Wave, Edge Edge, Spot Spot, Wave Wave, Ripple Ripple.

4) *Fractal* [27]: This includes features include Hurst Coefficient at Resolution1, Hurst Coefficient at Resolution2.

4) *Fourier Power Spectrum (FPS)* [28]: This includes features like Radial Sum, Angular Sum.

TABLE I: VARIOUS UNI-MODAL BENCHMARK FUNCTIONS

Func	Equation	lb	ub	dim
F1	$f(x) = \sum_{i=1}^{n} x_i^2$	-100	100	30
F2	$f(x) = \sum_{i=1}^{n} |x_i| + \prod_{i=1}^{n} |x_i|$	-10	10	30
F3	$f(x) = \sum_{i=1}^{n} (\sum_{j-1}^{i} x_i)^2$	-100	100	30
F4	$f(x) = \max_i \{|x_i|, 1 \leq i \leq n\}$	-100	100	30
F5	$f(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	-30	30	30
F6	$f(x) = \sum_{i=1}^{n} ([x_i + 0.5])^2$	-100	100	30
F7	$f(x) = \sum_{i=1}^{n} ix_i^4 + random[0, 1]$	-1.28	1.28	30

TABLE II: VARIOUS MULTI-MODAL BENCHMARK FUNCTIONS

Func.	Equation	lb	ub	dim
F8	$f(x) = \sum_{i=1}^{n} -x_i \sin(\sqrt{|x_i|})$	-500	500	30
F9	$f(x) = \sum_{i=1}^{n} [x_i^2 - 10\cos(2\pi x_i) + 10]$	-5.12	5.12	30
F10	$f(x) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right)$ $- \exp(\frac{1}{n}\sum_{i=1}^{n} \cos(2\pi x_i)) + 20 + e$	-32	32	30
F11	$f(x) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	-600	600	30
F12	$f(x) = \frac{\pi}{n}\{10\sin^2(\pi y_1) + \sum_{i=1}^{n-1}(y_i - 1)^2$ $[1+10\sin^2(\pi y_{i+1})] + (y_n-1)^2\} + \sum_{i=1}^{n} u(x_i, 10, 100, 4)$ $U(x_i, a, k, m) = \begin{cases} k(x_i - a)^m, & x_i > a \\ 0, & -a \leq x_i \leq a \\ k(-x_i - a)^m, & x_i < -a \end{cases}$	-50	50	30
F13	$f(x) = 0.1\{\sin^2(3\pi x_1) + \sum_{i=1}^{n}(x_i - 1^2)$ $[1+\sin^2(3\pi x_i + 1)] + (x_n - 1)^2[1$ $+\sin^2(2\pi x_n)]\} + \sum_{i=1}^{n} u(x_i, 5, 100, 4)$	-50	50	30

C. Evaluation Criteri a

To quantify the performance of the proposed algorithms, the following criteria are used [5]:

1) *Average Accuracy*: It is the number of samples correctly Classified from the total number of samples in all runs.

2) *Average fitness*: It is considered as the mean of all the fitness values obtained, by an algorithm in various runs.

3) *Worst fitness*: It gives the maximum of all fitness values found in all runs.

4) *Best fitness*: It gives the minimum of all fitness values found in all runs.

5) *Standard Deviation*: It gives the divergence of the finest solutions found after running the algorithm in all runs.

5) *Average Features subset size*: It is the total number of features selected averaged on all the runs.

6) *F-Measure*: It is an assessment of classifier's accuracy, from precision as well as the recall as a harmonic mean.

D. Parameter Settings

The basic parameter settings for the algorithms considered can be seen in Table III, where Ub and Lb for specific benchmark functions may be different [17].

TABLE III: PARAMETER SETTINGS

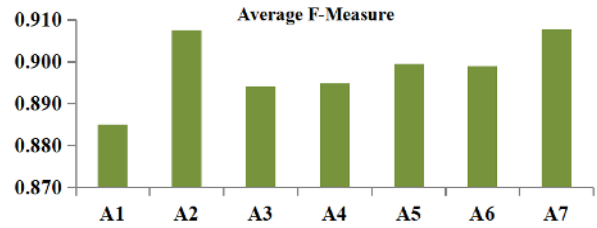| Parameter	Value
Search Agents	10 for FS 30 for Benchmark Functions
Runs	10 for FS30 for Benchmark Functions
Number of Iterations	100 for FS300 for Benchmark Functions
Lb(Lower Bound)	0 for FS
Ub(Upper Bound)	1 for FS
Dimensions	Total features
IC_2 (fitness function)	0.99
IC_1 (fitness function)	0.01

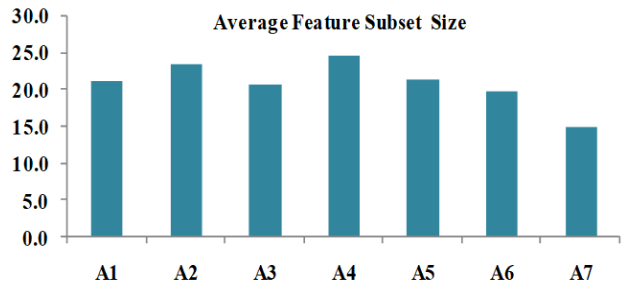Fig. 2. Analysis of f-measure breast density classification.

Fig. 3. Analysis of average feature subset size.

VI. RESULTS AND DISCUSSIONS

The results in Table IV give us a comparative study between the proposed algorithm and simple binary versions of GWO [17], Multi-verse optimization (MVO) [29], Sine cosine algorithm (SCA) [30], and Salp Swarm Algorithm (SSA) [31] and Particle Swarm Optimization (PSO) [32] on various benchmark functions [17]. The results indicate that OEGWO outperforms on seven benchmark functions and GWO outperforms on two benchmark functions PSO on three and MVO on one. Further on implementing the binary version of the proposed algorithm for breast density classification, clearly, the results indicate the outperformance of the proposed algorithm in Table V. in terms of Average classification accuracy value, Average fitness value, Worst fitness value, and Best fitness value. Also, when the binary version of the proposed algorithm is compared with the binary versions of the other algorithms in terms of Average Feature Subset Size (Number of Features Selected) and Average F-Measure value in the graphs shown in Fig. 2 and Fig. 3. The proposed algorithm outperforms the other algorithms. For average classification accuracy the Performance of the MVO and the proposed algorithm is the same but for average number of features selected the

proposed algorithm outperforms the other algorithms.

TABLE IV(1): RESULTS ON VARIOUS BENCHMARK FUNCTIONS

		PSO	SSA	GWO	OEGWO
F1	AvFitn	1.29E-02	5.90E-01	4.18E-15	**2.49E-34**
	Sdev	1.65E-02	4.85E-01	4.48E-15	7.90E-34
F2	AvFitn	5.86E-01	4.33E+00	1.98E-09	**4.90E-25**
	Sdev	4.27E-01	2.33E+00	1.02E-09	6.63E-25
F3	AvFitn	1.81E+02	3.73E+03	**7.01E-02**	1.01E-01
	Sdev	7.08E+01	2.70E+03	2.34E-01	3.21E-01
F4	AvFitn	1.82E+00	1.44E+01	9.06E-04	**1.90E-05**
	Sdev	6.33E-01	3.49E+00	6.55E-04	2.43E-05
F5	AvFitn	1.63E+02	8.73E+02	2.75E+01	**2.72E+01**
	Sdev	2.01E+02	1.14E+03	5.80E-01	7.85E-01
F6	AvFitn	**1.06E-02**	3.50E-01	9.68E-01	1.40E+00
	Sdev	1.33E-02	4.96E-01	4.66E-01	4.91E-01
F7	AvFitn	1.79E-01	2.68E-01	3.34E-03	**3.63E-04**
	Sdev	7.47E-02	1.04E-01	1.44E-03	2.68E-04
F8	AvFitn	-4.43E+03	**-7.25E+03**	-5.84E+03	-3.36E+03
	Sdev	1.17E+03	7.88E+02	1.13E+03	3.53E+02
F9	AvFitn	8.11E+01	5.46E+01	8.75E+00	**8.48E-01**
	Sdev	2.21E+01	1.84E+01	5.59E+00	4.65E+00
F10	AvFitn	1.12E+00	3.58E+00	1.39E-08	**9.41E-15**
	Sdev	6.87E-01	1.03E+00	7.82E-09	3.56E-15
F11	AvFitn	2.69E-02	3.84E-01	1.16E-02	**7.50E-13**
	Sdev	2.13E-02	1.94E-01	1.19E-02	4.11E-12
F12	AvFitn	7.76E-02	7.92E+00	**6.97E-02**	9.36E-02
	Sdev	1.51E-01	4.27E+00	4.07E-02	3.95E-02
F13	AvFitn	**6.27E-02**	3.74E+01	8.81E-01	1.24E+00
	Sdev	8.71E-02	1.36E+01	2.48E-01	2.09E-01

TABLE IV(2): RESULTS ON VARIOUS BENCHMARK FUNCTIONS

		SCA	MVO	GWO	OEGWO
F1	AvFitn	8.77E+01	3.38E+00	4.18E-15	**2.49E-34**
	Sdev	7.48E+01	1.10E+00	4.48E-15	7.90E-34
F2	AvFitn	2.97E-01	2.89E+01	1.98E-09	**4.90E-25**
	Sdev	3.88E-01	4.32E+01	1.02E-09	6.63E-25
F3	AvFitn	1.35E+04	7.81E+02	**7.01E-02**	1.01E-01
	Sdev	6.48E+03	4.01E+02	2.34E-01	3.21E-01
F4	AvFitn	4.75E+01	4.02E+00	9.06E-04	**1.90E-05**
	Sdev	1.15E+01	1.83E+00	6.55E-04	2.43E-05
F5	AvFitn	9.08E+05	6.56E+02	2.75E+01	**2.72E+01**
	Sdev	1.58E+06	7.99E+02	5.80E-01	7.85E-01
F6	AvFitn	1.13E+02	3.26E+00	**9.68E-01**	1.40E+00
	Sdev	1.09E+02	8.50E-01	4.66E-01	4.91E-01
F7	AvFitn	2.77E-01	5.59E-02	3.34E-03	**3.63E-04**
	Sdev	2.91E-01	1.88E-02	1.44E-03	2.68E-04
F8	AvFitn	-3.53E+03	**-7.68E+03**	-5.84E+03	-3.36E+03
	Sdev	2.11E+02	7.56E+02	1.13E+03	3.53E+02
F9	AvFitn	7.75E+01	7.75E+01	8.75E+00	**8.48E-01**
	Sdev	5.39E+01	5.39E+01	5.59E+00	4.65E+00
F10	AvFitn	1.66E+01	2.32E+00	1.39E-08	**9.41E-15**
	Sdev	6.74E+00	7.25E-01	7.82E-09	3.56E-15
F11	AvFitn	3.09E+00	1.02E+00	1.16E-02	**7.50E-13**
	Sdev	3.25E+00	2.71E-02	1.19E-02	4.11E-12
F12	AvFitn	9.89E+05	3.19E+00	**6.97E-02**	9.36E-02
	Sdev	2.07E+06	1.29E+00	4.07E-02	3.95E-02
F13	AvFitn	2.27E+06	**4.55E-01**	8.81E-01	1.24E+00
	Sdev	3.76E+06	3.23E-01	2.48E-01	2.09E-01

TABLE V: RESULTS ON VARIOUS CRITERIA'S FOR FEATURE SELECTION

	Algorithm considered	Avg Acc	Avg Fitn	Worst Fitn	Best Fitn	Std Dev
A1	SCA	0.847	0.156	0.183	0.127	0.018
A2	MVO	**0.875**	0.129	0.165	0.109	**0.018**
A3	PSO	0.863	0.140	0.183	0.097	0.026
A4	SSA	0.860	0.144	0.183	0.110	0.022
A5	GWO	0.866	0.138	0.165	0.114	0.018
A6	OEGWO	0.865	0.138	**0.164**	0.103	0.019
A7	S_OEGWO	**0.875**	**0.128**	0.181	**0.082**	0.031

Clearly, the results indicate the outperformance of the proposed algorithm in terms of various different parameters, viz. average fitness value and standard deviation value for

benchmark functions. For solving the feature selection problem the algorithm show outperformance on the basis of parameters like average classification accuracy, average fitness value and the average number of features selected. The improvement in the performance of the algorithm is due to various reasons viz. addition of opposition based strategy aids to improve better solutions from the set of random solutions and the opposite of the solutions. Improving the better solutions in every iteration rather than random solutions leads the algorithm to converge faster. The parameter 'a' is modified which controls the exploration and exploitation, for the position update part the proposed algorithm adds importance factors, in which it gives highest importance to the alpha wolf.

Similarly, for getting into the bottom of the feature selection problem, which is a binary problem. The solutions are represented in the form of strings of 0's and 1's. The conversion of the continuous version of the proposed algorithm to its corresponding binary version is an important task. For this the proposed algorithm incorporated the concept of transfer function viz. sigmoid in this research which is implemented at the position update step for the conversion of continuous version of the algorithm to its binary version.

VII. CONCLUSIONS AND FUTURE SCOPE

Breast Cancer is considered as the leading cause of death among women globally. An increase in the density of breast may lead to increase the chances of having breast cancer. A machine learning-based Computer-aided diagnosis system is used to assist the radiologist for the diagnosis of breast cancer or can be considered as secondary readers.

To increase the diagnosis accuracy of computer-aided diagnosis systems with a minimum number of features. This research proposes an Opposition based enhanced version of Grey Wolf Optimization (OEGWO) Algorithm particularly for solving the feature selection problem in breast density classification. The proposed OEGWO algorithm incorporates three different strategies viz. opposition-based population initialization, modification of exploration and exploitation controlling parameter and modification of position updating step. To conduct this research 322 mammogram images are considered and from each image, a Region of interest is extracted. From each region of interest a set of forty-five texture features is extracted and proposed algorithm is used to extract optimal set of features from the bigger set of features. From the experimental analysis, results indicate that the proposed continuous version of algorithm outperforms the other algorithms on various benchmark functions considered in this research; furthermore, the binary version of the proposed algorithm outperforms the other algorithms for solving feature selection problem in breast density classification. The improvement in the results is due to the fact that the opposition based strategy ads on the capability to start the algorithm from some of the good solutions rather than selecting the random solutions in every iteration. Similarly adding the importance factor gives the

priority to the better solutions and all the other solutions (omega wolfs) update their position while giving the highest priority to the alpha wolf, rather than giving equal priority to alpha, beta and delta wolf which was actually done in the normal grey wolf optimization algorithm.

In near future work can be further extended for solving feature selection problem in breast masses classification (other causes of breast cancer). Furthermore, in terms of metaheuristics, work can be further extended for reducing the computational time of the algorithm. To further improve the performance, some other strategies like quasi-oppositional learning should be considered for the performance enhancement of the algorithms in terms of average fitness value, average classification accuracy.

Moreover, the Grey Wolf Optimization algorithm can also be hybridized with some other algorithms. Instead of considering the opposition along with a random population, algorithms can be hybridized by considering the best solutions from two different algorithms and taking union of them to form a newly updated population for the next iterations. Another approach for improvement can be, by using some local search mechanisms to increase the exploitation and further performance should be validated on some other benchmark functions and for feature selection problems some standard datasets must be considered for performance validation. Apart from feature selection the binary version of the algorithm must be implemented for solving the other binary optimization problems from different problem domains.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

The author Harjot Kaur laid the foundation for the enhancement of Grey wolf optimization Algorithm for solving the feature selection problem and suggested various changes in the manuscript formation. The author Rahul Hans worked under the Guidance of Harjot Kaur and worked for the improvement of the Grey Wolf Optimization Algorithm for solving the feature selection problem in Breast Density Classification and gathered results after solving the problem.

ACKNOWLEDGMENT

REFERENCES

[1] V. Sharma, "CFS–SMO based classification of breast density using multiple texture models," *Medical & Biological Engineering & Computing*, vol. 52, no. 6, pp. 521-529, 2014.

[2] X. S. Yang, "Metaheuristic algorithms: optimal balance of intensification and diversification," *Applied Mathematics & Information Sciences*, vol. 8, no. 3, p. 977, 2014.

[3] D. H. Wolpert, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67-82, 1997.

[4] H. M. Zawbaa, "Feature selection approach based on moth-flame optimization algorithm," *Evolutionary Computation (CEC)*, pp. 4612-4617, July 2016.

[5] R. Hans and H. Kaur, "Binary multi-verse optimization (BMVO) approaches for feature selection," *International Journal of Interactive Multimedia and Artificial Intelligence*, July 2019.

[6] M. M. Mafarja, "Hybrid whale optimization algorithm with simulated annealing for feature selection," *Neurocomputing*, vol. 260, pp. 302-312, 2017.

[7] E. Emary, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371-381, 2016.

[8] A. A. Ewees, M. A. El Aziz, and A. E. Hassanien, "Chaotic multi-verse optimizer-based feature selection," *Neural Computing and Applications*, vol. 31, no. 4, pp. 991-1006, 2019.

[9] S. Gupta and K. Deep, "A novel random walk grey wolf optimizer," *Swarm* and *Evolutionary Computation*, vol. 44, pp. 101-112, 2019.

[10] S. Saha, "A novel quasi-oppositional chaotic antlion optimizer for global optimization," *Applied Intelligence*, vol. 48, no. 9, pp. 2628-2660, 2018.

[11] M. Mafarja, I. Aljarah *et al.*, "Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems," *Knowledge-Based Systems*, vol. 145, pp. 25-45, 2018.

[12] S. Mirjalili, "Autonomous particles groups for particle swarm optimization," *Arabian Journal for Science and Engineering*, vol. 39, no. 6, pp. 4683-4697, 2014.

[13] W. Long, "An exploration-enhanced grey wolf optimizer to solve high-dimensional numerical optimization," *Engineering Applications of Artificial Intelligence*, vol. 68, pp. 63-80, 2018.

[14] S. Mirjalil, "S-shaped versus V-shaped transfer functions for binary particle swarm optimization," *Swarm and Evolutionary Computation*, vol. 9, pp. 1-14, 2013.

[15] M. A. Elaziz, "An improved opposition-based sine cosine algorithm for global optimization," *Expert Systems with Applications*, vol. 90, pp. 484-500, 2017.

[16] W. Long, "Solving high-dimensional global optimization problems using an improved sine cosine algorithm," *Expert Systems with Applications*, vol. 123, pp. 108-126, 2019.

[17] S. Mirjalili, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46-61, 2014.

[18] H. R. Tizhoosh, "Opposition-based learning: A new scheme for machine intelligence," in *Proc. International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce*, vol. 1, pp. 695-701, November 2005.

[19] H. R. Tizhoosh, "Reinforcement learning based on actions and opposite actions," in *Proc. International Conference on Artificial Intelligence and Machine Learning*, December 2005, vol. 414.

[20] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175-185, 1992.

[21] J. Suckling, "The mammographic image analysis society digital mammogram database exerpta medica," *International Congress Series*, vol. 1069, pp. 375-378, 1994.

[22] R. M. Haralick, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, p. 610, 1973.

[23] J. S. Weszka, "A comparative study of texture measures for terrain classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 6, p. 269, 1976.

[24] I. N. Bankman, "Two-dimensional shape and texture quantification," *Handbook of Medical Image Processing and Analysis*, 2nd ed., pp. 270–272, 2009.

[25] C. M. Wu, "Statistical feature matrix for texture analysis," *Computer Vision Graphical Models and Image Processing*, vol. 54, no. 407, 1992.

[26] K. I. Laws, "Rapid texture identification," *SPIE Proceedings Image Processing for Missile Guidance*, San Diego, vol. 238, pp. 376–380, 1980.

[27] C. M. Wu, "Texture features for classification of ultrasonic liver images," *IEEE Transaction on Medical Imaging*, vol. 11, p. 141, 1992.

[28] G. O. Lendaris, "Diffraction pattern sampling for automatic pattern recognition," *Proceedings of IEEE*, February, vol. 58, pp. 198, 1970.

[29] S. Mirjalili, "Multi-verse optimizer: A nature-inspired algorithm for global optimization," *Neural Computing and Applications*, vol. 27, no. 2, pp. 495-513, 2016.

[30] S. Mirjalili, "SCA: A sine cosine algorithm for solving optimization problems," *Knowledge-Based Systems*, vol. 96, pp. 120-133, 2016.

[31] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili *et al.*, "Salp swarm algorithm: A bio-inspired optimizer for engineering design problems," *Advances in Engineering Software*, vol. 114, pp. 163-191, 2017.

[32] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. the IEEE International Conference on Neural Network*, Perth, Australia, 1995, pp. 1942–1948.

Rahul Hans has received his B.Tech degree in computer science and engineering from Punjab Technical University, Jalandhar and M.Tech. degree in computer science and engineering from Guru Nanak Dev University, Amritsar. He is currently pursuing the Ph.D. from Guru Nanak Dev University, Amritsar. His predominant research areas include mobile agent systems, machine learning and metaheuristics. He has published research papers in various reputed international conferences and journals.

Harjot Kaur is presently working as an assistant professor in the Department of Computer Science and Engineering at Guru Nanak Dev University, Regional Campus Gurdaspur, Punjab, India. She has completed her Ph.D. degree from the Department of Computer Science and Engineering, Guru Nanak Dev University, Amritsar, Punjab, India. Her predominant area of research is distributed artificial intelligence. She has published research papers in various reputed International Journals. Also she has attended various national and international conferences and presented many research papers.