

Wrapper Based Feature Selection for Disease Diagnosis using Optimization Algorithms

Ms. Keerthi Priya. N
Assistant Professor,
Department of CSE,
Kongu Engineering College,
Perundurai, Erode – 638060

Hema Latha. K, Hari Gokul Prasad. V,
Arun Kumar. T,
Department of CSE,
Kongu Engineering College, Perundurai,
Erode – 638060

Abstract—Data mining is the process of analyzing data from different perspectives and summarizing it into useful information that can be used to increase revenue, cuts costs, or both. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. Selection of relevant attributes for sample classification is a common task in most gene expression studies, where researchers try to identify the smallest possible set of features that can still achieve good predictive performance. Feature selection or data dimension refers to the process of identifying most important attributes which help to reduce the cost of prediction and also provide better classification accuracy due to finite sample size. Here the nature inspired algorithms like Particle Swarm Optimization (PSO) and Whale Optimization Algorithm(WOA) with Simulated Annealing(SA) algorithms are used. The performance of the proposed approaches is evaluated on benchmark datasets like hepatitis and breast cancer from UCI repository and compared with wrapper feature selection methods in the literature. The experimental results confirm that the efficiency of the proposed approaches in improving the classification accuracy varied depending on the datasets used. The ability of WOA algorithm in searching the feature space and selecting the most informative attributes for classification tasks was proved and also the simplicity of PSO in the exploration of the population was also proved.

Keywords- *Datamining, Metaheuristic, Feature selection, Optimization, Accuracy.*

I. INTRODUCTION

Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis. Feature selection is used to finds the minimum subset of features that are useful for the classification process. The data contains many features that are either redundant or irrelevant data[1]. It can be removed without incurring much loss of information. The selected subset of features could be evaluated using two different techniques: filter and wrapper. A wrapper-based approach includes a learning algorithm (e.g. classification algorithm) in the selection process [2]. In literature, there are many studies that conduct a comparison between many feature selection algorithms as in [3] where eight standard feature selection methods have been tested and evaluated using three different classifiers.

Searching an optimal subset is a crucial issue in feature selection. All possible subsets could be generated in an exhaustive manner to generate the best subset [4]. This approach is inapplicable for the huge datasets because if a dataset contains N features, then 2^N solutions should be generated and evaluated resulting in an extremely high computational cost [5]. Another approach is to search for a minimal reduct randomly [6]. The best

case in that approach is when finding the minimal reduct in early stages while the worst case of this approach is that it might perform a complete search. Furthermore, the heuristic, as the name suggests, uses heuristic information to guide the search. Although a heuristic search strategy does not guarantee finding the best subset, it normally finds an acceptable solution in a reasonable time [7]. Heuristic methods may be classified into two families: specific heuristics developed to solve a certain problem and general-purposed metaheuristics to solve a wide range of problems [7].

In the past two decades, metaheuristics have showed their efficiency and effectiveness in solving challenging and large-scale problems in engineering design, machine learning, data mining scheduling, and production problems. Nature-inspired algorithms are mostly metaheuristics inspired from nature [8]. There are three main sources of inspiration: evolutionary-based (e.g. evolutionary algorithms and artificial immune systems), swarm-based (e.g. ant colony, bees colony, and particle swarm optimization), and physics-based-based (e.g. simulated annealing [7]). Two contradictory criteria that are common in all these techniques are: diversification (exploration of the search space) and intensification (exploitation of the best solutions found)[7].As mentioned above, swarm-based algorithms mimic the collective behavior of natural creatures such as bat, grey wolf, ant lion, moth, etc. [9]. Recently, new evolutionary algorithms are proposed and shown a good performance when dealing with the. Feature Selection (FS) problem. For instance, Ant Lion Optimizer (ALO) was utilized to tackle this problem in [10] and [11] as a wrap-per feature selection model. In addition, a binary version of the ALO algorithm was presented in [12]. This approach has been enhanced by using a set of chaotic functions in adapting the single parameter that controls the balance between exploration and exploitation in the original algorithm [13]. Grey Wolf Optimizer (GWO) is a recent algorithm [14] that has been successfully employed for solving feature selection problems in [15,16].

In this paper Particle Swarm Optimization (PSO) and Whale Optimization Algorithm(WOA) are used. The results are being compared with the benchmark datasets to calculate the accuracy in predicting the disease. A stochastic swarm intelligence algorithm, known as Particle Swarm Optimization (PSO)proposed by Kennedy and Eberhart in 1995[17]. It have been applied to solve the energy minimization problem. It is simple, easy to implement, and requires only a small number of user-defined parameters, but it also suffers from premature convergence. Whale Optimization Algorithm (WOA) is a novel nature-inspired meta-heuristic optimization algorithm proposed by Seyedali Mirjalili and Andrew Lewis in 2016, which mimics the social behavior of humpback whales.

II. RELATED WORKS

In recent years, hybrid metaheuristics have been used by many researchers in the field of optimization [22]. Hybrid algorithms showed superior performance in solving many practical or academic problems [23]. In 2004, the first hybrid metaheuristic algorithm for feature selection was proposed [24]. In this approach, lo- cal search techniques were embedded into GA algorithm to control the search process. In feature selection domain, many hybrid metaheuristic algorithms have been proposed with much success. Mafarja and Abdullah proposed a filter feature selection hybrid algorithm that used SA to enhance the local search capability of genetic algorithm in [25]. The algorithm was tested on eight UCI datasets and showed a good performance in terms of the number of selected attributes when compared with the state-of-the-art approaches. In [26], a hybrid GA and SA has been proposed and tested on the hand-printed Farsi characters. Another wrapper feature selection hybrid algorithm was proposed in [27], by incorporating the metropolis acceptance criterion of simulated annealing into crossover operator of GA. Again, GA was hybridized with SA to produce a hybrid wrap- per feature selection algorithm to classify the power disturbance in the Power Quality (PQ) problem, and to optimize the SVM parameters for the same problem [28]. Olabiyisi et al. in 2012

[29] proposed a novel hybrid GA-SA metaheuristic algorithm for feature extraction in timetabling problem, where the GA selection process replaced by the one in SA to avoid stacking at the local optima. GA was hybridized with TS in a wrapper feature selection which used the FUZZY ARTMAP NN classifier as an evaluator [30]. Two filter feature selection memtic algorithms were proposed by Mafarja et al. [31]. In the scope of this work, this can be stated as: none of the heuristic wrapper feature selection is able to solve *all* feature selection problems. In other words, there is always room for the improvements of the current techniques to better solve the current of new feature selection problems. This motivated our attempts to propose yet another memetic algorithm for feature selection in the next section.

III. METHODS

In PSO, there are two vectors: position and velocity. The velocity vector is the main vector, which considers the best solution obtained so far by the particle and the entire swarm. With the velocity vector, the position vector is calculated for every solution. In WOA, however, there is only one vector to store the position of each solution in a search space. Also, the solutions are update during a spiral equation towards a promising solution or randomly in the search space. The WOA is more efficient compared to PSO in terms of memory since this algorithm only stores the best estimation of the global optimum in each iterations. However, PSO stores the best solution and personal best for all particles. WOA algorithm is a recent optimization algorithm that shows superior results in many optimization problems. The native algorithm uses a blind operator to play the role of exploitation regardless of the fitness value of the current solution and the operated one. We replaced this operator with a local search which considers a solution as its initial state, work on it, and replace the original solution by the enhanced one. This approach represents a hybridization between global search (WOA) and local search algorithm (SA).

A. Particle Swarm optimization

Particle swarm optimization (PSO) is a population based method that inspired from the

behavior (information exchange) of the birds in a swarm[18] . In PSO the population is called a swarm and the individuals are called particles. In the search space, each particle moves with a

Algorithm 1

Pseudo-code of the native PSO algorithm.

```

P = Particle_Initialization();
For i=1 to it_max
  For each particle p in P do
    fp = f(p);
    If fp is better than f(pBest)
      pBest = p;
  end
end
gBest = best p in P;
For each particle p in P do
  Update velocity of the particle
   $V_{t+1} = v_t + c1*rand*(pBest - p_t) + c2*rand*(gBest - p_t)$ ;
  Update position of the particle
   $P_{t+1} = p_t + V_{t+1}$ ;
end
end

```

velocity. The particle adapts this velocity due to the information exchange between it and other neighbors. At each iteration, the particle uses a memory in order to save its best position and the overall best particle positions. The best particle position is saved as a best local position, which was assigned to a neighborhood particles, while the overall best particle position is saved as a best global position, which was assigned to all particles in the swarm.

Each particle is represented by a D dimensional vectors,

$$x_i = (x_{i1}, x_{i2}, \dots, x_{iD}) \quad (1)$$

Each particle has a velocity, which is represented as,

$$v_i = (v_{i1}, v_{i2}, \dots, v_{iD}) \quad (2)$$

During the movement, each particle updates its position and velocity according to its own experience and that of its neighbors. The best previous position of the particle is recorded as the personal best *pbest*, and the best position obtained by the population thus far is called *gbest*. Based on *pbest* and *gbest*, PSO searches for the optimal solutions by updating the velocity and the position

of each particle according to the following equations:

$$x_i^{(t+1)} = x_i^{(t)} + v_i^{(t+1)} \quad (3)$$

$$v_i^{(t+1)} = v_i^{(t)} + c_1 r_{i1} * pbest_i^{(t)} - x_i^{(t)} + c_2 r_{i2} * gbest_i^{(t)} - x_i^{(t)} \quad (4)$$

where c_1, c_2 are two acceleration constants called cognitive and social parameters r_1, r_2 are random vector $\in [0, 1]$.

B. Whale Optimization Algorithm

Algorithm 2

Pseudo-code of the native WOA algorithm.

```

Generate Initial Population  $X$  ( $i = 1, 2, \dots, n$ )
Calculate the fitness of each solution
 $X^* =$  the best search agent
while ( $t < \text{Max\_Iteration}$ )
    for each solution
        Update  $a, A, C, l$ , and  $p$ 
        if 1 ( $p < 0.5$ )
            if 2 ( $|A| < 1$ )
                Update the position of the current solution
                by Eq. (6)
            else if 2 ( $|A| > 1$ )
                Select a random search agent ( $i$ )
                Update the position of the current search
                agent by the Eq. (11)
            end if 2
        else if 1 ( $p \geq 0.5$ )
            Update the position of the current search
            by the Eq. (9)
        end if 1
    end for
    Check if any solution goes beyond the search space and
    amend it
    Calculate the fitness of each solution
    Update  $X^*$  if there is a better solution  $t = t + 1$ 
end while
return  $X^*$ 

```

WOA belongs to the family of stochastic population-based algorithms proposed by Mirjalili and Lewis in 2016 [19]. It mimics the bubble-net feeding in the foraging behavior of the humpback whales [19]. The humpback whales hunt close to the surface with trapping the prey in a net of bubbles. They create this net when swimming in a '6'-shaped path.

The algorithm mimics two phases: the first phase (exploitation phase) is encircling a prey and

spiral bubble-net attacking method, and the second phase (exploration phase) is searching randomly for a prey. The details of each phase are presented in the following subsections.

Exploitation phase

To update a solution, Eqs. (5) and (6) are used, which mathematically model the movement of a whale around a prey

$$\vec{D} = |\vec{C} \cdot \vec{X}^* - \vec{X}(t)| \quad (5)$$

$$\vec{X}(t+1) = \vec{X}^* - \vec{A} \vec{D} \quad (6)$$

$$\vec{A} = 2\vec{a}\vec{r} - \vec{a} \quad (7)$$

$$\vec{C} = 2\vec{r} \quad (8)$$

where t represents the current iteration, X^* represents the best solution obtained so far, X is the current solution, $||$ is the absolute value. A and C are coefficient vectors that are calculated as in Eqs. Then a spiral equation is created between the current solution and the best (leading) solution as in Eq. (9).

$$\vec{X}(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^* \quad (9)$$

Where \vec{D} is the distance between a whale X and a prey ($\vec{D} = |\vec{X}^* - \vec{X}(t)|$) b defines the spiral's shape of the spiral, and l is a random number in $[-1, 1]$.

Exploration phase

In order to enhance the exploration in WOA, instead of requiring the solutions to search randomly based on the position of the best solution found so far, a randomly chosen solution is used to update the position accordingly. So, a vector A with the random values greater than 1 or less than -1 is used to force a solution to move far away from the best known search agent. This mechanism can be mathematically modeled as in Eqs. (10) and (11).

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}| \quad (10)$$

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A} \vec{D} \quad (11)$$

Where \vec{X}_{rand} is a random whale chosen from the current population

Algorithm 2 shows the pseudo code of WOA algorithm. It may be seen that WOA creates a random, initial population and evaluates it using a fitness function once the optimization process starts.

After finding the best solution, the algorithm repeatedly executes the following steps until the satisfaction of an end criterion. Firstly, the main coefficients are updated. Secondly, a random value is generated. Based on this random value, the algorithm updates the position of a solution using either Eqs. (6)/(11) or Eq.(9). Thirdly, the solutions are prevented from going outside the search landscape. Finally, the algorithm returns the best solution obtained as an approximation of the global optimum.

C. Simulated Annealing

Algorithm 3

Pseudo-code for SA algorithm.

```

Input:  $iteration_{max}$ 
Output:  $S_{best}$ 
 $S_{current} \leftarrow create\_initial\_solution()$ 
 $S_{best} \leftarrow S_{current}$ 
For ( $i=1$  to  $iteration_{max}$ )
     $S_i \leftarrow create\_neighbor\_solution(S_{current})$ 
     $Temp_{curr} \leftarrow calculate\_temperature(i)$ 
    If  $f(S_i) \geq f(S_{current})$ 
         $S_{current} \leftarrow S_i$ 
        If  $f(S_i) \geq f(S_{best})$ 
             $S_{best} \leftarrow S_i$ 
        End
    Elseif  $exp(\frac{f(S_{current}) - f(S_i)}{temp_{current}}) > rand()$ 
         $S_{current} \leftarrow S_i$ 
    End
End
Return ( $S_{best}$ )

```

Simulated annealing, proposed by Kirkpatrick et al. [20], is a single-solution metaheuristic algorithm based on the hill climbing method that. To overcome the problem of stagnating in local optima, SA utilizes a certain probability to accept a worse solution.

Algorithm 3 shows the pseudo code of the SA heuristic. In each step, SA considers some neighboring state S_i of the current state $S_{current}$, and decides between moving to state S_i or staying in state $S_{current}$ with some probability. The new state (S_i) will be accepted if it has a better fitness compared to the current state ($S_{current}$). If, however,

the new state has lower fitness, it will be accepted with the probability showed.

IV. THE PROPOSED WORK

PSO is similar to a genetic algorithm (GA) in that the system is initialized with a population of random velocity, and the potential solutions, called *particles*, are then “flowed” through the problem solutions. It is unlike a GA, however, in that each potential solution is also assigned a randomized space. Each particle keeps track of its coordinates in the problem space which are associated with the best solution (fitness) it has achieved so far. (The fitness value is also stored) This value is called *pbest*. Another “best” value that is tracked by the global version of the particle swarm optimizer is the overall best value, and its location, obtained so far by any particle in the population. This location is called *gbest*. The particle swarm optimization concept consists of, at each time step, changing the velocity (accelerating) each particle toward its *pbest* and *gbest* locations (global version of PSO). Acceleration is weighted by a random term, with separate random numbers being generated for acceleration toward *pbest* and *gbest* locations.

In WOA, each feature subset can be seen as a position of a whale. Each subset may contain N features, where N is the number of features in the original set. The less the number of features in the solution and the higher the classification accuracy, the better is the solution. The algorithm starts with a set of randomly generated solutions (subsets) called population. Each solution is then evaluated using the proposed fitness function. The fittest solution in the population is marked as X^* (prey). The main loop in WOA is iterated a number of times. In each iteration, solutions update their positions according to each other to mimic the bubble-net attacking and searching for prey methods. To mimic the bubble-net attacking, a probability of 50% is assumed to choose between the shrinking encircling mechanism (Eq. (6)) and the spiral model (Eq. (11)) to update the position of the solutions (whales). When the shrinking encircling mechanism is employed, a balance between exploration and exploitation is required. A random vector A which contains values >1 or <1 is used for this purpose. If $A > 1$ then the

exploration (searching for prey methods) is employed by searching in the neighborhood of a randomly selected solution, while the neighborhood of best solution so far is exploited when $A < 1$. This process is repeated until satisfying the stopping criteria, which is usually the maximum number of iterations.

WOA algorithm is a recent optimization algorithm that shows superior results in many optimization problems. The native algorithm uses a blind operator to play the role of exploitation regardless of the fitness value of the current solution and the operated one. We replaced this operator with a local search which considers a solution as its initial state, work on it, and replace the original solution by the enhanced one. This approach represents a hybridization between global search (WOA) and local search algorithm (SA). In the proposed approach, two hybridization models between the two algorithms are considered, (1) Low-Level Team-work Hybrid (LTH) and (2) High-Level Relay Hybrid (HRH). On one hand, in LTH, SA algorithm is embedded in WOA algorithm to search for the best solution in the neighbour of both the randomly selected solution (to replace Eq. (11)) and the neighbour of the best known solution (to replace Eq. (6)) and replace the original one. This process improves the exploitation ability of WOA algorithm. SA algorithm in this approach acts as an operator in WOA algorithm. On the other hand, HRH model uses SA algorithm after applying WOA algorithm and finding the best solution, then SA is used to enhance that final solution. In both process, WOA uses random selection mechanism to select the random solution that enables the algorithm to explore the feature space.

V. EXPERIMENTS

The implementation of the proposed algorithm is done using Matlab. MATLAB (MATrix LABoratory) is a multi-paradigm numerical computing environment and fourth-generation programming language. Developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in

other languages, including C, C++, Java, Fortran and Python.

Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing capabilities. An additional package, Simulink, adds graphical multi-domain simulation and Model-Based Design for dynamic and embedded systems. MATLAB is a high-level language and interactive environment for numerical computation, visualization, and programming. Using MATLAB, user can analyze data, develop algorithms, and create models and applications. The language, tools, and built-in math functions enable to explore multiple approaches and reach a solution faster than with spreadsheets or traditional programming languages, such as C/C++ or Java. MATLAB can be used for a range of applications, including signal processing and communications image and video processing control systems test and measurement, computational finance, and computational biology. More than a million engineers and scientists in industry and academia use MATLAB, the language of technical computing.

A. Datasets

In order to assess the performance of the proposed algorithms, the experiments are performed on Feature Selection (FS) benchmark datasets (Table 1) from the UCI data repository [21]

Table 1: List of datasets used in the experiments

Datasets	No of attributes	No of objects
Breastcancer	9	699
Hepatitis	20	155

B. Results and discussion

All results obtained from the proposed approaches are reported in this section. The native WOA are compared to assess the effect of hybridizing SA algorithm. Then PSO algorithm is being compared with WOA-SA algorithm. To find out the best approach among the proposed approaches, all approaches are compared together in one table.

The proposed approaches are compared based on the following criteria:

1) Classification accuracy by using the Selected features on the test dataset.

2) The average selection size is the second comparison that is presented in this section.

3) Then the fitness values obtained from each approach is reported.

Table 2: Comparison of results between the proposed approaches (average accuracy for 100 iterations)

Algorithms	Accuracy for breast cancer dataset	Accuracy for hepatitis dataset
PSO	97.57	98.65
WOA-SA	98.00	75.64

Table 3: Comparison of results between the proposed approaches (average accuracy for 500 iterations)

Algorithms	Accuracy for breast cancer dataset	Accuracy for hepatitis dataset
PSO	97.48	98.65
WOA-SA	97.70	75.64

Table 4: Comparison of results between the proposed approaches (average accuracy for 1000 iterations)

Algorithms	Accuracy for breast cancer dataset	Accuracy for hepatitis dataset
PSO	96.57	98.577
WOA-SA	97.42	71.79

Table 5: Comparison of results between the proposed approaches (average accuracy for 1500 iteration)

Algorithms	Accuracy for breast cancer dataset	Accuracy for hepatitis dataset
PSO	97.55	98.61
WOA-SA	96.57	75.79

C. Comparison of proposed approaches

This section compares the two proposed approaches to analyze the impact of enhancing the exploration, exploitation and both of them combined. By analyzing the results in the Tables and Figures. The performance of the proposed approaches shows better results on varying datasets.

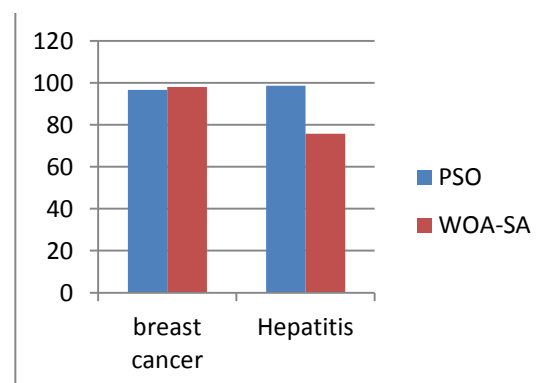


Figure 1: Average accuracy plotted for PSO and WOA-SA algorithm for 100 iterations

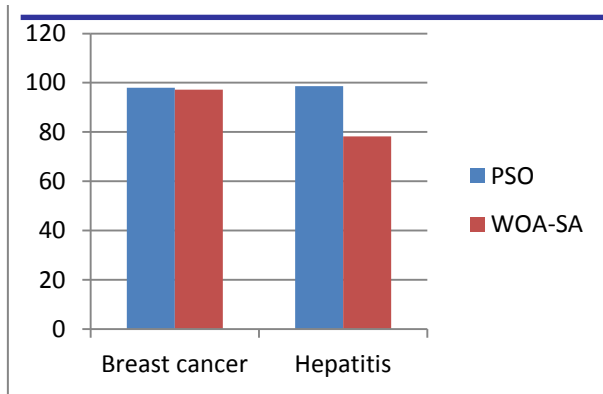


Figure 2: Average accuracy plotted for PSO and WOA-SA algorithm for 500 iterations

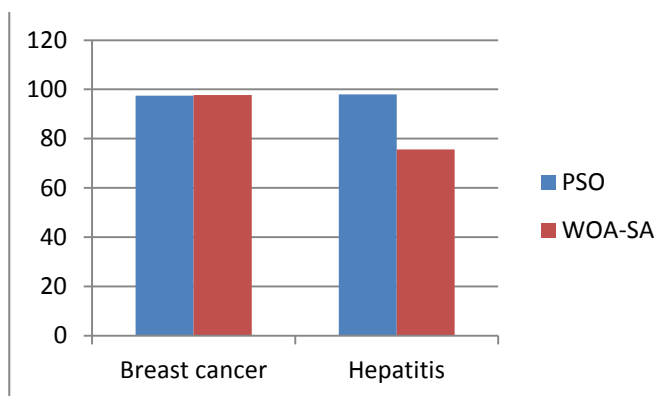


Figure 3: Average accuracy plotted for PSO and WOA-SA algorithm for 1000 iterations

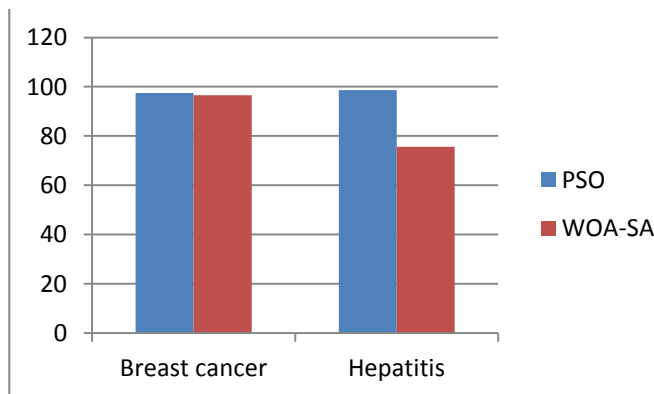


Figure 4: Average accuracy plotted for PSO and WOA-SA algorithm for 1500 iteration

VI. CONCLUSION

Feature selection is one of the key factors in enhancing the classifier abilities in the classification

problem. Literature review was carried out for PSO (Particle Swarm Optimization), SA (Simulated Annealing) and WOA (Whale Optimization Algorithm). It was concluded that the PSO algorithm is used in many engineering applications for optimizing the various design variables and it gives better performance than many other optimization algorithms. WOA mimicked the behavior of the humpback whales. Standard benchmark datasets were considered for the present work. The results of PSO and WOA-SA were compared in the previous section. Future work can be extended by considering more data sets with yet more metaheuristics algorithms that are recently proposed.

References

- [1] H. Liu, H. Motoda, Feature Selection for Knowledge Discovery and Data Mining, Kluwer Academic Publishers, Boston, 1998.
- [2] R. Kohavi, G.H. John, Wrappers for feature subset selection, *Artif. Intell.* 97 (1)(1997) 273–324.
- [3] H. Abusamra, A comparative study of feature selection and classification methods for gene expression data of glioma, *Procedia Comput. Sci.* 23 (2013) 5–14.
- [4] N. Zhong, J. Dong, S. Ohsuga, Using rough sets with heuristics for feature selection, *J. Intell. Inf. Syst.* 16 (3) (2001) 199–214.
- [5] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, *J. Mach. Learn. Res.* 3 (March) (2003) 1157–1182
- [6] C. Lai, M.J. Reinders, L. Wessels, Random subspace method for multivariate feature selection, *Pattern Recognit. Lett.* 27 (10) (2006) 1067–1076.
- [7] E.G. Talbi, *Metaheuristics From Design to Implementation*, Wiley Online Library, 2009.
- [8] I. Fister Jr., et al., A Brief Review of Nature-inspired Algorithms for Optimization, *arXiv preprint arXiv:1307.4186*, 2013.

- [9] F. Valdez, in: J. Kacprzyk, W. Pedrycz (Eds.), *Bio-Inspired Optimization Methods*, in Springer Handbook of Computational Intelligence, Springer Berlin Heidelberg, Berlin, Heidelberg, 2015, pp. 1533–1538.
- [10] S. Mirjalili, The ant lion optimizer, *Adv. Eng. Softw.* 83 (2015) 80–98.
- [11] H.M. Zawbaa, E. Emary, B. Parv, Feature selection based on antlion optimization algorithm, 2015 Third World Conference on Complex Systems (WCCS) (2015) 1–7.
- [12] E. Emary, H.M. Zawbaa, A.E. Hassanien, Binary ant lion approaches for feature selection, *Neurocomputing* (2016).
- [13] H.M. Zawbaa, E. Emary, C. Grosan, Feature selection via chaotic antlion optimization, *PLoS One* 11 (3) (2016) e0150652.
- [14] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, *Adv. Eng. Softw.* 69(2014) 46–61.
- [15] E. Emary, H.M. Zawbaa, A.E. Hassanien, Binary grey wolf optimization approaches for feature selection, *Neurocomputing* 172 (2016) 371–381.
- [16] E. Emary, et al., Feature subset selection approach by gray-wolf optimization, in: *Afro-European Conference for Industrial Advancement*, Springer, 2015.
- [17] Bansal JC, Shashi, Deep K, Katiyar VK. Minimization of molecular potential energy function using particle swarm optimization. *Int J Appl math mech* 2010;6(9):1-9
- [18] Goldberg DE. *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley; 1989.
- [19] S. Mirjalili, A. Lewis, The whale optimization algorithm, *Adv. Eng. Softw.* 95(2016) 51–67.
- [20] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, *Science* 220 (4598) (1983) 671–680.
- [21] Blake, C.L. and C.J. Merz. UCI Repository of machine learning databases. 1998 [cited 2016 1 June]; Available from: <http://www.ics.uci.edu/~mllearn/>.
- [22] E.G. Talbi, *Metaheuristics From Design to Implementation*, Wiley Online Library, 2009.
- [23] E.-G. Talbi, A taxonomy of hybrid metaheuristics, *J. Heuristics* 8 (5) (2002) 541–564.
- [24] I.-S. Oh, J.-S. Lee, B.-R. Moon, Hybrid genetic algorithms for feature selection, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (11) (2004) 1424–1437.
- [25] M. Mafarja, S. Abdullah, Investigating memetic algorithm in solving rough set attribute reduction, *Int. J. Comput. Appl. Technol.* 48 (3) (2013) 195–202.
- [26] R. Azmi, et al., A hybrid GA and SA algorithms for feature selection in recognition of hand-printed Farsi characters, in: *Proceeding of the 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems*, 2010.
- [27] J. Wu, Z. Lu, A novel hybrid genetic algorithm and simulated annealing for feature selection and kernel optimization in support vector regression, in: *Proceeding of the 2012 IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI)*, IEEE, 2012.
- [28] K. Manimala, K. Selvi, R. Ahila, Hybrid soft computing techniques for feature selection and parameter optimization in power quality data mining, *Appl. Soft Comput.* 11 (8) (2011) 5485–5497.
- [29] O. Olabiyisi Stephen, et al., Hybrid metaheuristic feature extraction technique for solving timetabling problem, *Int. J. Sci. Eng. Res.* 3 (8) (2012).
- [30] W.C. Tang, Feature Selection For The Fuzzy Artmap Neural Network Using A Hybrid Genetic Algorithm And Tabu Search, USM, 2007.

- [31] M. Majdi , S. Abdullah , N.S. Jaddi , Fuzzy Population-based meta-heuristic approaches for attribute reduction in rough set theory, World Acad. Sci. Eng. Technol. Int. J. Comput. Electr. Autom. Control Inf. Eng. 9 (12) (2015) 2289–2297 .