

A NEW APPROACH FOR FEATURES SELECTION BASED ON BINARY SALP SWARM ALGORITHM

¹MUNTADHER KHAMEES, ²ABBAS ALBAKRY, ³KHALID SHAKER

¹Department of Computer Science, College of Sciences, Diyala University, Diyala, Iraq

²University of Information Technology & Communications, Baghdad, Iraq

³Department of Computer Science, College of Computer Sciences & Information Technology, University Of Anbar, Anbar, Iraq

E-mail: ¹alkarawis@gmail.com, ²abbasm.albakry@uoitc.edu.iq, ³khalidalhity@gmail.com

ABSTRACT

Metaheuristic techniques become considerably popular in solving feature selection (FS) problems due to their flexibility and ability to avoid the local optimum problem. Features selection is important and essential mean to tackle the classification problems through choosing an optimal features subset according to a certain criterion. FS is used to reduce dimensionality and remove noise from data, these are given rise to speed of learning, simplicity of rules, visualizes the data and predictive accuracy. Salp Swarm Algorithm (SSA) is a new metaheuristic algorithm that emulates the inbred behaviour of the Salp chain. In this study, a new FS approach applies the native SSA in machine learning domain to select the optimal feature group on the basis of wrapper mode. Subsequently, SSA is hybridised with a mutation operator. Mutation is embedded to act as an internal operator and consequently maintain diversity and improve the exploration ability within the SSA. The performance of SSA with mutation operator (SSAMUT) on 16 datasets from UCI Machine Learning repository is evaluated and compared with that of the native SSA and other related approaches in the literature. Experimental results proved the efficiency of the proposed approaches in solving search space problems. SSAMUT presents the most excellent performance compared with those of other approaches on all datasets.

Keywords: *Metaheuristics, Salp Swarm Algorithm, Feature Selection, Wrapper Model, Mutation Operator*

1. INTRODUCTION

Knowledge Discovery in Databases (KDD) extracts useful information from a large amount of data. Data mining is an essential part in KDD. This database consists of four major types of processes, namely, data warehousing, pre-processing, data mining and evaluation process [1]. Dimensionality reduction techniques in data mining can be applied in high-dimensional data to improve the predictive accuracy, learning speed and simplicity of rules; the redundant and irrelevant features in the original data are eliminated to generate a subset of attributes with improved discriminant power [2], [3].

Features are also called properties, attributes, dimensions or characteristics. The advanced data processing technology and the growing demand of taking advantages of data stored form a new

challenge for data mining, one of the solutions to create a predictive feature selection model [4]. The procedure selects the optimal feature subset from a large dataset according to a certain criterion called feature selection (FS) [4]. Selecting the most suitable subset of features (M) from the large dataset (N) minimises the size of the search space problem. This step makes learning algorithm remarkably easy and learning high-dimensional data fast [4]. FS technique can be used in many applications, such as machine learning [5], image processing [6], [7], signal processing [8], pattern recognition [9] and data mining [10].

FS presents two models: wrapper and filter models [11]. The wrapper approach establishes a classifier to obtain the highest accuracy of prediction and subsequently selects the optimal features. The filter approach is based on separate

measures: classic and consistency measures. Filter approach ranks all attributes in the original dataset and sorts them, but it provides no number regarding the most remarkable features. A filter approach is fast. This approach can also handle large dataset and evaluate the features on the basis of the data. We can hybridise between filter and wrapper approaches through reducing the search dimension areas by using the filtration technique and the wrapping approach [4], [12].

FS is considered a search space problem. Search strategies can be summarised as follows: complete search, heuristic search and nondeterministic search. Complete search is time consuming because it covers all feature combinations to determine the optimal solution. By contrast, the heuristic strategy is much faster than the complete strategy; this strategy only searches a particular path to determine the near optimal solution. However, heuristic search is commonly time consuming [4].

Metaheuristic-based search and optimisation are applied in computer science to determine the near-optimum solution [13]. Meta and heuristic are both Greek words: meta refers to high perspective, and heuristic indicates the art of discovering modern techniques. Metaheuristic design shows two contradictory criteria, namely, exploration (diversification) of the search space and exploitation (intensification), to determine the optimal solutions. Single-based solution algorithms are diversified, whereas population-based algorithms are intensified. Each metaheuristic algorithm shares many search techniques [14].

A binary antlion optimisation algorithm (ALO) adjusts the continuous antlion to deal with FS as a binary problem [15]. ALO is utilised to search the feature space for optimal feature combination adaptively, thereby maximising the classification performance. Furthermore, an extended version of particle swarm optimisation (PSO) algorithm is used to deal with binary FS problems [16]. Many other studies also applied bioinspired techniques [17], [18], [19], [20], [21], [22].

Recently, hybrid metaheuristics algorithm showed high-performance in solving data mining and machine learning problems [23], [24], [25], [26], [27], [28], [29].

The first hybrid approach proposed in November 2004, Oh and Lee introduced a hybrid algorithm between local search method were embedded within genetic algorithm to enhance the search space by searching the most promising regions located by GA algorithm [30].

In 2010, simulated annealing (SA) has been hybridised with genetic algorithm (GA), the hybrid approached was tested on eight datasets from UCI Machine Learning repository [31]. It showed a better performance when compared with other search algorithms. Many hybrid metaheuristics techniques have been used in feature selection domain with much success. [32] Proposed a new wrapper-based approach for features selection by hybrid SA into crossover operator to classify the power disturbance. Moreover, in [33], proposed genetic algorithm was hybridised with PSO algorithm which used the SVM classifier as an evaluator called GPSO. This approach applied as microarray data classification method. For recent reviews about metaheuristic and feature selection techniques we refer to [34], [35], [36], [37], [38].

In the domain of this work, there is no metaheuristic-based approach is able to solve all problems of features selection. Nevertheless, improvements can be made to current approaches to get better performance. This motivation underlies most our attempts to create a predictive model based on hybrid algorithm for selecting optimal features from original dataset.

This paper is organised as follows; Section 1 presents the introduction. Section 2 describes the continuous salp swarm algorithm (CSSA). In Section 3, the SSA with mutation operator (SSAMUT) is applied for FS. Section 4 discusses the datasets, parameters and test results. Section 5 presents the conclusion and main findings of this study and suggests work for future studies.

2. PRELIMINARIE

2.1. CSSA

Salps are marine creatures belonging to the Salpidae family, and they exhibit a barrel-shaped transparent body. Their texture is considerably similar to that of a jellyfish. The salp pumps water through its body to propel itself forward. In deep seas, Salps often form into a chain-shaped group, which is called a salp chain. This behaviour achieves the most excellent movement by using rapid harmonised changes and foraging. The salp shape and salp chain are shown in Figure 1(a) and 1(b), respectively [39], [40].

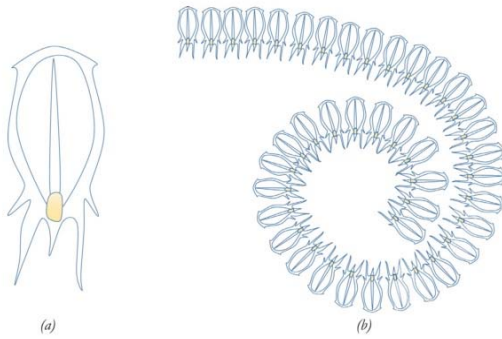


Figure 2. 1. (a) Individual salp. (b) Salp chain.

In 2017, [41] suggested a swarm model based on salp chains to solve optimisation problem. Firstly, the population is divided into two classes: leader Salps are at the forefront, and followers are the remaining Salps of the chain. The leader starts the swarm, and the followers trace back one another. The salp position is defined in a 2D matrix called x . F is the target food source in the search space. Only the leader position updates with regard to the food source, as shown in Eq. (2.1):

$$x_j^1 = \begin{cases} F_j + c_1((ub_j - lb_j)c_2 + lb_j) & c_3 \geq 0 \\ F_j - c_1((ub_j - lb_j)c_2 + lb_j) & c_3 < 0 \end{cases} \quad (2.1)$$

where x_j^1 shows the first leader position in the j th dimension, F_j is the position of the food source, ub_j and lb_j indicate the upper and lower bounds of the dimension, respectively, and c_1 is a random number parameter used to balance between diversification and intensification as follows:

$$c_1 = 2e^{-\left(\frac{4l}{L}\right)^2} \quad (2.2)$$

where parameter l indicates the current iteration, and L is the ultimate number of iterations. c_2 and c_3 are random values between [0,1]. The next place of Salp chain is towards either positive or negative infinity. The following equation based on Newton's law of motion is used to update the position of follower Salps:

$$x_j^i = \frac{1}{2}at^2 + v_0t \quad (2.3)$$

where $i \geq 2$, x_j^i is the position of the follower salps in j th dimension, v_0 is the initial speed value, t is time, where a is calculated as follows

$$a = \frac{v_{final}}{v_0} \text{ where } v = \frac{x-x_0}{t} \quad (2.4)$$

$a = v$ final divided by v_0 , where $v = x-x_0$ divided by time value. In optimisation, the time is iteration, and the discrepancy value is equal to 1, where $v_0 = 0$:

$$x_j^i = \frac{1}{2}(x_j^i + x_j^{i-1}) \quad (2.5)$$

where $i \geq 2$ and x_j^i indicate the position of i th follower Salp in j - dimension.

```

Initialize the Salp population  $X_i (i= 1, 2, \dots, n)$ 
considering  $Ub$  and  $Lb$ 
While (end condition is not satisfied)
    Calculate the fitness of each search agent (Salp)
     $F$ = the best search agent
    Update  $c_1$  by Eq. (2.2)
    For each Salp ( $x_i$ )
        If ( $i=1$ )
            Update the position of the leading
            Salp by Eq. (2.1)
        else
            Update the position of the follower
            Salp by Eq. (2.5)
        end
    end
    Amend the Salp beads on the upper and lower
    bounds of variables
end
Return  $F$ 
    
```

Figure2. 2. SSA pseudocode.

Figure2.2 shows the SSA pseudocode. The algorithm starts initiating Salps with random positions to approximate the global optimum, subsequently computes the fitness value for each individual of initial population and assigns the most remarkable position to variable F as the food source. Equations (2.1) and (2.5) can be used to simulate the Salp chains [41].

3. PROPOSED BINARY SSA FOR FS APPROACH

The SSA is used in feature subset selection to classify problems on the basis of wrapper-based mode. The wrapper-based methodology applies the classification approach as an evidence in feature subset selection technique based on some optimising algorithms. In a binary search problem,

the salp population is considered a binary bound of dimension. When SSA is used as a feature subset selection technique, a binary version of SSA should be sophisticated. In our approach, we assumed that each solution of the population is represented in one-dimensional vector, and the vector length is equal to the dimension number of the dataset. Each cell inside the vector is symbolised by one or zero: 1 indicates the selection of a corresponding attribute, and 0 indicates the neglect of a corresponding attribute.

```

Initialize the Salp population  $X_i (i= 1, 2, \dots n)$ 
considering to the binary bound of dimension when
upper=1 & lower=0
While (end condition is not satisfied)
    Calculate the fitness of each search agent (Salp)
    based on error rate
     $F$  = the best search agent
    Update  $c_r$  by Eq. (2.2)
    For each Salp ( $x_i$ )
        If ( $i=1$ )
            Update the position of the leading
            Salp by Eq. (2.1)
            Mutate the position of the leading salp
            with probability  $p_m$  by Eq. (3.2)
        else
            Update the position of the follower
            Salp by Eq. (2.5)
        end
    end
    Amend the Salp beads on the upper and lower
    bounds of variables
end
Return  $F$ 
    
```

Figure 3.1. Proposed pseudocode approach (SSA with mutation operator).

In this paper, we used k-nearest neighbour (KNN) as classifier, where $k = 5$. The KNN is a nonparametric method, which determines the nearest neighbours as optimal solutions. KNN is a predictor of weight variables at a distance based on trial and error processes [42]. KNN method is one of the simplest supervised learning techniques, in which classifying a new unknown sample depends on the distance between the new sample and the training dataset samples [43].

To balance between the classification accuracy (maximum value) and the number of selected features (minimum value) in each solution, the following fitness function in Eq. (3.1) is applied to evaluate the search agents of each SSAMUT, native SSA, ALO and PSO position:

$$F = \alpha * ERR(D) + \beta \frac{|L|}{|T|} \quad (3.1)$$

where $ERR(D)$ indicates the error rate of KNN, $|L|$ is the length of selected feature subset, $|T|$ is the total number of attributes in the original dataset and α and β are used as constant parameters to control the classification quality and feature subset selection; α belongs to a binary dimension between $[0, 1]$, and β is $1-\alpha$ adopted by [42].

This approach represents a hybridisation between SSA and the mutation operator. The mutation operator is embedded within SSA to maintain diversity through exploring many areas and allowing the SSA to avoid local minima by preventing the similarity between individuals of the population. The mutation operation is controlled by an amplification weight factor (a factor lies between 1 and 0), which is called a mutation rate and defined by the user. The optimal mutation rate is a common problem in this field, and it should be set at a low rate. On the contrary, if this rate is set at a high value, then the search will deflect into a random search and prevent the algorithm to converge to any optimal solution. Equation (3.2) shows the mutation schemes.

$$X_i^{t+1} = Mutation(X_i^t) \quad (3.2)$$

In SSA, the population is divided to two classes: leader Salps are at the forefront, and followers are the remaining Salps of the chain. Only the leader position updates with regard to the target source, whereas the position of follower Salps updates on the basis of Newton's law of motion. Therefore, we used the mutation operator on the leader position regardless of the follower position. Given that the exploration in the algorithm depends on the changing of position of each search agent, we selected one or more bits and flipped them. Mutation is essential to the convergence of the SSA. Figure 3.1 illustrates the SSAMUT pseudocode.

4. EMPIRICAL RESULT AND DISCUSSION

4.1. Dataset and parameters

The performance of the proposed approaches was evaluated on 16 benchmark datasets from UCI Machine Learning repository [44]. Given that SSAMUT is a population-based technique, we assumed that each individual in its population represents an index vector for the features in data.

In the basic SSAMUT, we saved only the most remarkable individual and its fitness in each iteration. We compared the remaining individuals with the quality or the classification accuracy for the given feature subset. In this study, the results were compared with native SSA and FS methods, namely, ALO and PSO, to select the optimal feature subset from the original dataset and prove the performance by using all features in the dataset. The global and specific parameter settings are summarised in Table 2.

Table 1. Dataset description

Dataset	No. of Attributes	No. of Objects
Breastcancer	9	699
Exactly	13	1000
Exactly2	13	1000
Breast_EW	30	699
Congress_EW	16	535
Heart_EW	13	270
Ionosphere_EW	34	351
Krvskp_EW	36	3196
Lymphography	18	148
M-of-n	13	1000
Sonar_EW	60	208
Spect_EW	22	267
Tic-tac-toe	9	958
Waveform_EW	40	5000
Wine_EW	13	178
Zoo	16	101

Table 2. Setting of experimental parameter

Parameter	Value
Search agent number	5
Iteration number	100
Repetitions of runs	20
Dimension	No attribute in the dataset
Search domain	Binary [0 1]
α parameter	0.01
β parameter	1- α
Mutation rate	0.05

4.2 Evaluation criteria

All datasets were split into three different and equal sections, namely, training dataset to fit the selection model, validation dataset to infer prediction error and testing dataset to provide an equitable evaluation of a final model. We used the

following indicators to evaluate and compare the different optimisation methods with the proposed approach based on SSA and determine the most remarkable approach. Best, worst, mean, average selection size and standard deviation (std) are represented as the following mathematical equations:

Statistical best fitness: represents the most upbeat solution gained, is the minimum value of fitness function at the different M operations of an optimization method that obtained for a given optimizer. Best fitness can be formulated in Eq (4.1).

$$best = \min_{i=1}^M g_*^i, \quad (4.1)$$

where M is the times number to run the optimizer to select subset of features, and g_*^i is the best solution obtained from run number i .

Statistical worst fitness: is represents the worst solution among all possible solution that can obtain for running optimizer. Worst can be formulated in equation (4.2).

$$Worst = \min_{i=1}^M g_*^i, \quad (4.2)$$

where M is the times number to run the optimizer to select subset of features, and g_*^i is the best solution obtained from run number i .

Statistical mean fitness: is the average performance, indicate to the average of solutions obtained from running an optimizer with different M running, Mean can be represented in equation (4.3).

$$Mean = \frac{1}{M} \sum_{i=1}^M g_*^i, \quad (4.3)$$

where M is the times number to run the optimizer to select subset of features, and g_*^i is the best solution obtained from run number i .

standard deviation (std) : is indicate to the variation of the acquired optimal solutions from running a stochastic optimizer with M different runs. Std is used as an indicator for the stability and robustness of optimizer, Std value, if smaller that indicate the optimizer converges to same solution. On the contrary, while std value is larger that mean much random results and can be formulated as in Eq (32).

$$Std = \sqrt{\frac{1}{M-1} \sum (g_i^* - Mean)^2}, \quad (4.4)$$

where M is the times number to run the optimizer to select subset of features, and g_i^* is the best solution obtained from run number i .

Average selection size: represents the average selected features size to the total number of features. Average selection size can be defined in Eq (4.5).

$$Selection = \frac{1}{M} \sum_{i=1}^M \frac{size(g_i^*)}{D}, \quad (4.5)$$

where M is the times number to run the optimizer to select subset of features, and g_i^* is the best solution obtained from run number i . $size(x)$ is the features number of the vector x , and D is the total number of features in the original data set.

4.3 Results and discussion

In the proposed approach, the SSAMUT embedded a mutation operator in SSA to act as an internal operator. The SSAMUT was compared with the native SSA and other FS algorithms, including PSO and ALO, on the basis of the following evaluation criteria: classification accuracy; statistically best, worst and mean fitness; standard deviation (std); and the average selection size. All results from evaluation criteria were computed with an average of 20 runs using Matlab framework. The SSAMUT performance was compared with that of native SSA on the basis of two objectives, namely, classification accuracy and average selected size.

As shown in Table 3, the hybrid mutation operator in native Salp swarm algorithm is considerably better than that of the native SSA in terms of both objectives: number of selected features and classification accuracy. Comparison of SSAMUT and SSA indicated that SSAMUT performs better than native SSA over all datasets in terms of classification accuracy. In terms of average selected features, SSAMUT outperforms native SSA over many datasets. In the Table 4, SSAMUT performance was also compared with that of other related approaches in the literature, including PSO and ALO. According to the results in Tables 3, 4, the accuracy performance by using complete features is worse than selecting optimal features by using the approaches proposed. Moreover,

SSAMUT outperforms other optimisers on all datasets.

Table 3. Comparison between the proposed approach and native SSA approach in terms of classification accuracy

DATASET	FULL	SSAMUT	SSA
BREASTCANCER	0.94	0.9680	0.9577
EXACTLY	0.67	0.8834	0.7166
EXACTLY2	0.74	0.7278	0.6847
BREAST_EW	0.96	0.9525	0.9370
CONGRESS_EW	0.92	0.9537	0.9362
HEART_EW	0.82	0.8122	0.7737
IONOSPHERE_EW	0.87	0.8960	0.8832
KRVSKP_EW	0.92	0.9496	0.9011
LYMPHOGRAPHY	0.68	0.8372	0.7708
M-OF-N	0.85	0.9374	0.8276
SONAR_EW	0.62	0.8918	0.8519
SPECT_EW	0.83	0.8157	0.7896
TIC-TAC-TOE	0.72	0.7891	0.7646
WAVEFORM_EW	0.77	0.7283	0.7077
WINE_EW	0.93	0.9752	0.9522
ZOO	0.79	0.9647	0.9400

Table 4. Comparison between the proposed approach and other optimiser approaches in terms of classification accuracy

DATASET	SSAMUT	PSO	ALO
BREASTCANCER	0.9680	0.9559	0.9597
EXACTLY	0.8834	0.7115	0.7047
EXACTLY2	0.7278	0.6695	0.6918
BREAST_EW	0.9525	0.9326	0.9395
CONGRESS_EW	0.9537	0.9321	0.9378
HEART_EW	0.8122	0.7752	0.7793
IONOSPHERE_EW	0.8960	0.8639	0.8835
KRVSKP_EW	0.9496	0.9069	0.9044
LYMPHOGRAPHY	0.8372	0.7649	0.7767
M-OF-N	0.9374	0.8283	0.8249
SONAR_EW	0.8918	0.8385	0.8538
SPECT_EW	0.8157	0.7653	0.7869
TIC-TAC-TOE	0.7891	0.7332	0.7603
WAVEFORM_EW	0.7283	0.7040	0.7095

WINE_EW	0.9752	0.9489	0.9539
Zoo	0.9647	0.9370	0.9429

Tables 5, 6 and 7 provide a summary of the results obtained from the different measurements (best, worst, mean, respectively) on the 20 runs of the approaches on all the datasets. SSAMUT approach was compared with native SSA and other FS methods. Results showed that SSAMUT outperforms native SSA, PSO and ALO in statistical best, worst and mean criteria on all datasets. Notably, the proposed approach obtains the lowest values in all datasets. In Table 8, the results that obtained from standard deviation measure indicate to the variation of the acquired optimal solutions from running stochastic optimizers with 20 different runs. SSAMUT approach outperforms other related approaches in the literature over thirteen datasets. Average selected size in Table 9 demonstrated that the SSAMUT performs better than those of other approaches over fourteen datasets.

Table 5. Statistical best fitness results

DATASET	SSAMUT	PSO	ALO
BREASTCANCER	0.0299	0.0332	0.0299
EXACTLY	0.0192	0.1554	0.2686
EXACTLY2	0.2423	0.2585	0.2699
BREAST_EW	0.0411	0.0890	0.0884
CONGRESS_EW	0.0250	0.0354	0.0407
HEART_EW	0.1536	0.1918	0.1926
IONOSPHERE_EW	0.0849	0.0894	0.0894
KRVSKP_EW	0.0346	0.0607	0.0562
LYMPHOGRAPHY	0.1276	0.1806	0.1639
M-OF-N	0.0054	0.0841	0.1384
SONAR_EW	0.0723	0.0937	0.1048
SPECT_EW	0.1564	0.1680	0.1745
TIC-TAC-TOE	0.1959	0.2051	0.2167
WAVEFORM_EW	0.2634	0.2815	0.2829
WINE_EW	0.0165	0.0211	0.0211
ZOO	0.0063	0.0254	0.0250

Table 6. Statistical worst fitness results

DATASET	SSAMUT	PSO	ALO
BREASTCANCER	0.0445	0.0570	0.0626
EXACTLY	0.2334	0.3577	0.3407
EXACTLY2	0.3047	0.3546	0.3399
BREAST_EW	0.0663	0.0890	0.0883
CONGRESS_EW	0.0717	0.0860	0.0917
HEART_EW	0.2203	0.2651	0.2644
IONOSPHERE_EW	0.1530	0.1631	0.1569
KRVSKP_EW	0.0719	0.1710	0.1296
LYMPHOGRAPHY	0.2068	0.2893	0.2904
M-OF-N	0.1206	0.2900	0.2120
SONAR_EW	0.1576	0.2044	0.2002
SPECT_EW	0.2331	0.2862	0.2759
TIC-TAC-TOE	0.2351	0.3032	0.2808
WAVEFORM_EW	0.2886	0.3275	0.3189
WINE_EW	0.0506	0.0848	0.0990
ZOO	0.1027	0.1252	0.1264

Table 7. Statistical mean fitness results

DATASET	SSAMUT	PSO	ALO
BREASTCANCER	0.0383	0.0468	0.0626
EXACTLY	0.1221	0.2793	0.3023
EXACTLY2	0.2761	0.3099	0.3115
BREAST_EW	0.0541	0.0680	0.0665
CONGRESS_EW	0.0521	0.0663	0.0681
HEART_EW	0.1923	0.2226	0.2271
IONOSPHERE_EW	0.1090	0.1198	0.1217
KRVSKP_EW	0.0571	0.0880	0.1038
LYMPHOGRAPHY	0.1676	0.2234	0.2286
M-OF-N	0.0686	0.1666	0.1826
SONAR_EW	0.1140	0.1447	0.1532
SPECT_EW	0.1886	0.2135	0.2178
TIC-TAC-TOE	0.2156	0.2582	0.2464
WAVEFORM_EW	0.2766	0.2934	0.2971
WINE_EW	0.0312	0.0546	0.0540
ZOO	0.0411	0.0655	0.0654

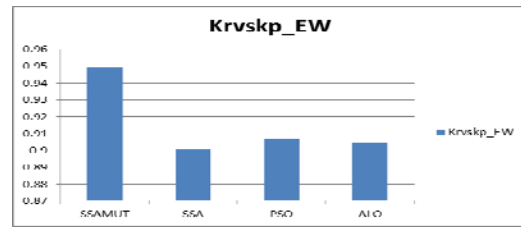
Table 8. Most remarkable solution in terms of standard deviation of all optimisers

DATASET	SSAMUT	PSO	ALO
BREASTCANCER	0.0045	0.0062	0.0473
EXACTLY	0.0643	0.0461	0.0191
EXACTLY2	0.0150	0.0213	0.0178
BREAST_EW	0.0080	0.0106	0.0129
CONGRESS_EW	0.0106	0.0118	0.0148
HEART_EW	0.0182	0.0231	0.0179
IONOSPHERE_EW	0.0169	0.0198	0.0197
KRVSKP_EW	0.0103	0.0262	0.0216
LYMPHOGRAPHY	0.0237	0.0351	0.0331
M-OF-N	0.0385	0.0392	0.0207
SONAR_EW	0.0245	0.0288	0.0328
SPECT_EW	0.0230	0.0312	0.0306
TIC-TAC-TOE	0.0108	0.0247	0.0159
WAVEFORM_EW	0.0069	0.0115	0.0074
WINE_EW	0.0095	0.0159	0.0180
ZOO	0.0311	0.0356	0.0330

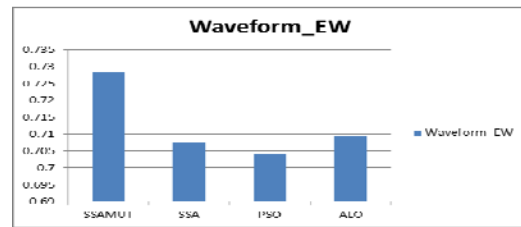
Table 9. Average selected size for the different approaches

Dataset	SSAMUT	PSO	ALO
Breastcancer	0.6611	0.8556	0.7444
Exactly	0.6692	0.9077	0.9923
Exactly2	0.6654	0.8077	0.6423
Breast_EW	0.7033	0.7366	0.6583
Congress_EW	0.6218	0.7531	0.6531
Heart_EW	0.6385	0.8385	0.8654
Ionosphere_EW	0.6132	0.7559	0.6352
Krvskp_EW	0.7153	0.8722	0.9153
Lymphography	0.6389	0.7667	0.7528
M-of-n	0.6692	0.9038	0.9346
Sonar_EW	0.6917	0.7800	0.8508
Spect_EW	0.6136	0.7000	0.6840
Tic-tac-toe	0.6889	0.8222	0.9111
Waveform_EW	0.7625	0.870	0.9500
Wine_EW	0.6730	0.8462	0.8462
Zoo	0.6125	0.8219	0.8937

In this study, we used two high-dimensional datasets, namely, Krvskp_EW (36 attributes with 3196 objects) and Waveform_EW (40 attributes with 5000 objects), which are large datasets. Results obtained from different measures showed that the SSAMUT approach evidently outperforms other FS algorithms. This approach is also better than native SSA, PSO and ALO in terms of the obtained solution on standard deviation.



(a)



(b)

Figure 4.1 Average performance over the test data averaged over high-dimensional data sets (a) & (b) using the different optimizers

Generally, on the basis of the final results, we can conclude that the SSAMUT model considerably enhances the native SSA performance. SSAMUT investment also improves the exploration ability in SSA and allows this algorithm to avoid local minima by preventing the similarity between the individuals of the population. Result showed the capability of SSAMUT in locating high-performance regions in the search space.

5. CONCLUSIONS

This study aimed to use the SSA to reduce dimensionality through selecting an optimal feature subset on the basis of certain criteria and improving classification accuracy by using all features of the dataset. In first approach, native SSA was applied in a wrapper-based mode to solve FS problem in data mining tasks. The proposed second approach (SSAMUT) was hybrid with mutation operator and compared with the native SSA and other optimisation algorithms, including PSO and ALO.

In Figure 5.1, obviously, the native SSA approach got stuck in a local optimum in the earlier step. Whereas, the proposed SSAMUT approach can rapidly expand the search spot in the search space by tweaking the values in a higher level of mutation. Since feature selection is an expensive problem, less function evaluation times (fes) with high exploration are desired. That makes the second proposed SSAMUT approach more suitable for feature selection than the first one.

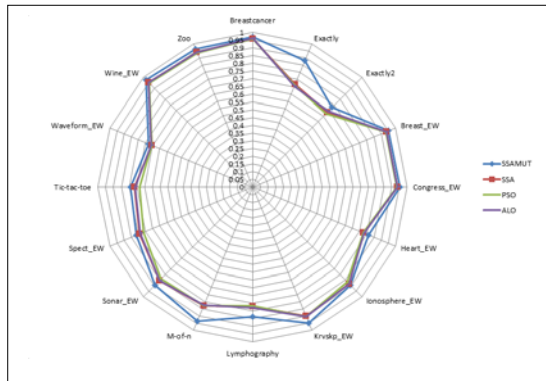


Figure 5.1: classification accuracy over the test data averaged over all the datasets using the SSAMUT and SSA optimizer.

On the basis of common evaluation criteria. We found that SSAMUT showed evidently remarkable performance in all selected features. The SSAMUT approach maintained diversity and improved the exploration ability within the SSA. SSA can be applied as a filter-based method or hybridised with other metaheuristic algorithms, such as ALO, in future works.

REFERENCES:

- [1] Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI magazine*. 1996 Mar 15;17(3):37.
- [2] Zhao ZA, Liu H. *Spectral feature selection for data mining*. CRC Press; 2011 Dec 14.
- [3] Zhang Z, Ning Y. Effective semi-supervised nonlinear dimensionality reduction for wood defects recognition. *Computer Science and Information Systems*. 2010;7:127-38.
- [4] Liu H, Motoda H. *Feature selection for knowledge discovery and data mining*. Springer Science & Business Media; 2012 Dec 6.
- [5] Wang Y, Dahnoun N, Achim A. A novel system for robust lane detection and tracking. *Signal Processing*. 2012 Feb 29;92(2):319-34.
- [6] Tsai JS, Huang WB, Kuo YH, Horng MF. Joint robustness and security enhancement for feature-based image watermarking using invariant feature regions. *Signal Processing*. 2012 Jun 30;92(6):1431-45.
- [7] Zanaty EA, Ghiduk AS. A novel approach based on genetic algorithms and region growing for magnetic resonance image (MRI) segmentation. *Computer Science and Information Systems*. 2013;10(3):1319-42.
- [8] Choi SI, Oh J, Choi CH, Kim C. Input variable selection for feature extraction in classification problems. *Signal Processing*. 2012 Mar 31;92(3):636-48.
- [9] Wang Y, Dahnoun N, Achim A. A novel system for robust lane detection and tracking. *Signal Processing*. 2012 Feb 29;92(2):319-34.
- [10] Lutu PE, Engelbrecht AP. A decision rule-based method for feature selection in predictive data mining. *Expert Systems with Applications*. 2010 Jan 31;37(1):602-9.
- [11] Liu Y, Zheng YF. FS_SFS: A novel feature selection method for support vector machines. *Pattern recognition*. 2006 Jul 31;39(7):1333-45.
- [12] Sikora R, Piramuthu S. Framework for efficient feature selection in genetic algorithm based data mining. *European Journal of Operational Research*. 2007 Jul 16;180(2):723-37.
- [13] Du KL, Swamy MN. *Search and Optimization by Metaheuristics*.
- [14] Talbi EG. *Metaheuristics: from design to implementation*. John Wiley & Sons; 2009 May 27.
- [15] Zawbaa HM, Emary E, Parv B. Feature selection based on antlion optimization algorithm. In *Complex Systems (WCCS), 2015 Third World Conference on 2015 Nov 23* (pp. 1-7). IEEE.
- [16] Kennedy J, Eberhart RC. A discrete binary version of the particle swarm algorithm. In *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on 1997 Oct 12* (Vol. 5, pp. 4104-4108). IEEE.
- [17] Kanan H, Faez K, Taheri S. Feature selection using ant colony optimization (ACO): a new method and comparative study in the application of face recognition system. *Advances in Data Mining. Theoretical Aspects and Applications*. 2007:63-76.

- [18] Huang J, Cai Y, Xu X. A hybrid genetic algorithm for feature selection wrapper based on mutual information. *Pattern Recognition Letters*. 2007 Oct 1;28(13):1825-44.
- [19] Banati H, Bajaj M. Firefly based feature selection approach. *International Journal of Computer Science Issues*. 2011 Jul;8(4):473-80.
- [20] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. *Advances in Engineering Software*. 2014 Mar 31;69:46-61.
- [21] Zhang Y, Wang S, Phillips P, Ji G. Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. *Knowledge-Based Systems*. 2014 Jul 31;64:22-31.
- [22] Chuang LY, Tsai SW, Yang CH. Improved binary particle swarm optimization using catfish effect for feature selection. *Expert Systems with Applications*. 2011 Sep 15;38(10):12699-707.
- [23] Martin OC, Otto SW. Combining simulated annealing with local search heuristics. *Annals of Operations Research*. 1996 Feb 1;63(1):57-75.
- [24] Lin Y, Bian Z, Liu X. Developing a dynamic neighborhood structure for an adaptive hybrid simulated annealing-tabu search algorithm to solve the symmetrical traveling salesman problem. *Applied Soft Computing*. 2016 Dec 31;49:937-52.
- [25] Lenin K, Reddy BR, Suryakalavathi M. Hybrid Tabu search-simulated annealing method to solve optimal reactive power problem. *International Journal of Electrical Power & Energy Systems*. 2016 Nov 30;82:87-91.
- [26] Azmi R, Pishgoo B, Norozi N, Koohzadi M, Baesi F. A hybrid GA and SA algorithms for feature selection in recognition of hand-printed Farsi characters. In *Intelligent Computing and Intelligent Systems (ICIS)*, 2010 IEEE International Conference on 2010 Oct 29 (Vol. 3, pp. 384-387). IEEE.
- [27] Olabiyisi Stephen O, Fagbola Temitayo M, Omidiora Elijah O, Oyeleye Akin C. Hybrid MetaHeuristic Feature Extraction Technique for Solving Timetabling Problem.
- [28] Tang WC. Feature Selection For The Fuzzy Artmap Neural Network Using A Hybrid Genetic Algorithm And Tabu Search (Doctoral dissertation, USM).
- [29] Hadi AS, Fergus P, Dobbins C, Al-Bakry AM. A machine learning algorithm for searching vectorised RDF data. In *Advanced Information Networking and Applications Workshops (WAINA)*, 2013 27th International Conference on 2013 Mar 25 (pp. 613-618). IEEE.
- [30] Oh IS, Lee JS, Moon BR. Hybrid genetic algorithms for feature selection. *IEEE Transactions on pattern analysis and machine intelligence*. 2004 Nov;26(11):1424-37.
- [31] Vasant P. Hybrid simulated annealing and genetic algorithms for industrial production management problems. *International Journal of Computational Methods*. 2010 Jun;7(02):279-97.
- [32] Wu J, Lu Z. A novel hybrid genetic algorithm and simulated annealing for feature selection and kernel optimization in support vector regression. In *Advanced Computational Intelligence (ICACI)*, 2012 IEEE Fifth International Conference on 2012 Oct 18 (pp. 999-1003). IEEE.
- [33] Talbi EG, Jourdan L, Garcia-Nieto J, Alba E. Comparison of population based metaheuristics for feature selection: Application to microarray data classification. In *Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on 2008 Mar 31* (pp. 45-52). IEEE.
- [34] Yusta SC. Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognition Letters*. 2009 Apr 1;30(5):525-34.
- [35] Chizi B, Rokach L, Maimon O. A survey of feature selection techniques. In *Encyclopedia of Data Warehousing and Mining, Second Edition 2009* (pp. 1888-1895). IGI Global.
- [36] Molina LC, Belanche L, Nebot À. Feature selection algorithms: A survey and experimental evaluation. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on 2002* (pp. 306-313). IEEE.
- [37] Chandrashekar G, Sahin F. A survey on feature selection methods. *Computers & Electrical Engineering*. 2014 Jan 31;40(1):16-28.
- [38] Boussaïd I, Lepagnot J, Siarry P. A survey on optimization metaheuristics. *Information Sciences*. 2013 Jul 10;237:82-117.
- [39] Madin LP. Field observations on the feeding behavior of salps (Tunicata: Thaliacea). *Marine Biology*. 1974 May 1;25(2):143-7.
- [40] Anderson PA, Bone Q. Communication between individuals in salp chains II. *Physiology*. *Proceedings of the Royal Society of London B: Biological Sciences*. 1980 Nov 28;210(1181):559-74.
- [41] Mirjalili S, Gandomi AH, Mirjalili SZ, Saremi S, Faris H, Mirjalili SM. Salp swarm algorithm:

- a bio-inspired optimizer for engineering design problems. *Advances in Engineering Software*. 2017 Jul 24.
- [42] Emary E, Zawbaa HM, Hassanien AE. Binary grey wolf optimization approaches for feature selection. *Neurocomputing*. 2016 Jan 8;172:371-81.
- [43] Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*. 1992 Aug 1;46(3):175-85.
- [44] Frank A, Asuncion A. UCI Machine Learning Repository. Irvine, CA: University of California. School of information and computer science. 2010;213.