

# S-Shaped versus V-Shaped Transfer Functions for Binary Manta Ray Foraging Optimization in Feature Selection Problem

Kushal Kanti Ghosh · Ritam Guha · Suman Kumar Bera\* · Neeraj Kumar · Ram Sarkar

the date of receipt and acceptance should be inserted later

**Abstract** Feature selection (FS) is considered as one of the core concepts in the areas of machine learning and data mining which immensely impacts the performance of classification model. Through FS, irrelevant or partially relevant features can be eliminated which in turn helps in enhancing the performance of the model. Over the years, researchers have applied different meta-heuristic optimization techniques for the purpose of FS as these overcome the limitations of traditional optimization approaches. Going by the trend, we introduce a new FS approach based on a recently proposed meta-heuristic algorithm called Manta Ray Foraging Optimization (MRFO) which is developed following the food foraging nature of the Manta rays, one of the largest known marine creatures. As MRFO is apposite for continuous search space problems, we have adapted a binary version of MRFO to fit it into the problem of FS by applying eight different transfer functions belonging to two different families: S-shaped and V-shaped. We have evaluated the eight binary versions of MRFO on 18 standard UCI datasets. Of these, the best one is considered for comparison with 16 recently proposed meta-heuristic FS approaches. The results show that MRFO outperforms the state-of-art methods in terms of both classification accuracy and number of features selected.

---

Kushal Kanti Ghosh, ORCID: 0000-0003-0929-5928

E-mail: kushalkanti1999@gmail.com

Department of Computer Science and Engineering, Jadavpur University, Kolkata, India

Ritam Guha, ORCID: 0000-0002-1375-777X

E-mail: ritamguha16@gmail.com

Department of Computer Science and Engineering, Jadavpur University, Kolkata, India

Suman Kumar Bera, ORCID: 0000-0001-6968-2079

E-mail: sumankumarbera.cse.rs@jadavpuruniversity.in

Department of Computer Science and Engineering, Jadavpur University, Kolkata, India

Neeraj Kumar, ORCID: 0000-0002-3020-3947

E-mail: neeraj.kumar@thapar.edu

Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Punjab, India

Ram Sarkar, ORCID: 0000-0001-8813-4086

E-mail: rsarkar@ieee.org

Department of Computer Science and Engineering, Jadavpur University, Kolkata, India

**Keywords** Manta Ray Foraging Optimization, Feature Selection, Optimization, Classification, Transfer Functions, Metaheuristic

## 1 Introduction

In recent times, due to the huge amount of data collected every minute and the need to convert such data into useful information, data mining is considered one of the fastest growing fields of Information Technology [1]. Data mining involves: pre-processing, knowledge representation and pattern evaluation [2]. One of the main data pre-processing steps is feature selection (FS). FS can be defined as the process of finding the optimal subset of features that can retain suitably high classification accuracy in representing the original feature set [2]. Through FS, we can remove irrelevant and redundant features from the datasets. In the learning process, irrelevant and redundant features not only mislead the learning algorithm and reduce the performance but also result in increased computational complexity and high storage requirement [3]. FS techniques are broadly classified into two categories [4]: filter and wrapper. Filter methods do not require any learning algorithm, rather they use statistical methods to evaluate a feature subset or determine correlation between variables. On the other hand, wrapper models use learning algorithm, like classifiers for the evaluation of feature subsets. Filters are much faster than wrappers but wrappers are capable to achieve higher classification accuracies than filters [4].

Over the last two decades, several meta-heuristic optimization techniques have been proposed in the field of FS to overcome the limitations of traditional optimization approaches [5]. Nature inspired algorithms have shown high performance in solving the search problems in general [5]. In the literature, many methods have been proposed to mimic the behaviours of animals, birds, fish, wolves etc. The first proposed FS technique using meta-heuristic approach is Genetic Algorithm (GA) [6]. A chaotic genetic feature selection optimization method (CGFSO) is proposed by the authors in [7]. Particle Swarm Optimization (PSO) is proposed in [8] by mimicking the social synergy of a flock of birds. Many FS approaches based on PSO are proposed in [9], [10], [11]. Ant Colony Optimization (ACO) [12] is another meta-heuristic algorithm that mimics the behaviour of real ants when searching for the shortest path to a food source. ACO based FS approaches are proposed in [13], [14]. Following the biological behaviours of bees, Artificial Bee Colony (ABC) optimization algorithm [15] and its FS counterpart [16] are also proposed. Some recent optimization algorithms and their FS counterparts are Grey Wolf Optimizer (GWO) [4], Whale Optimization Algorithm (WOA) [17], Ant Lion Optimizer (ALO) [5], Gravitational Search Algorithm [18].

The authors of the work reported in [19] have proposed two binary variants of ALO using S-Shaped and V-Shaped transfer functions. In [20], the authors have proposed six variants of ALO using three S-Shaped and three V-Shaped transfer functions. In [21], authors have proposed binary version of Grasshopper Optimization Algorithm (BGOA) using S-Shaped and V-Shaped functions. In [22], the authors proposed six transfer functions and eight binary versions of PSO.

Since there already exists so many optimization algorithms in literature that perform quite well, the question is whether we need another new optimization approach. According to *No Free Lunch* theorem for optimization [23], there cannot

exist a single algorithm to solve all optimization problems. This, in turn, implies that currently proposed algorithms for FS are not able to solve all FS problems. This motivated us to propose a new FS approach based on a recently proposed meta-heuristic algorithm, Manta Ray Foraging Optimization (MRFO) [24]. We have proposed the binary version of MRFO for FS problems. We have proposed eight different versions of the FS technique based on four S-Shaped and four V-Shaped transfer functions. The methods are evaluated on 18 standard datasets and compared with different recently proposed meta-heuristic FS techniques to validate the performance of the same.

## 2 Present Work

### 2.1 Manta Ray Foraging Optimization: A Brief Overview

MRFO is first proposed in [24], inspired by Manta rays, one of the largest known marine creatures. Manta rays feed on plankton, mostly made of microscopic animals from the water. They have different types of foraging strategies:

1. **Chain foraging** [25]: Manta rays observe the plankton position and swim towards it by forming an orderly line. So, the plankton that is missed by the previous manta rays will be devoured by the following manta rays. The higher the concentration of plankton in a position, the better that position is. Now, except the first, all manta rays move towards both the best position and the one in front of it. The mathematical model of chain foraging is given by Equation 1.

$$x_i^d(t+1) = \begin{cases} x_i^d(t) + r * (x_{best}^d(t) - x_i^d(t)) + \alpha * (x_{best}^d(t) - x_i^d(t)) & \text{if } i = 1 \\ x_i^d(t) + r * (x_{i-1}^d(t) - x_i^d(t)) + \alpha * (x_{best}^d(t) - x_i^d(t)) & \text{if } i = 2, \dots, N \end{cases} \quad (1)$$

$$\alpha = 2 * r * \sqrt{|\log(r)|} \quad (2)$$

where,  $x_i^d(t)$  is the position of  $i^{th}$  individual in  $t^{th}$  iteration at  $d^{th}$  dimension.  $r$  is a random number,  $r \in [0, 1]$ .  $\alpha$  is a weight coefficient.  $x_{best}^d$  is the position with highest plankton concentration. Although the actual best solution is not known, MRFO assumes the position with highest concentration found so far as the best solution.

2. **Cyclone foraging** [26]: When a school of manta rays finds a plankton with high concentration in deep water, they form a spiral by creating head-to-tail links. In this formation, each manta ray moves towards the plankton position as well as the one in front of it. The mathematical model of cyclone foraging is given by Equation 3.

$$x_i^d(t+1) = \begin{cases} x_{best}^d(t) + r * (x_{best}^d(t) - x_i^d(t)) + \beta * (x_{best}^d(t) - x_i^d(t)), & \text{if } i = 1 \\ x_{best}^d(t) + r * (x_{i-1}^d(t) - x_i^d(t)) + \beta * (x_{best}^d(t) - x_i^d(t)), & \text{if } i = 2, \dots, N \end{cases} \quad (3)$$

$$\beta = 2e^{r_1 \frac{T-t+1}{T}} * \sin(2\pi r_1) \quad (4)$$

where,  $\beta$  is a weight coefficient,  $T$  is the maximum number of iterations,  $r_1$  is a random number,  $r_1 \in [0, 1]$ . All individuals perform a random search considering

the best so far plankton position as the reference position. So, it works as a good exploitation for the region found best so far. This can also be modified to improve exploration capability of the MRFO algorithm. We can force the manta rays to look for a new position, by assigning a new random position as their reference position. This results in extensive global search. The mathematical model is given by:

$$x_{rand}^d = Lb^d + r * (Ub^d - Lb^d) \quad (5)$$

$$x_i^d(t+1) = \begin{cases} x_{rand}^d(t) + r * (x_{rand}^d(t) - x_i^d(t)) + \beta * (x_{rand}^d(t) - x_i^d(t)), & \text{if } i = 1 \\ x_{rand}^d(t) + r * (x_{i-1}^d(t) - x_i^d(t)) + \beta * (x_{rand}^d(t) - x_i^d(t)), & \text{if } i = 2, \dots, N \end{cases} \quad (6)$$

where,  $x_{rand}^d$  is a random position in the search space,  $Lb^d$  and  $Ub^d$  are lower and upper limits of the  $d^{th}$  dimension respectively.

**3. Somersault foraging** [27]: This is a random, frequent, local and cyclical movement that helps manta rays optimize plankton intake. The best so far plankton position is used as a pivot and each individual swims to and fro around the pivot and somersault to a new position. The mathematical model is given as:

$$x_i^d(t+1) = x_i^d(t) + S \cdot (r_2 \cdot x_{best}^d - r_3 \cdot x_i^d(t)), i = 1, 2, \dots, N \quad (7)$$

where,  $S$  is the Somersault factor,  $r_2$  and  $r_3$  are random numbers in  $[0, 1]$ . As per [24], we have fixed  $S = 2$ .

## 2.2 Proposed Binary MRFO

### 2.2.1 Manta Ray Representation

An FS problem has its own intrinsic discrete binary search space [22]. Hence, we need a vector of 0's and 1's to represent any solution in a FS problem, where 0 represents the corresponding feature not selected and 1 represents selected. The length of the vector is equal to the feature dimension of the dataset under consideration. Accordingly, a manta ray is represented as a binary vector. Problems with continuous real search space can be converted to binary problems by converting their variables to binary variables. In the continuous version of MRFO, manta rays can move around the search space using position vectors within the continuous real domain. In discrete binary search space, update of positions implies switching between 0 and 1. Therefore, we have converted the real position of manta rays to binary values by applying a transfer function to the real position [10].

Transfer function defines the probability of updating the values of binary solution from 0 to 1 and vice-versa. Transfer functions force manta rays to move in a binary space. There are two different types of transfer functions [20]: S-Shaped and V-Shaped. In S-Shaped functions, the positions are updated based on Equation 8.

$$x_i^d(t+1) = \begin{cases} 1 & \text{if } rand < Sfunction(x_i^d(t+1)) \\ 0 & \text{else} \end{cases} \quad (8)$$

**Algorithm 1** Pseudocode of MRFO algorithm

---

```

while Stop Criterion not Satisfied do
  for  $i = 1 \dots N$  do
    if  $rand < 0.5$  then //Cyclone Foraging
      if  $t/T < rand$  then
         $x_{rand} = x_l + rand * (x_u - x_l)$ 
         $x_i^d(t+1) = \begin{cases} x_{rand}^d(t) + r * (x_{rand}^d(t) - x_i^d(t)) + \beta * (x_{rand}^d(t) - x_i^d(t)) & i = 1 \\ x_{rand}^d(t) + r * (x_{i-1}^d(t) - x_i^d(t)) + \beta * (x_{rand}^d(t) - x_i^d(t)) & i = 2, \dots, N \end{cases}$ 
      else
         $x_i^d(t+1) = \begin{cases} x_{best}^d(t) + r * (x_{best}^d(t) - x_i^d(t)) + \beta * (x_{best}^d(t) - x_i^d(t)) & i = 1 \\ x_{best}^d(t) + r * (x_{i-1}^d(t) - x_i^d(t)) + \beta * (x_{best}^d(t) - x_i^d(t)) & i = 2, \dots, N \end{cases}$ 
      end if
    else //Chain Foraging
       $x_i^d(t+1) = \begin{cases} x_i^d(t) + r * (x_{best}^d(t) - x_i^d(t)) + \alpha * (x_{best}^d(t) - x_i^d(t)) & i = 1 \\ x_i^d(t) + r * (x_{i-1}^d(t) - x_i^d(t)) + \alpha * (x_{best}^d(t) - x_i^d(t)) & i = 2, \dots, N \end{cases}$ 
    end if
    Compute fitness for  $i^{th}$  individual  $f(x_i(t+1))$ 
    if  $f(x_i(t+1)) < f(x_{best})$  then
       $x_{best} = x_i(t+1)$ 
    end if
  end for
  //Somersault Foraging
  for  $i = 1 \dots N$  do
     $x_i^d(t+1) = x_i^d(t) + S \cdot (r_2 \cdot x_{best}^d - r_3 \cdot x_i^d(t)), i = 1, 2, \dots, N$ 
    Compute fitness for  $i^{th}$  individual  $f(x_i(t+1))$ 
    if  $f(x_i(t+1)) < f(x_{best})$  then
       $x_{best} = x_i(t+1)$ 
    end if
  end for
end while

```

---

where,  $rand$  is a random number in  $[0, 1]$ . In case of V-Shaped functions, the positions are updated based on Equation 9.

$$x_i^d(t+1) = \begin{cases} \sim x_i^d(t) & \text{if } rand < Vfunction(x_i^d(t+1)) \\ x_i^d(t) & \text{else} \end{cases} \quad (9)$$

In this work, we have studied the impacts of different transfer functions on the performance of MRFO. We have used eight transfer functions [22], four S-Shaped and four V-Shaped transfer functions. Table 1 shows the mathematical formulas of the 8 transfer functions used in this work. Figure 1 shows their corresponding graphs.

Table 1: S-Shaped and V-Shaped families of Transfer functions

Family of S-Shaped functions		Family of V-Shaped functions	
Name	Transfer function	Name	Transfer function
S1	$T(x) = \frac{1}{1+e^x}$	V1	$T(x) =  tanh(x) $
S2	$T(x) = \frac{1}{1+e^{2x}}$	V2	$T(x) =  erf(\frac{\sqrt{\pi}}{2}x) $
S3	$T(x) = \frac{1}{1+e^{3x}}$	V3	$T(x) =  \frac{x}{\sqrt{1+x^2}} $
S4	$T(x) = \frac{1}{1+e^{x/2}}$	V4	$T(x) =  \frac{2}{\pi}arctan(\frac{\pi}{2}x) $

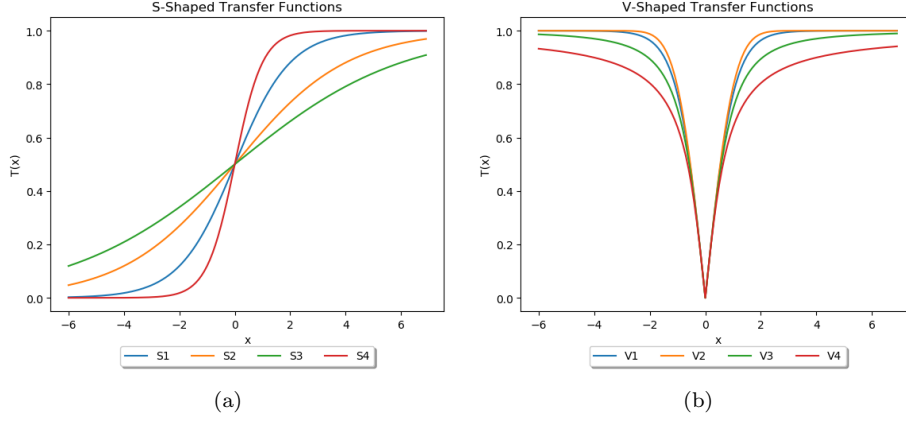


Fig. 1: (a) S-Shaped and (b) V-Shaped family of transfer functions

### 2.2.2 Fitness Function

One important aspect of FS techniques is evaluation of the selected feature subset. Since the proposed method is a wrapper based FS technique, a learning algorithm (classifier) is involved in the evaluation process. We have used a well-known classifier, K-Nearest Neighbour (KNN) classifier [28] for this purpose. FS, in general, has two objectives [29]: higher classification accuracy and lower number of selected features. Both the higher classification accuracy and lower number of selected features indicate that the selected subset is better. We have considered both while creating the fitness function for the proposed FS technique. Since the number of features needs to be minimized, we have chosen classification error, complement of classification accuracy. In this work, the fitness function is defined by Equation 10.

$$\downarrow Fitness = \omega\gamma + (1 - \omega)\frac{|R|}{|D|} \quad (10)$$

where,  $\gamma$  is the classification error of the used classifier,  $|R|$  is number of selected features,  $|D|$  is the total number of features (dimension),  $\omega$  is weight [19],  $\omega \in [0, 1]$ .

## 3 Experimental Results and Discussion

### 3.1 Datasets

The proposed eight binary MRFO algorithms are evaluated on 18 benchmark UCI datasets [30]. The datasets are described in terms of number of attributes, number of samples, number of classes and dataset domain in Table 2.

Table 2: Description of the datasets used in the present work

Sl. No.	Dataset	No. of Attributes	No. of Samples	No. of Classes	Dataset Domain
1	Breastcancer	9	699	2	Biology
2	BreastEW	30	569	2	Biology

3	CongressEW	16	435	2	Politics
4	Exactly	13	1000	2	Biology
5	Exactly2	13	1000	2	Biology
6	HeartEW	13	270	2	Biology
7	IonosphereEW	34	351	2	Electromagnetic
8	KrvskpEW	36	3196	2	Game
9	Lymphography	18	148	2	Biology
10	M-of-n	13	1000	2	Biology
11	PenglungEW	325	73	2	Biology
12	SonarEW	60	208	2	Biology
13	SpectEW	22	267	2	Biology
14	Tic-tac-toe	9	958	2	Game
15	Vote	16	300	2	Politics
16	WaveformEW	40	5000	3	Physics
17	WineEW	13	178	3	Chemistry
18	Zoo	16	101	6	Artificial

### 3.2 Parameter Tuning

The convergence graphs for the 18 datasets with all the eight binary MRFO versions are given in Figure 2. These graphs depict the value of fitness function *vs* no. of iterations for all eight versions of the proposed method. It can be noticed that all the methods have converged within 30 iterations. So, we have set the number of iterations as 30 for all the subsequent experiments.

We have also observed the effect of population size on the performance of MRFO in terms of classification accuracy. For this, the proposed approaches are evaluated with population sizes of 5, 10, 20, 30 and 50. Figure 3 depicts the impact of different population sizes on the attained classification accuracy for all the said datasets. It can be noticed that increasing population size does not always increase classification accuracy. Besides, increased population size means higher run time. So, for all our subsequent experiments, the value of the population size is set to 20 as a trade off between the classification accuracy and the running time of the algorithm. All the proposed approaches are implemented in Python3 [31] and experimented on Intel Core-i3 CPU with 4 GB of RAM.

Table 3 represents the classification accuracies achieved (in percent) by the proposed eight FS methods. *Total* column indicates the classification accuracy of the dataset without applying any FS method.

In Table 3, we have ranked the proposed methods for each dataset according to the achieved classification accuracy and from this, we have obtained average rank of each proposed method. MRFOv3 has achieved rank one *i.e.*, it achieves higher accuracy than other seven proposed methods for most of the datasets used in this work and MRFOv3 performs in a consistent way *i.e.*, does not fluctuate in achieving higher accuracy.

Table 4 shows the number of features selected using the proposed eight FS methods. *Total* column indicates the total number of features in the dataset.

In Table 4, we have ranked the proposed methods for each dataset according to the number of selected features and from this, we have obtained average rank

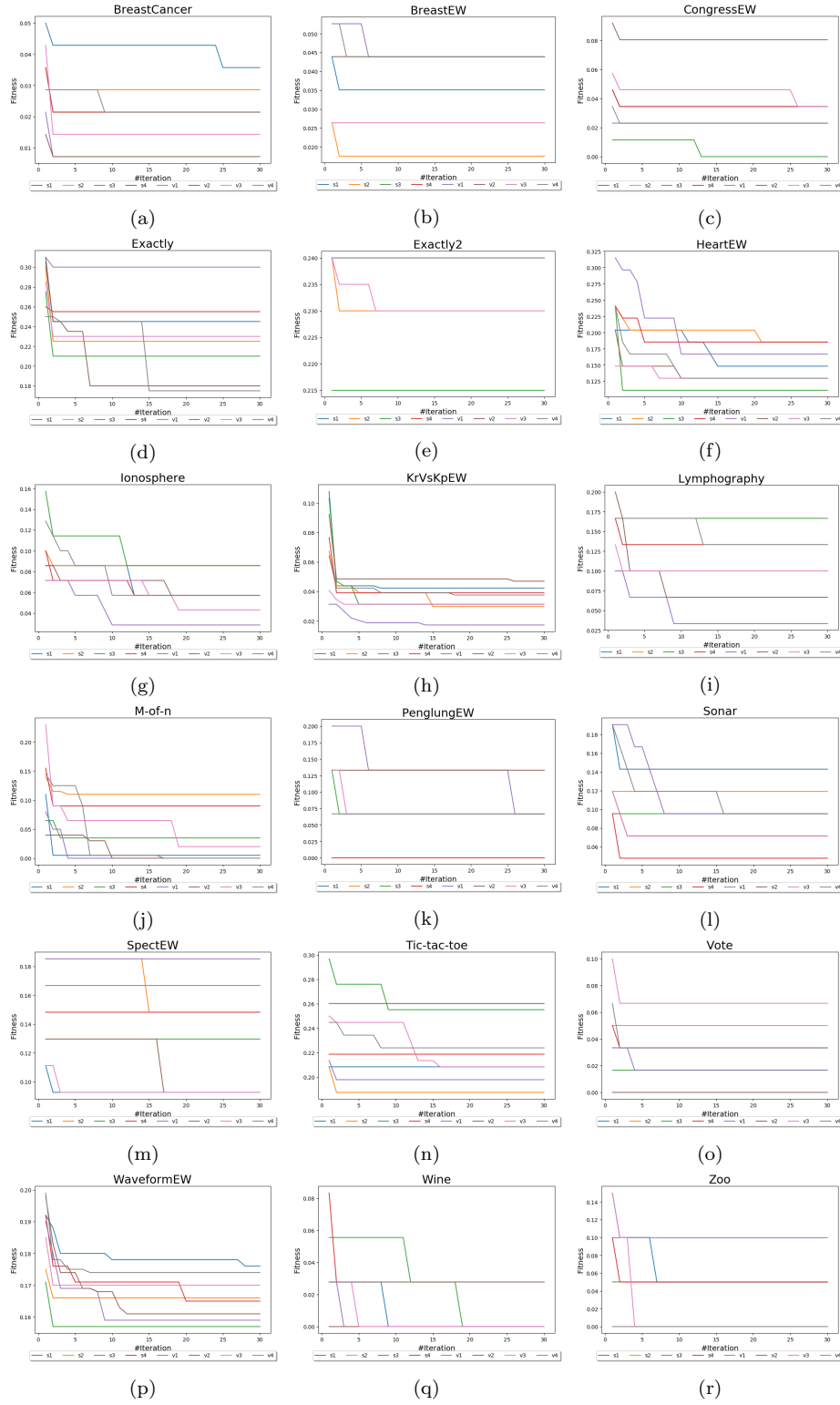


Fig. 2: Convergence graphs for 18 UCI datasets used in present work



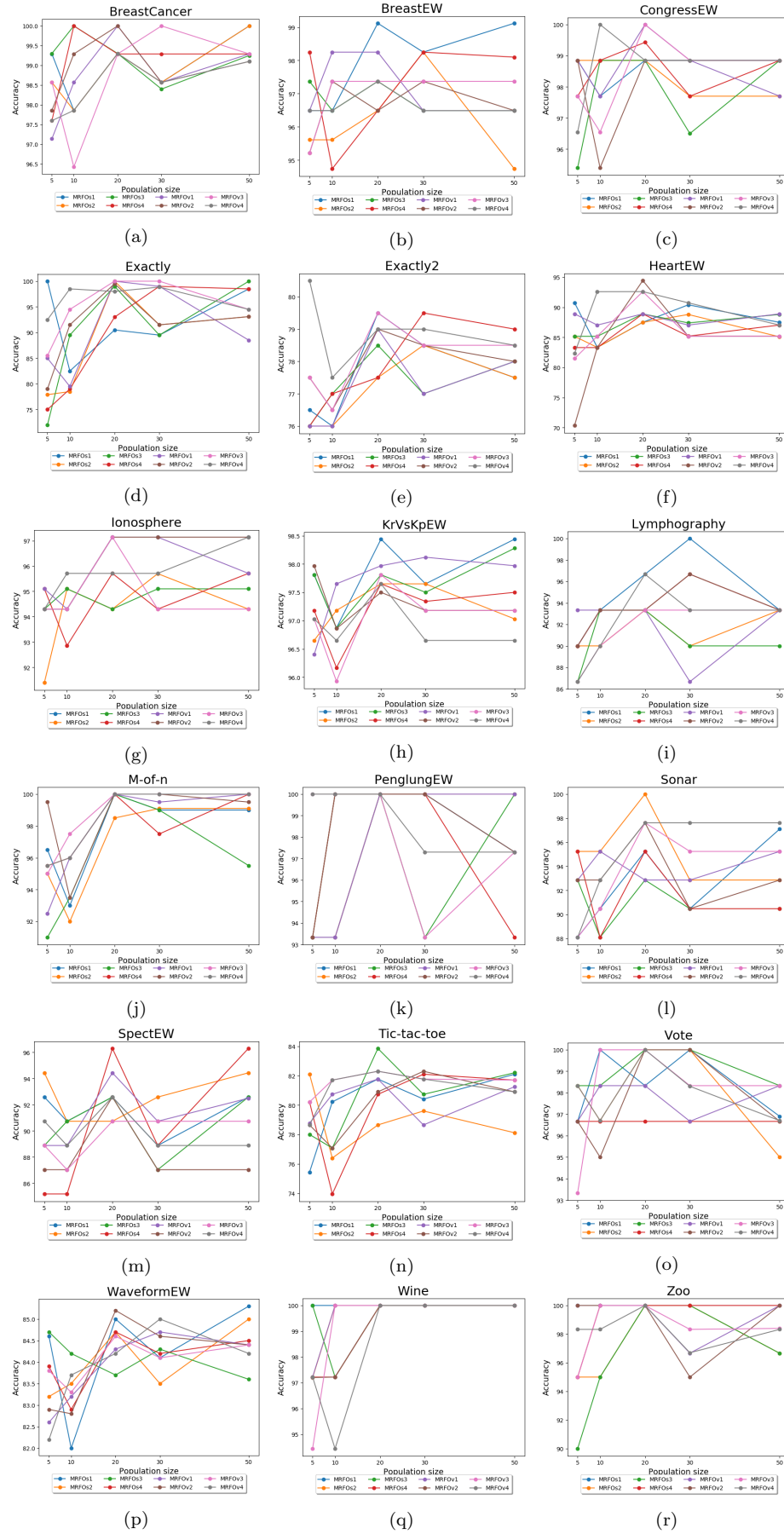


Fig. 3: Accuracy vs Population size for 18 UCI datasets used in present work

Table 3: Classification accuracies (in percentage) of the proposed FS method using eight different transfer functions

	Without FS	MRFOs1	MRFOs2	MRFOs3	MRFOs4	MRFOv1	MRFOv2	MRFOv3	MRFOv4
Breastcancer	94.4	99.29	100	99.29	99.29	<b>100</b>	<b>100</b>	99.29	99.29
BreastEW	96.3	<b>99.12</b>	96.49	97.37	96.49	98.25	96.49	97.37	97.37
CongressEW	91.7	98.85	98.85	98.85	<b>100</b>	<b>100</b>	98.85	<b>100</b>	98.85
Exactly	67.3	90.5	100	99	93	<b>100</b>	99.5	<b>100</b>	<b>100</b>
Exactly2	74.3	<b>79.5</b>	77.5	78.5	77.5	79	79	<b>79.5</b>	79
HeartEW	81.5	87.04	87.04	88.89	85.19	88.89	<b>94.44</b>	92.59	92.59
IonosphereEW	86.6	<b>97.14</b>	94.29	94.29	95.71	<b>97.14</b>	<b>97.14</b>	<b>97.14</b>	95.71
KrvskpEW	91.5	<b>98.44</b>	97.65	97.81	97.65	97.97	97.5	97.81	97.65
Lymphography	68.3	<b>96.67</b>	93.33	93.33	93.33	93.33	93.33	93.33	96.67
M-of-n	84.9	<b>100</b>	98.5	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
PenglungEW	95.1	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
SonarEW	62	95.24	<b>100</b>	92.86	95.24	92.86	97.62	97.62	97.62
SpectEW	83.1	92.59	90.74	92.59	<b>96.3</b>	94.44	92.59	90.74	92.59
Tic-tac-toe	71.5	79.69	78.65	83.85	80.73	81.77	80.73	<b>82.29</b>	82.29
Vote	87.7	98.33	<b>100</b>	<b>100</b>	96.67	98.33	<b>100</b>	<b>100</b>	<b>100</b>
WaveformEW	76.8	84.9	84.7	83.5	84.7	84.3	<b>85.2</b>	84.6	84.2
WineEW	93.2	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
Zoo	83.1	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
sum of ranks	79	37	45	44	46	35	35	33	37
avg of ranks	4.39	2.06	2.5	2.44	2.56	1.94	1.94	1.83	2.06
overall rank	7	3	5	4	6	2	2	1	3

Table 4: No. of selected features with the proposed FS method using eight different transfer functions

	Entire dimension	MRFOs1	MRFOs2	MRFOs3	MRFOs4	MRFOv1	MRFOv2	MRFOv3	MRFOv4
Breastcancer	9	<b>2</b>	5	6	6	5	6	<b>2</b>	6
BreastEW	30	7	14	3	3	18	7	4	<b>2</b>
CongressEW	16	5	6	5	9	2	4	2	<b>1</b>
Exactly	13	8	7	7	8	<b>6</b>	7	<b>6</b>	<b>6</b>
Exactly2	13	7	9	8	7	9	5	<b>1</b>	9
HeartEW	13	5	<b>4</b>	7	<b>4</b>	5	5	<b>4</b>	<b>4</b>
IonosphereEW	34	7	<b>2</b>	5	6	10	5	<b>2</b>	<b>2</b>
KrvskpEW	36	28	30	29	26	28	29	<b>13</b>	19
Lymphography	18	8	10	<b>4</b>	5	5	10	5	12
M-of-n	13	7	9	7	7	<b>6</b>	<b>6</b>	<b>6</b>	7
PenglungEW	325	<b>13</b>	14	110	71	59	65	15	134
SonarEW	60	42	28	24	32	18	33	<b>8</b>	38
SpectEW	22	17	14	17	5	13	17	<b>4</b>	11
Tic-tac-toe	9	5	7	5	4	6	5	<b>3</b>	5
Vote	16	6	10	6	6	4	6	<b>1</b>	<b>1</b>
WaveformEW	40	33	35	37	28	27	31	<b>20</b>	31
WineEW	13	10	3	6	4	4	<b>2</b>	<b>2</b>	4
Zoo	16	8	11	8	8	7	7	<b>5</b>	8
sum of ranks	696	218	218	294	239	232	250	103	300
avg of ranks	38.67	12.11	12.11	16.33	13.3	12.9	13.8	5.7	16.7
overall rank	7	2	2	8	4	3	5	1	6

of each method. MRFOv3 has achieved rank one *i.e.*, it selects lower number of features in comparison to other seven methods for most of the datasets used in this work and MRFOv3 performs in a consistent way *i.e.*, does not fluctuate in selecting lower number of features.

Observing both of the tables, on an average, MRFOv3 gives highest classification accuracy and selects the least number of features, and it performs consistently. So, amongst the proposed eight methods, MRFOv3 is considered as the best and we have used for comparing with different state-of-the-art methods.

#### 4 Comparison

In this section, we have compared the classification accuracy achieved and number of features selected using the proposed MRFOv3 approach with some methods reported recently in literature. We have compared our method with ALO based methods: bALO-QR, bALOS-QR, bALOV-QR, ALO, bALO-1, bALO-S, bALO-V, bALO-CE, bALOS-CE, bALOV-CE reported in [5]; GSA based methods: BGSA, HGSA reported in [32] and hybrid methods using GWO and WOA: HSGW, RSGW, ASGW reported in [33]. In this work we have used 18 UCI datasets and those were used in all the works reported in [33, 5, 32].

Figure 4 shows the classification accuracy achieved by the proposed MRFOv3 method and other methods for each dataset.

From Figure 4, it is clear that MRFOv3 achieves the highest accuracy in 14 out of 18 datasets. It is also worth mentioning that not only MRFOv3 has outperformed other methods in 77.8% of the datasets, but also it is able to achieve **100% accuracy** in 7 cases: CongressEW, Exactly, M-of-n, PenglungEW, Vote, WineEW and Zoo. Other four cases where MRFOv3 fails to achieve the highest accuracy are: BreastEW, Exactly2, IonosphereEW and SpectEW. For SpectEW, MRFOv3 has achieved second highest accuracy. From Figure 5, we can see that, in case of BreastEW, Exactly2 and IonosphereEW the number of selected features using MRFOv3 is significantly low. Figure 5 shows the number of features selected by the proposed MRFOv3 and other methods for each dataset used in this work.

From Figure 5, it is clear that MRFOv3 selects the lowest number of features in 15 out of 18 datasets *i.e.*, 83.3% cases. It can also be noted that for BreastEW, IonosphereEW, PenglungEW, SpectEW and WineEW datasets, the number of selected features by MRFOv3 is quite low compared to other methods. Now, the three cases where MRFOv3 fails to select the lowest number of features are: Exactly, M-of-n and WaveformEW. But for all these datasets, MRFOv3 has achieved the highest accuracy. Moreover, for Exactly and M-of-n, MRFOv3 is able to achieve 100% accuracy. So, observing Figure 4 and Figure 5, we can conclude that no other method outperforms MRFOv3 *w.r.t.* both in terms of classification accuracy and number of features even for a single dataset.

Taken together, this comparative study establishes that the proposed method outperforms the state-of-the-art methods considered here for comparison. Again, it justifies the *No Free Lunch* theorem by showing that there can always be a better algorithm for an optimization problem such as FS. It also shows the impact of transfer functions on the meta-heuristic algorithm used for FS. In Table 5, we have provided the Wilcoxon signed pair test [34] for all possible pair of the methods. Table 5 shows whether the difference between results (in terms of classification accuracy) of two algorithms is statistically significant or not according to Wilcoxon signed rank test. A p-value less than 0.05 is 1 (0 otherwise) which implies that difference between results is statistically significant, 0 means it is not.

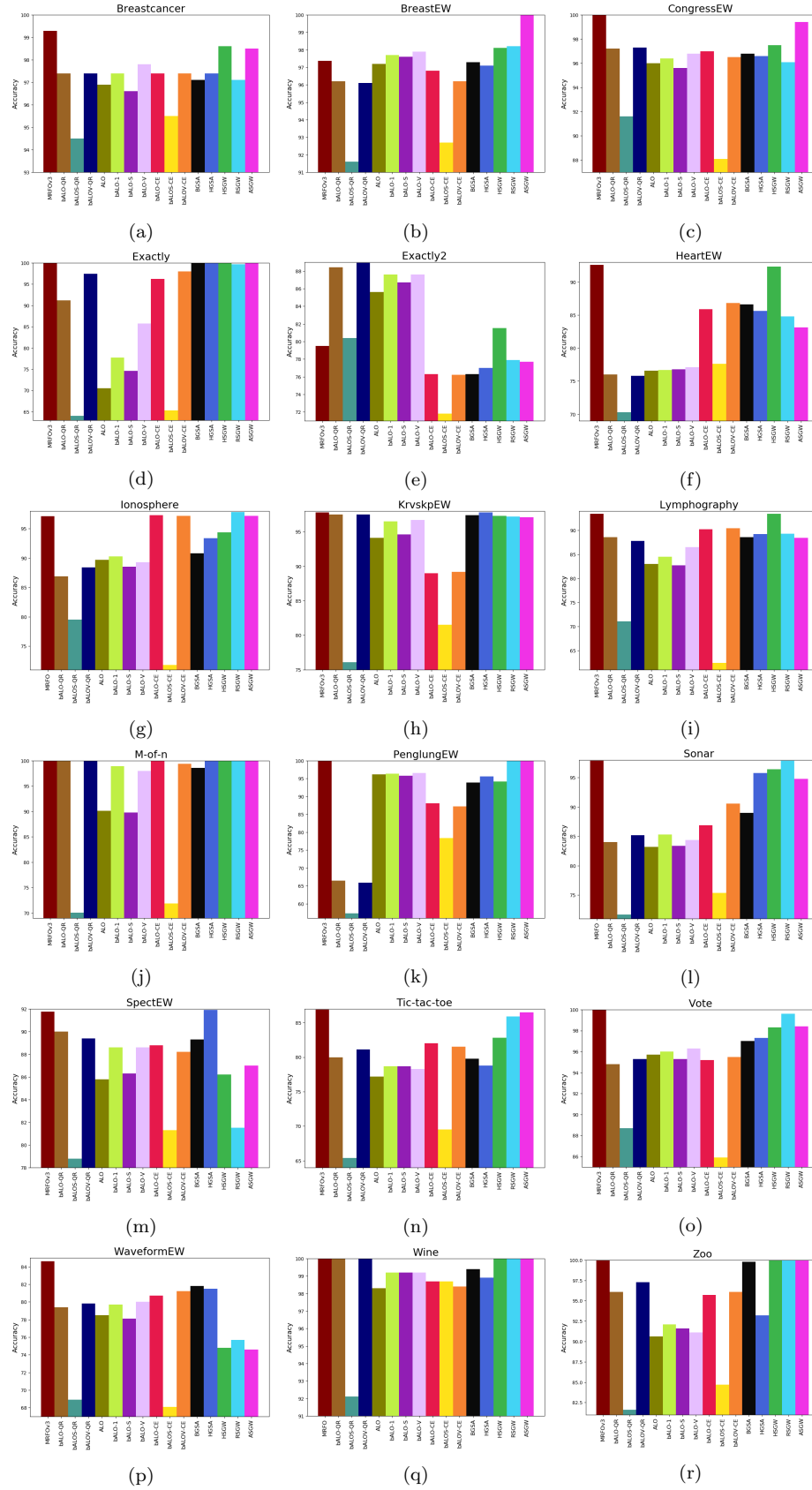


Fig. 4: Comparison between MRFO and state-of-the-art methods based on accuracy

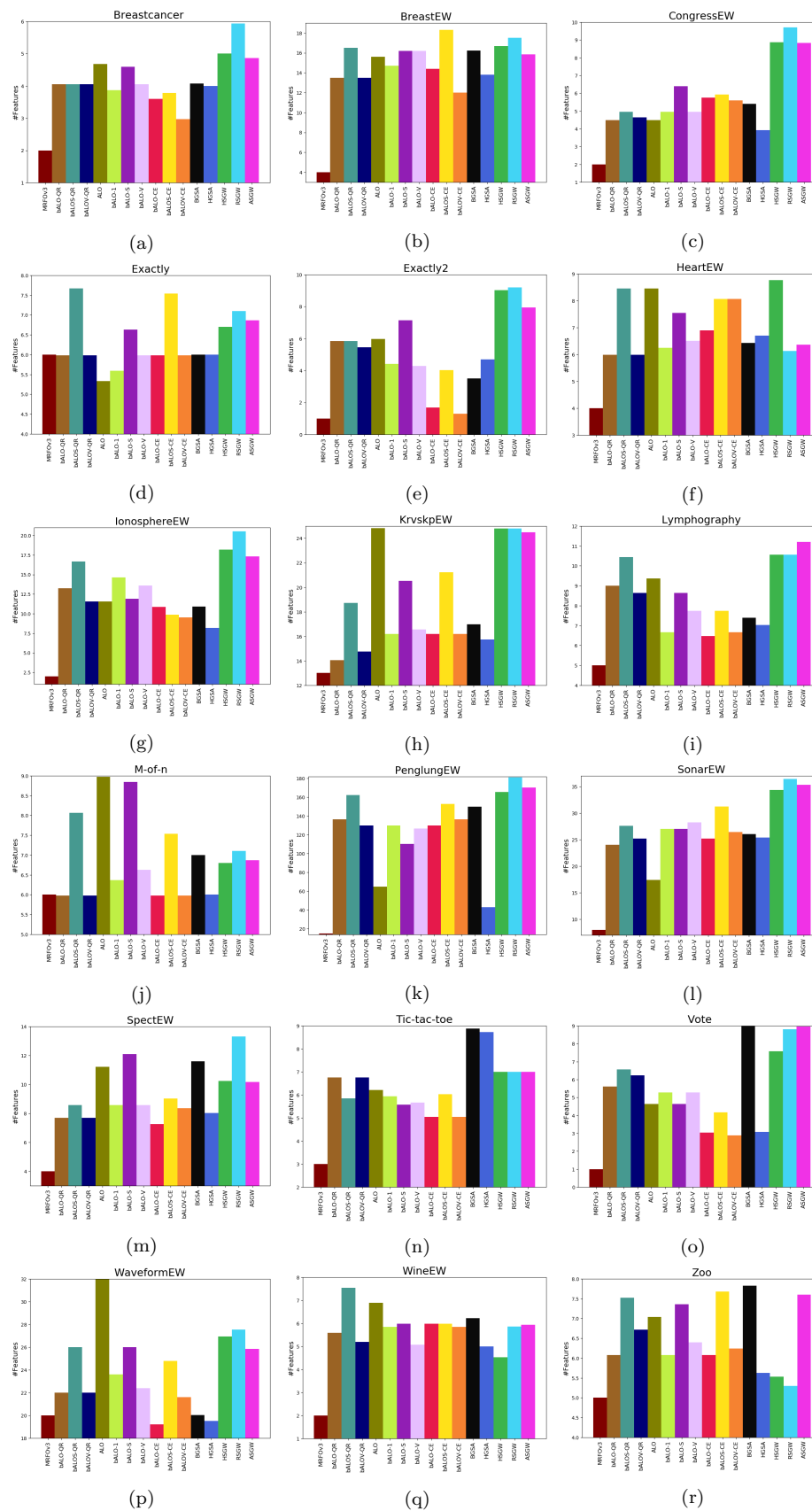


Fig. 5: Comparison between MRFO and state-of-the-art methods based on no. of selected features

Table 5: Wilcoxon signed rank test result for all pair of methods

	MRFO	bALO-QR	bALOS-QR	bALOV-QR	ALO	bALO-1	bALO-S	bALO-V	bALO-CE	bALOS-CE	bALOV-CE	BGSA	HGSA	HSGW	RSGW	ASGW
MRFO	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
bALO-QR	1	-	1	0	1	0	1	0	1	1	0	0	0	1	0	0
bALOS-QR	1	1	-	1	1	1	1	1	0	1	1	1	1	1	1	1
bALOV-QR	1	0	1	-	1	0	1	0	1	1	0	0	0	0	0	0
ALO	1	1	1	1	-	1	0	1	1	1	0	1	1	1	1	1
bALO-1	1	0	1	0	1	-	1	0	1	1	0	1	1	0	1	1
bALO-S	1	1	1	1	0	1	-	1	1	1	0	1	1	1	1	1
bALO-V	1	0	1	0	1	0	1	-	1	1	0	0	1	0	0	1
bALO-CE	1	0	1	0	0	0	0	0	1	1	0	0	0	1	0	1
bALOS-CE	1	1	0	1	1	1	1	1	-	1	1	1	1	1	1	1
bALOV-CE	1	0	1	0	1	0	0	0	1	1	-	0	0	1	0	1
BGSA	1	0	1	0	0	0	0	0	1	1	0	0	0	1	0	1
HGSA	1	0	1	0	1	1	1	0	1	1	0	-	0	1	0	0
HSGW	1	1	1	0	1	0	1	0	1	1	1	1	-	0	0	0
RSGW	1	0	1	0	1	1	1	0	1	1	0	0	0	-	0	0
ASGW	1	0	1	0	1	1	1	1	1	1	1	0	0	0	0	-

## 5 Conclusion and Future Work

FS is considered as an important pre-processing step in the domain of Machine Learning and Data Mining. In this article, we have proposed the binary version of MRFO for selection of optimal subset of features. Since MRFO is reported to be suitable for continuous search space problems, we have modified MRFO for the binary search space to fit it into the problem of FS using eight different transfer functions which belong to two different families: S-shaped and V-shaped. We have applied the eight binary versions of MRFO methods on 18 standard UCI datasets. We have shown the impact of transfer functions on the classification accuracy obtained and number of features selected.

We have considered both the classification accuracy and number of features selected while designing the fitness function. To measure the effectiveness of the proposed approaches, we have considered the best among the eight versions of MRFO and compared it with 16 recently proposed meta-heuristic FS approaches. The results show that MRFO outperforms the state-of-art methods in terms of both classification accuracy and number of selected features.

In future, we plan to propose an embedded FS method where binary MRFO can be united with a compatible filter method. We can also aim to hybridize MRFO with other meta-heuristic FS method. Besides, as the proposed FS version of MRFO is a general approach, so this can easily be applied to other standard pattern classification problems where usually large dimension of feature vectors are used.

**Conflict of interest:** We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

## References

- [1] Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.
- [2] Huan Liu and Hiroshi Motoda. *Feature selection for knowledge discovery and data mining*. Vol. 454. Springer Science & Business Media, 2012.
- [3] Soulib Ghosh et al. “A filter ensemble feature selection method for handwritten numeral recognition”. In: *EMR* (2016), p. 007213.
- [4] Qasem Al-Tashi et al. “Binary optimization using hybrid grey wolf optimization for feature selection”. In: *IEEE Access* 7 (2019), pp. 39496–39508.
- [5] Majdi M Mafarja and Seyedali Mirjalili. “Hybrid binary ant lion optimizer with rough set and approximate entropy reducts for feature selection”. In: *Soft Computing* 23.15 (2019), pp. 6249–6265.
- [6] Il-Seok Oh, Jin-Seon Lee, and Byung-Ro Moon. “Hybrid genetic algorithms for feature selection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.11 (Nov. 2004), pp. 1424–1437. DOI: 10.1109/tpami.2004.105. URL: <https://doi.org/10.1109/tpami.2004.105>.
- [7] Hao Chen et al. “A heuristic feature selection approach for text categorization by using chaos optimization and genetic algorithm”. In: *Mathematical problems in Engineering* 2013 (2013).

- [8] James Kennedy and Russell Eberhart. "Particle swarm optimization". In: *Proceedings of ICNN'95-International Conference on Neural Networks*. Vol. 4. IEEE. 1995, pp. 1942–1948.
- [9] B. Chakraborty. "Feature subset selection by particle swarm optimization with fuzzy fitness function". In: *2008 3rd International Conference on Intelligent System and Knowledge Engineering*. Vol. 1. Nov. 2008, pp. 1038–1042. DOI: 10.1109/ISKE.2008.4731082.
- [10] Sangwook Lee et al. "Modified binary particle swarm optimization". In: *Progress in Natural Science* 18.9 (Sept. 2008), pp. 1161–1166. DOI: 10.1016/j.pnsc.2008.03.018. URL: <https://doi.org/10.1016/j.pnsc.2008.03.018>.
- [11] Xiangyang Wang et al. "Feature selection based on rough sets and particle swarm optimization". In: *Pattern recognition letters* 28.4 (2007), pp. 459–471.
- [12] Marco Dorigo, Vittorio Maniezzo, and Alberto Coloni. "Ant system: optimization by a colony of cooperating agents". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 26.1 (1996), pp. 29–41.
- [13] Liangjun Ke, Zuren Feng, and Zhigang Ren. "An efficient ant colony optimization approach to attribute reduction in rough set theory". In: *Pattern Recognition Letters* 29.9 (2008), pp. 1351–1357.
- [14] Yumin Chen, Duoqian Miao, and Ruizhi Wang. "A rough set approach to feature selection based on ant colony optimization". In: *Pattern Recognition Letters* 31.3 (Feb. 2010), pp. 226–233. DOI: 10.1016/j.patrec.2009.10.013. URL: <https://doi.org/10.1016/j.patrec.2009.10.013>.
- [15] Dervis Karaboga. *An idea based on honey bee swarm for numerical optimization*. Tech. rep. Technical report-tr06, Erciyes university, engineering faculty, computer engineering department, 2005.
- [16] Mohammad Shokouhifar and Shima Sabet. "A hybrid approach for effective feature selection using neural networks and artificial bee colony optimization". In: *3rd international conference on machine vision (ICMV 2010)*. 2010, pp. 502–506.
- [17] Majdi Mafarja and Seyedali Mirjalili. "Whale optimization approaches for wrapper feature selection". In: *Applied Soft Computing* 62 (2018), pp. 441–453.
- [18] Esmat Rashedi, Hossein Nezamabadi-Pour, and Saeid Saryazdi. "BGSA: binary gravitational search algorithm". In: *Natural Computing* 9.3 (2010), pp. 727–745.
- [19] Eid Emary, Hossam M Zawbaa, and Aboul Ella Hassanien. "Binary grey wolf optimization approaches for feature selection". In: *Neurocomputing* 172 (2016), pp. 371–381. DOI: 10.1016/j.neucom.2015.06.083.
- [20] Majdi Mafarja et al. "S-shaped vs. V-shaped transfer functions for ant lion optimization algorithm in feature selection problem". In: *Proceedings of the international conference on future networks and distributed systems*. 2017, pp. 1–7.
- [21] Majdi Mafarja et al. "Binary grasshopper optimisation algorithm approaches for feature selection problems". In: *Expert Systems with Applications* 117 (2019), pp. 267–286.
- [22] Seyedali Mirjalili and Andrew Lewis. "S-shaped versus V-shaped transfer functions for binary particle swarm optimization". In: *Swarm and Evolutionary Computation* 9 (2013), pp. 1–14.



- [23] D. H. Wolpert and W. G. Macready. “No free lunch theorems for optimization”. In: *IEEE Transactions on Evolutionary Computation* 1.1 (Apr. 1997), pp. 67–82. ISSN: 1941-0026. DOI: 10.1109/4235.585893.
- [24] Weiguo Zhao, Zhenxing Zhang, and Liying Wang. “Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications”. In: *Engineering Applications of Artificial Intelligence* 87 (2020), p. 103300. DOI: 10.1016/j.engappai.2019.103300.
- [25] J. Rizzo. *Ocean Animals: Who’s Who in the Deep Blue*. Animals Series. National Geographic Society, 2016. ISBN: 9781426325069. URL: <https://books.google.co.in/books?id=Klu4DQAAQBAJ>.
- [26] Gene Helfman and George H Burgess. *Sharks*. JHU Press, 2014.
- [27] Henry Bryant Bigelow. “Sawfishes, guitarfishes, skates and rays”. In: *Sawfishes, guitarfishes, skates and rays, and chimaeroids* (1953), pp. 1–514.
- [28] Naomi S Altman. “An introduction to kernel and nearest-neighbor nonparametric regression”. In: *The American Statistician* 46.3 (1992), pp. 175–185. DOI: 10.1080/00031305.1992.10475879.
- [29] Manosij Ghosh et al. “Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods”. In: *Medical & Biological Engineering & Computing* 57.1 (Jan. 2019), pp. 159–176. ISSN: 1741-0444. DOI: 10.1007/s11517-018-1874-4.
- [30] Dheeru Dua and Casey Graff. *UCI Machine Learning Repository*. 2017. URL: <http://archive.ics.uci.edu/ml>.
- [31] Guido Van Rossum and Fred L Drake. *The python language reference manual*. Network Theory Ltd., 2011.
- [32] Mohammad Taradeh et al. “An evolutionary gravitational search-based feature selection”. In: *Information Sciences* 497 (2019), pp. 219–239.
- [33] Majdi Mafarja et al. “Efficient hybrid nature-inspired binary optimizers for feature selection”. In: *Cognitive Computation* (2019), pp. 1–26.
- [34] Frank Wilcoxon, SK Katti, and Roberta A Wilcox. “Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test”. In: *Selected tables in mathematical statistics* 1 (1970), pp. 171–259.