

Received May 13, 2020, accepted May 26, 2020, date of publication June 1, 2020, date of current version June 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2999093

# Hybrid of Harmony Search Algorithm and Ring Theory-Based Evolutionary Algorithm for Feature Selection

SHAMEEM AHMED<sup>1</sup>, KUSHAL KANTI GHOSH<sup>1</sup>, PAWAN KUMAR SINGH<sup>2</sup>, (Member, IEEE), ZONG WOO GEEM<sup>3</sup>, (Member, IEEE), AND RAM SARKAR<sup>1</sup>, (Senior Member, IEEE)

<sup>1</sup>Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India

<sup>2</sup>Department of Information Technology, Jadavpur University, Kolkata 700106, India

<sup>3</sup>Department of Energy IT, Gachon University, Seongnam 13120, South Korea

Corresponding author: Zong Woo Geem (geem@gachon.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant through the Korea Government (MSIT) under Grant 2020R1A2C1A01011131, and in part by the Energy Cloud Research and Development Program through the National Research Foundation of Korea (NRF) through the Ministry of Science, ICT, under Grant 2019M3F2A1073164.

**ABSTRACT** Feature Selection (FS) is an important pre-processing step in the fields of machine learning and data mining, which has a major impact on the performance of the corresponding learning models. The main goal of FS is to remove the irrelevant and redundant features, resulting in optimized time and space requirements along with enhanced performance of the learning model under consideration. Many meta-heuristic optimization techniques have been applied to solve FS problems because of its superiority over the traditional optimization approaches. Here, we have introduced a new hybrid meta-heuristic FS model based on a well-known meta-heuristic Harmony Search (HS) algorithm and a recently proposed Ring Theory based Evolutionary Algorithm (RTEA), which we have named as Ring Theory based Harmony Search (RTHS). Effectiveness of RTHS has been evaluated by applying it on 18 standard UCI datasets and comparing it with 10 state-of-the-art meta-heuristic FS methods. Obtained results prove the superiority of RTHS over the state-of-the-art methods considered here for comparison.

**INDEX TERMS** Ring theory based harmony search, feature selection, harmony search, ring theory based evolutionary algorithm, meta-heuristic, hybrid optimization, UCI datasets.

## I. INTRODUCTION

In this era of computer and technology, with the advancements in the fields of image processing, pattern recognition, financial analysis, business management, medical studies [1], [2] and many more, we have to deal with huge amount of data, whose dimensions are increasing everyday. This has a great impact on the performances of different algorithms used in the field of machine learning and data mining in terms of time and space requirements. There may be numerous features in a dataset, but not all of which are useful or important for a particular task. Feature selection (FS), a data pre-processing step, can be used to remove the irrelevant and redundant features [3], resulting in optimized time and space requirements. Basically, these redundant features act

as noise, and removal of those results in better performing ability of the corresponding machine learning or data mining algorithm [4]. There are two different categories of FS techniques based on evaluation criteria of features [3]: Filter and Wrapper. A filter method evaluates features based on pre-defined mathematical or statistical criteria, (e.g., Relief [5], Information Gain [6], Laplacian Score [7], Chi-Square [8], Fisher Score [9], etc.) and selects most important features according to that. Whereas, a wrapper method uses a learning algorithm to evaluate feature subsets and selects the optimum subset for the corresponding task [10]. Filter methods are relatively faster than wrapper methods as the former do not use learning algorithm, whereas the later, in general, achieves higher accuracy [11].

In recent times, the meta-heuristic methods have become popular in solving various optimization problems due its advantages over traditional optimization methods, such as

The associate editor coordinating the review of this manuscript and approving it for publication was Adnan Kavak<sup>1</sup>.

avoidance of local optima, non-derivative mechanism and flexibility [12]. Two important aspects of a meta-heuristic algorithm are [13]: exploration and exploitation. Exploration means the ability to search the solution space for new potential solution in each iteration avoiding local optima, and exploitation means finding a better solution in the neighborhood of the solution obtained so far. A good meta-heuristic algorithm has a characteristic of maintaining the balance between both exploration and exploitation phases.

In this work we have introduced a new hybrid meta-heuristic based on a well-known meta heuristic Harmony Search (HS) algorithm [14] and a recently proposed Ring Theory based Evolutionary Algorithm (RTEA) [15]. HS is inspired from an artificial phenomena, musical harmony. Just like musical performances seek a best state (fantastic harmony) which is determined by aesthetic estimation, HS seeks a best state (global optimum) determined by fitness or objective function. RTEA draws inspiration from algebraic theory on evolution process. Generally, there are two models followed for hybridizing meta-heuristic algorithms [16]: low level and high level. In low level hybridization, a function in a meta-heuristic is replaced by another meta-heuristic. In high level version, the base meta-heuristics are executed in sequence. We have hybridized HS and RTEA in a high level fashion, which follows the pipeline model, where each meta-heuristic optimization algorithm works on the output of previous optimization algorithm. To the best of our knowledge, this is the first time HS is hybridized with RTEA for solving FS problems. In a nutshell, the main contributions of this work are as follows:

- RTHS: A new FS method named Ring Theory based Harmony Search is introduced using a popular meta-heuristic HS and recently proposed RTEA.
- The proposed hybrid FS approach is evaluated on 18 standard UCI datasets [17] using K-nearest Neighbors (KNN), Random Forest, and Naive Bayes classifiers.
- The proposed FS approach is compared with 10 state-of-the-art meta-heuristic FS methods.

## II. LITERATURE REVIEW

Meta-heuristic algorithms are categorized differently in the literature: single solution based and population based [18], nature inspired and non-nature inspired [19], metaphor based and non-metaphor based [20]. These algorithms can also be divided into four different categories from ‘inspiration’ point of view [21]: Evolutionary, Swarm inspired, Physics based and Human behavior related.

FS is a binary optimization problem and most of the standard and popular optimization algorithms introduced so far in the literature have been applied for solving FS problems. Different applications of Genetic Algorithm (GA) for FS can be found in [22]–[25]. Particle Swarm Optimization (PSO) based FS methods can be found in [26]–[28]. Ant Colony Optimization (ACO) and Gravitational Search Algorithm (GSA) based FS methods can be found in [29]

and [30] respectively. Different types of meta-heuristics algorithms based on the source of inspiration are briefed here.

- Evolutionary algorithms are inspired from biological science. It uses the concept of mutation and crossover to evolve the randomly generated initial population over iterations and eliminates the worst solution in order to obtain a better solution. GA [31] is inspired from the biological evolution. *Mutation* and *crossover* are two of the most common operators used in GA. Mutation operates on a single solution and generally changes a feature randomly or following some pre-defined criterion. Crossover, on the other hand, operates on two parent solutions to produce two offspring, resulting in new and better solutions. Co-evolving algorithm [32], Differential evolution (DE) [33], Genetic Programming [34], Evolutionary programming [35], Bio-geography based optimizer [36], Stochastic fractal search [37] etc. are some of the well-known evolutionary algorithms.
- Swarm inspired algorithms mimic individual and social behavior of swarms, herds, schools, teams or any group of animals. Every individual has its own behavior, but the behavior of the accumulated individuals helps us to solve complex optimization problems. One of the most popular Swarm inspired algorithms is PSO [38], which is proposed by simulating social behavior, as representation of the movements of organisms in a bird flock or fish schools. This method performs the search to obtain the optimal solution through agents, referred to as particles. The movement of these particles is influenced by local optima in search space, and are updated if a better solution is found. Another approach of this category is ACO [39], inspired from the foraging method of ant species. Grey Wolf Optimizer (GWO) [12] is inspired by the leadership hierarchy and hunting mechanism of grey wolves. Four types of grey wolves such as  $\alpha$ ,  $\beta$ ,  $\delta$  and  $\omega$  are employed for replicating the leadership ranking. The three main steps of hunting, which are searching for prey, surrounding a prey, and attacking the prey are used as evolution. Ant Lion Optimizer (ALO) [40] proposed mimicking the hunting strategy of antlions in nature. The main steps of the hunting strategy includes the random walk of ants, building traps for ants by antlions, entrapment of ants in traps, catching preys, and rebuilding the traps. Some other methods belonging to this category are: Shuffled frog-leaping algorithm [41], Bacterial foraging [42], Artificial bee colony (ABC) [43], Firefly algorithm [44], Cuckoo search algorithm [45], Whale optimization algorithm (WOA) [46], Grasshopper optimization [47], Dolphin echolocation [48], Squirrel search algorithm [49] etc.
- Physics based algorithms are inspired by the physical processes in nature. The inspiring physical processes include music, metallurgy to mathematics, physics, chemistry, and complex dynamic systems. One of the oldest algorithms of this category is Simulated Annealing (SA) [50], developed by following the annealing [51]

process of metals present in metallurgy and materials sciences. Another popular method of this category is GSA [30], developed by following gravity and mass interaction. The search agents are considered as collection of masses, which interact with each other based on the Newton's gravitational law and the laws of motion. The used HS in the present work belongs to this category. Some other methods of this category are Self propelled particles [52], Black hole optimization [53], Charged system search [54], Sine Cosine algorithm [55], Multi-verse optimizer [56] etc.

- Human related algorithms are inspired from human behavior and interactions. Teaching-Learning-Based optimization [57] is a very popular method of this category, developed by following the enhancing procedure of class grade. Imperialistic competitive algorithm [58] is inspired from the human socio-political evolution process. Here, the populations are divided into two categories: colonies and imperialists states. The idea of this algorithm stands upon the competition among imperialists to take control of the colonies. At the end of the competition, only one imperialist stands out as victor and takes control over all the colonies, and the weak empires collapse. Some other methods of this category are: Society and civilization [59], League championship algorithm [60], Tug of war optimization [61], Volleyball premier league algorithm [62].

Nowadays, hybrid meta-heuristics algorithms have been used frequently for solving FS problems. Hybrid meta-heuristics have been proven to be an efficient approach to achieve better performance in various real-life problems [63]. In [64], the first hybrid meta-heuristic method is proposed for FS by combining GA with local search algorithm. The hybrid combination of Markov chain and SA is proposed in [65]. Memetic algorithm and Late acceptance hill climbing have been hybridized and used for FS for facial emotion recognition [66]. Spotted Hyena optimizer is combined with SA and used for FS on UCI datasets in [3]. Hybrid of GA and SA has been used for FS and applied on UCI datasets in [67]. In [68], Salp Swarm algorithm (SSA) is hybridized with Opposition Based Learning (OBL) and a local search method which is then applied on UCI datasets for FS. In [69], GA and PSO have been hybridized for FS and applied on Digital Mammogram datasets. In [70], the hybrid of GWO and PSO has been applied on UCI datasets for FS. Hybrid of PSO and GSA can be found in [71]. Hybrid of ACO and GA has been proposed in [72]. In [73], a hybrid version of DE and ABC for FS has been proposed and applied on UCI datasets.

Therefore, it is a time to raise a question. Why do we even need any new hybrid meta-heuristic algorithm for solving FS problems, as we have abundant of such algorithms? This question is quite logical and obvious. The question is best answered by a work reported in [74], which proposes *No Free Lunch* theorem and summarizes that there is not a single optimization algorithm which is capable for solving every type of optimization problem. With each new optimization

algorithm following any regular phenomena, researchers primarily focus to give some new facet to the algorithm where both exploration and exploitation will have a superior trade-off, so it ultimately gets away from the local optima and compasses to the global optima. Nevertheless, accomplishing these objectives are not simple, particularly in the event for which one needs to propose an algorithm that can be applicable to different domains. This practically motivates the researchers to come up with better optimization algorithms in comparison with the previously proposed algorithms. This is the inspiration which keeps the research active in the field of FS and motivates us to propose a new hybrid FS algorithm called RTHS algorithm based on HS method [14] and recently proposed RTEA [15].

### III. PRELIMINARIES

#### A. HARMONY SEARCH: AN OVERVIEW

HS algorithm [14] is inspired from an artificial phenomena, musical harmony. It transforms the qualitative improvisation process into quantitative optimization process with some well defined rules and thus turning the beauty and harmony of music into solution for several optimization problems. Just like the musical performers seek a fantastic harmony, this algorithm seeks the best state determined by fitness (objective) function. Just like the music can be improved for better aesthetic estimation, the fitness value can be improved in every iteration on order to find a better solution. Three important components of this algorithm are: harmony memory, pitch adjustment and randomization. HS algorithm does not require differential gradients, thus it can consider continuous as well as discontinuous functions. The basic steps used in HS algorithm are as follows:

Step 1: Randomly initialize initial population, here addressed as Harmony Memory (HM).

Step 2: Improvise a new harmony from it.

Step 3: If the new harmony is better than the worst harmony in HM, replace it.

Step 4: If stopping criterion is not met, goto Step 2.

Now, it assumes that all the parts of a global solution exist initially in HM, which is not necessarily the case always. So, to bring diversity, it utilizes Harmony Memory Considering Rate (HMCR),  $HMCR \in [0, 1]$ . If this rate is too low, then very few elite harmonies are selected and it may converge too slowly. On the other hand, if this rate is extremely high (near 1), the pitches in the HM are mostly used, and other ones are not explored well, not leading to good solutions. Therefore, typically, we use  $HMCR = [0.7, 0.95]$  [75]. Another factor called Pitch Adjustment Rate (PAR) is introduced by mimicking pitch adjustment procedure. This produces a new pitch by adding small random amount to the existing pitch.

Below, we have discussed application of HS algorithm as a meta-heuristic FS method in different domains as well as some of the modifications of it found in the literature. In [76], the authors have tried to generate a new solution vector that improves accuracy and convergence rate of HS algorithm and applied it on constrained functions (minimization of the

**Algorithm 1** Pseudocode of HS Algorithm

---

**Input:**  $popSize, maxIter, HMCR, PAR$   
**Output:** Best harmony  $H = (h_1, h_2, \dots, h_D)$   
 Randomly generate initial population  $HM(0)$   
**for**  $iter \leftarrow 1 \dots maxIter$  **do**  
   **if**  $random(0, 1) \geq HMCR$  **then**  
      $h^j(t+1) \leftarrow random(lb, ub) \forall j \in [1, D]$   
     // lb = lower bound, ub = upper bound  
   **else**  
     **for**  $j \leftarrow 1 \dots D$  **do**  
        $i \leftarrow random(1, popSize)$   
        $h^j(t+1) \leftarrow HM_i^j(t)$   
     **end for**  
   **end if**  
    $w$  be index of the worst harmony in  $HM(t)$   
   **if**  $fitness(h(t+1)) < fitness(HM_w(t))$  **then**  
      $HM_w(t) \leftarrow h(t+1)$   
   **end if**  
   **for**  $i \leftarrow 1 \dots popSize$  **do**  
     **if**  $random(0, 1) < PAR$  **then**  
        $HM_i^j(t+1) \leftarrow HM_i^j(t) + random(-1, 1) \times (ub - lb) \forall j \in [1, D]$   
     **end if**  
   **end for**  
**end for**

---

weight of the spring, Pressure vessel design, welded beam design etc.) and unconstrained functions. This work mainly discusses about the effect of constant parameters of HS algorithm. Another work reported in [77] describes a new version of HS algorithm for engineering optimization problems with continuous design variables. This algorithm has been applied on unconstrained function minimization problems (Rosenbrock function, Eason and Fenton’s gear train inertia function, Wood function, Powerwell quartic function etc.), constrained function minimization problems, structural engineering optimization problems (Pressure vessel design, Welded beam design etc.) and more. In [78], a new variant of HS algorithm has been proposed, which is known as Global-best Harmony Search (GHS). This algorithm takes the help of swarm intelligence to improve the performance of HS algorithm. It has been applied on Sphere function, Schwefel’s problem, Step function, Rosenbrock function, Rastrigin function etc. In [79] the authors present a cost minimization model for the the design of water distribution network. It is applied on five water distribution networks, which are Two-loop water distribution network, Hanoi water distribution network, New York City water distribution network, GoYang water distribution network, and BakRyun water distribution network. The work proposed in [75] reviews and analyzes the HS algorithm in the context of meta-heuristic algorithms.

**B. RING THEORY BASED EVOLUTIONARY ALGORITHM**

RTEA [15] is an recently proposed approach to solve combinatorial problems using algebraic theory. A global

exploration operator (R-GEO) and local development operator (R-LDO) are proposed using the addition, multiplication and inverse operation of direct product of the rings. Then by using R-GEO and R-LDO, new individuals are generated following a greedy strategy.

1) SOME PROPERTIES OF RING

*a: DEFINITION OF RING [80]:*

A ring  $(\mathcal{R}, +, \cdot)$  is a nonempty set  $\mathcal{R}$  together with two binary operations,  $+$  and  $\cdot$ , defined on  $\mathcal{R}$  such that:

- $\forall x, y \in \mathcal{R}, x + y = y + x$
- $\forall x, y, z \in \mathcal{R}, (x + y) + z = x + (y + z)$
- $\exists 0 \in \mathcal{R}$  such that  $\forall x \in \mathcal{R}, x + 0 = 0 + x = x$ , this 0 is called an additive identity
- $\forall x \in \mathcal{R} \exists y \in \mathcal{R}$  such that  $x + y = y + x = 0$ ,  $y$  is called the inverse of  $x$ , written as  $-x$
- $\forall x, y, z \in \mathcal{R}, (x \cdot y) \cdot z = x \cdot (y \cdot z)$  and
- $\forall x, y, z \in \mathcal{R}, x \cdot (y + z) = x \cdot y + x \cdot z; (y + z) \cdot x = y \cdot x + z \cdot x$

Let  $\mathbb{Z}_n = \{[0], [1], \dots, [n - 1]\}$  be a collection of residue classes of modulo  $n$ , where  $[i] = \{x \in \mathbb{Z} \mid x \equiv i \pmod{n}\}, 0 \leq i \leq n - 1$ , where  $n \geq 2$  and  $\mathbb{Z}$  is the set of all integers. We define two binary operations as following:

$$[i] \oplus [j] = [(i + j) \pmod{n}] \text{ and } [i] \odot [j] = [(i \cdot j) \pmod{n}] \forall [i], [j] \in \mathbb{Z}_n.$$

Hence,  $\mathbb{Z}_n$  is a ring with operations  $\oplus$  and  $\odot$ .

*b: DIRECT PRODUCT OF RINGS [80]:*

The direct product of rings is another ring, whose every element is an ordered  $m$ -tuple. If  $\mathcal{R}_i$  are rings,  $i \in \mathcal{I} = 1, 2, 3, \dots, m$ , then  $\prod_{i \in \mathcal{I}} \mathcal{R}_i = \mathcal{R}_1 \times \mathcal{R}_2 \times \dots \times \mathcal{R}_m$  is a ring with operations  $(a_1, a_2, \dots, a_m) \oplus (b_1, b_2, \dots, b_m) = (a_1 + b_1, a_2 + b_2, \dots, a_m + b_m)$  and  $(a_1, a_2, \dots, a_m) \odot (b_1, b_2, \dots, b_m) = (a_1 b_1, a_2 b_2, \dots, a_m b_m)$ .  $\prod_{i \in \mathcal{I}} \mathcal{R}_i$  is called as the direct product of  $\mathcal{R}_i, i \in \mathcal{I}$ .

2) RTEA OVERVIEW

So, we have  $\mathbb{Z}[r_1, r_2, \dots, r_D] = \{0, 1, \dots, r_1 - 1\} \times \{0, 1, \dots, r_2 - 1\} \times \dots \times \{0, 1, \dots, r_D - 1\}$ .

Now, let  $P_1 = (p_{11}, p_{12}, \dots, p_{1D}), P_2 = (p_{21}, p_{22}, \dots, p_{2D}), P_3 = (p_{31}, p_{32}, \dots, p_{3D})$  and  $P_4 = (p_{41}, p_{42}, \dots, p_{4D})$  be four different  $D$ -dimensional vector selected randomly  $\in \mathbb{Z}[r_1, r_2, \dots, r_D]$ . A new  $D$ -dimensional vector  $P = (p_1, p_2, \dots, p_D) \in \mathbb{Z}[r_1, r_2, \dots, r_D]$  is created from  $P_1, P_2, P_3, P_4$  using R-GEO, given by Equation 1.

$$p_k = \begin{cases} \{p_{1k} + p_{4k} \times [p_{2k} + (r_k - p_{3k})]\} \pmod{r_k} & \text{if } rndm(k) \leq 0.5; \\ \{p_{1k} + [p_{2k} + (r_k - p_{3k})]\} \pmod{r_k} & \end{cases} \quad (1)$$

R-GEO reflects global exploration ability and R-LDO which acts as the local search operator is given by 2.

Concisely R-LDO and R-GEO are used to generate new individuals, and a greedy strategy is used to select individuals to form the new generation.

**Algorithm 2** Pseudocode of R-LDO

---

**Input**  $X = (x_1, x_2, \dots, x_D) \in \mathbb{Z}[r_1, r_2, \dots, r_D]$   
 local search probability  $Prb_m \in (0, 0.5]$   
**Output**  $X = (x_1, x_2, \dots, x_D)$

**for**  $i = 1 \dots D$  **do**  
   **if**  $rndm1 < Prb_m$  **then**  
    **if**  $rndm2 < 0.5$  and  $x_i \neq 0$  **then**  
      $x_i \leftarrow r_i - x_i$ ;  
   **else**  
      $x_i \leftarrow rndm(\{0, 1, \dots, r_i - 1\} - \{x_i\})$ ;  
   **end if**  
**end if**  
**end for**

---

**IV. PROPOSED RTHS ALGORITHM**

Let us consider that we have the original feature set as  $\mathcal{F} = \{f_1, f_2, \dots, f_D\}$ , where  $n$  is the total number of attributes,  $\mathcal{C} = \{c_1, c_2, \dots, c_l\}$  be the class set, where  $l$  is the number of classes. Our target is to find a subset  $\mathcal{F}' = \{f'_1, f'_2, \dots, f'_o\}$ , where  $o < D$ ,  $\mathcal{F}' \subset \mathcal{F}$  and  $\mathcal{F}'$  has lower classification error rate than any other subset having same size or any proper subset of  $\mathcal{F}'$ . FS is a binary optimization problem, where 1 is indicates that the corresponding feature is selected whereas 0 indicates that the corresponding feature is discarded. So, we select the features having value 1 and discard those features having value 0. Our main goal is to decrease the number of 1's along with increasing the classification accuracy. In RTEA, which has been applied for solving Knapsack problem (KP) problem in the original paper, there are two cases. First, it is considered that the problem has a feasible solution, and second, it considers the opposite i.e. there is no feasible solution of the problem under consideration. For us, as we assume that there are potential solutions in the search space, so we consider this particular case [15] which is a binary problem in itself.  $\mathbb{Z}[r_1, r_2, \dots, r_D]$  in the previous section becomes  $\mathbb{Z}[2, 2, \dots, 2]$ , so we do not have to use any transfer function. Now, we need to consider another major aspect of FS, that is whenever we talk about achieving higher classification accuracy with lowest number of features by the algorithm, it can be observed that these two objectives are contradictory in nature. To get rid of this issue, classification error rate has been considered here. Using Equation 2, these two driving candidates have been combined.

$$\downarrow Fitness = \omega \zeta(\mathcal{F}') + (1 - \omega) \frac{|\mathcal{F}'|}{|\mathcal{F}|} \quad (2)$$

where  $\mathcal{F}'$  represents the set consisting of selected features,  $|\mathcal{F}'|$  represents number of selected features  $\zeta(\mathcal{F}')$  represents classification error rate of  $\mathcal{F}'$ ,  $|\mathcal{F}|$  is the original dimension of the dataset and  $\omega$  represents weight  $\in [0, 1]$ . HS algorithm finds the global optima by initiating HMCR and uses PAR to escape local optima. RTEA uses R-LDO for local exploration and R-GEO for global search. As FS is a binary optimization problem, R-LDO has been modified slightly as the population contains binary values. This modified version of R-LDO is

given by Equation 3.

$$\begin{aligned} &\text{for } d \leftarrow 1 \dots D \\ &\quad X^d(t+1) = 1 - X^d(t) \text{ if } random[0, 1] \leq Prb_m \\ &\text{endfor} \end{aligned} \quad (3)$$

where  $X^d(t)$  represents the  $d^{th}$  dimension in  $t^{th}$  iteration of the solution  $X$ .

The direct product of rings is a ring too. As we utilize the  $\{0, 1\}$  version of RTEA, the obtained results will be in  $\{0, 1\}$ . With the help of R-GEO, RTEA performs the global exploration, and with R-LDO, it performs the exploitation. So, there is a balance between exploration phase and exploitation phase. The flowchart of our proposed method is given in Figure 1.

The HS algorithm depends on the value of HMCR and PAR in order to find out the global best solution. HMCR [76] mainly helps to find the 'promising' areas where optimal solution (global best) may lie i.e., it ensures exploration. On the other hand, PAR [76] helps to properly search the areas already discovered i.e., it ensures exploitation. Therefore, chances of obtaining the best solution depend on this HMCR and PAR, but their values need to be defined at the beginning of the HS algorithm. Appropriate setting of these values is needed for convergence of this algorithm. Again the same set of values may not to produce optimum results for all problems. In [76], the authors have tried to address this issue with an iterative approach, but then precision becomes another problem to deal with. With increase in precision by 10% margin, the time requirement increases exponentially. Here, we have tried to solve this problem using a completely different approach. We have enhanced both the exploration and exploitation operations of HS algorithm by a recently proposed meta-heuristic RTEA. In each iteration, we not only 'enhance' a particular harmony by finding its fitter neighbor using the R-LDO operator but also we excel the 'improvise a new harmony' step used in the HS algorithm with the help of R-GEO operator. This implies that we are able to reduce the dependency of the HS algorithm on initial values of HMCR and PAR to obtain the global optima. R-GEO, aids in the exploration process by considering randomly 4 harmonies present in current HM (population) and checking whether is it possible to form a better harmony or not. R-LDO tries to improve the exploitation by finding any better neighbor for a particular harmony. The results shown in subsection V-C and section VI validate this claim.

**V. RESULTS AND DISCUSSION**

We have evaluated the proposed RTHS algorithm using three popular classifiers: KNN [81], Random Forest [82], Naive Bayes [83] for assessing the effectiveness of the same. For each dataset, 80% of the instances are used to train the model and the rest 20% are used for testing. We have applied the FS methods on the trained data, and determined the features which are useful. These features form the optimal feature subset. From test data, only those features are selected and the

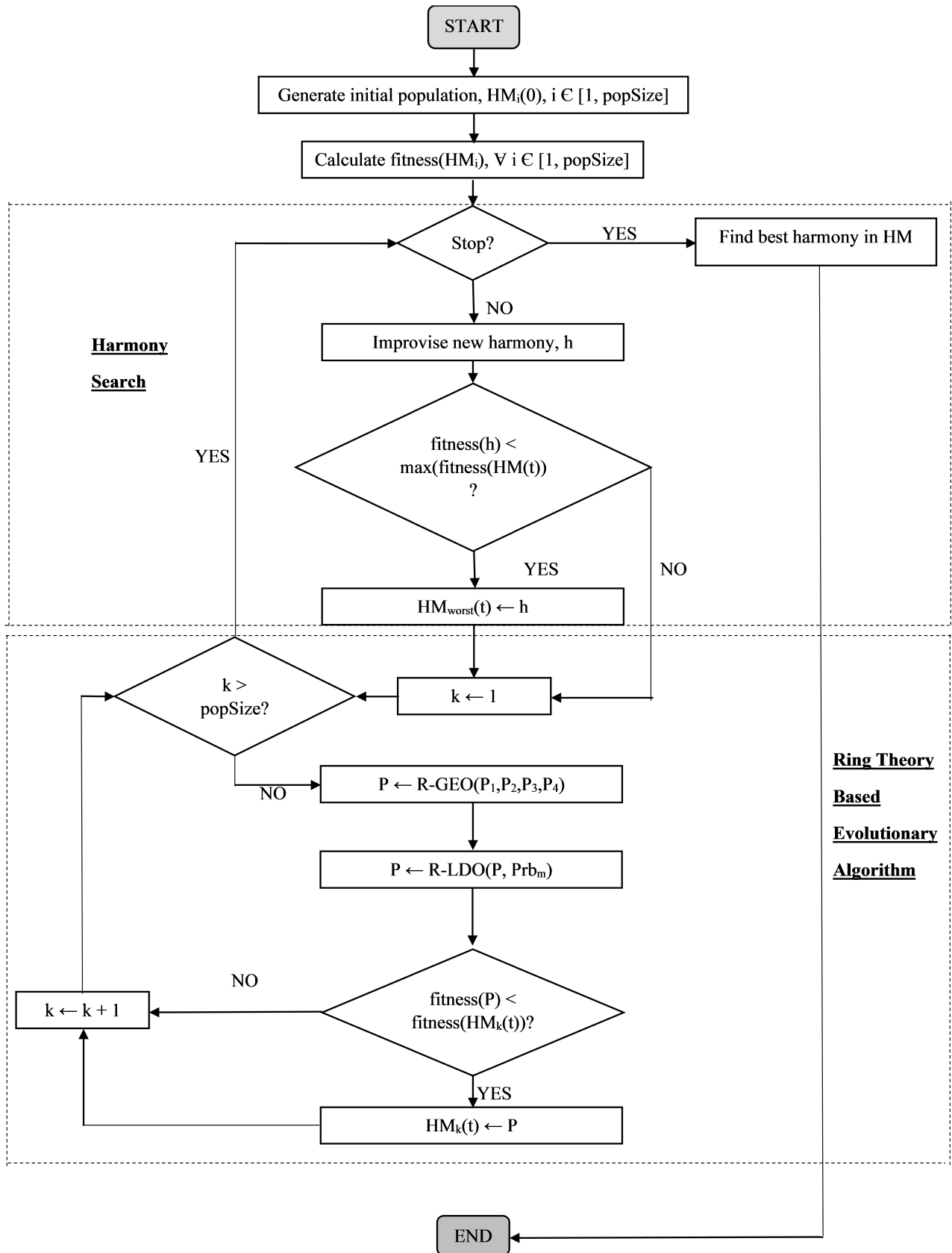


FIGURE 1. Pictorial representation of the proposed FS model named as RTHS: hybrid of Harmony Search (HS) algorithm and Ring Theory based Evolutionary Algorithm (RTEA).

**TABLE 1.** Description of the datasets used to evaluate the present FS method.

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	4	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

test classification accuracy is measured based on these using the above mentioned classifiers. The proposed FS method is implemented using Python3 [84] whereas the graphs are plotted using Matplotlib [85].

#### A. DATASET DESCRIPTION

In order to examine the performance of HS, RTEA and RTHS algorithms, 18 standard UCI datasets [17] have been considered. These datasets are selected from various backgrounds. The description of these datasets is presented in Table 1, which shows that there are 14 bi-class and 4 multi-class datasets. The datasets are diverse in terms of number of features and instances. These mixture helps us in establishing the robustness of the proposed FS method.

#### B. PARAMETER TUNING

When we talk about multi-agent evolutionary algorithm, both population size and maximum number of iterations play a significant role for characterizing the behavior of one agent's learning ability from others' experiences and the step-by-step evolution of the agents respectively. For finding the appropriate values of these two parameters, we have performed experiments by varying one parameter *w.r.t.* the other.

Figure 2 shows the effect of the size of the population on achieved classification accuracy using the proposed FS method. Considering that the time requirement increases with increase in population size and the effect of the same on classification accuracy, we have fixed the values of population size to 20 as the standard population size, the maximum number of iterations to 30, HMCR to 0.8, PAR to 0.2 and  $Prb_m$  to 0.005 (as suggested in [15]) for all further experiments. Figure 3 shows the best value of the fitness function in each iteration.

#### C. DISCUSSION

This section reports the results of the proposed FS method called RTHS for the datasets mentioned in Section V-A.

Table 2, Table 3 and Table 4 describe the results obtained by the proposed RTHS algorithm, as compared to RT and HS algorithms using KNN, Random Forest and Naive Bayes classifiers respectively. It can be concluded that the proposed RTHS algorithm performs the best over UCI datasets with KNN classifier. Besides, KNN is also widely used in the literature for FS purpose on UCI datasets [10], [86], [87]. Hence, for further experiments and analysis, we have used only KNN classifier with  $K = 5$ .

From Table 2, it is quite evident that the proposed FS method has performed significantly well. The RTHS algorithm has produced accuracy  $> 90\%$  (83.33%) for 15 datasets. Whereas it has achieved 100% accuracy for 9 (50%) datasets: CongressEW, Exactly, Ionosphere, M-of-n, PenglungEW, Sonar, Vote, WineEW, and Zoo, which is quite impressive. Out of 18 datasets, it achieves the highest accuracy in almost 17 cases (94.44%). Comparing these results with HS algorithm, it can be observed that the RTHS algorithm performs better than HS algorithm in exactly 15 cases and in 3 cases they produce equivalent result. Comparing RTHS algorithm with RTEA, it is found that in 12 cases, they achieve the same result, and but in 5 cases, the RTHS algorithm outperforms RTEA. However, in the case of Tic-tac-toe dataset, the RTHS algorithm could not outperform RTEA in terms of classification accuracy.

Now, if we focus on the number of features selected, then it is quite clear that the RTHS algorithm selects the least number of features in exactly 15 cases (83.33%). Careful observation of Table 2 reveals that the RTHS algorithm outperforms both HS algorithm and RTEA with significant margin in most of the cases. It outperforms HS algorithm in 11 cases and gives equivalent result in 4 cases (CongressEW, Exactly2, HeartEW and M-of-n datasets). In case of Exactly, Ionosphere and Tic-tac-toe datasets, HS algorithm is able to produce better results than RTHS algorithm. But RTEA is unable to show better performance than RTHS algorithm

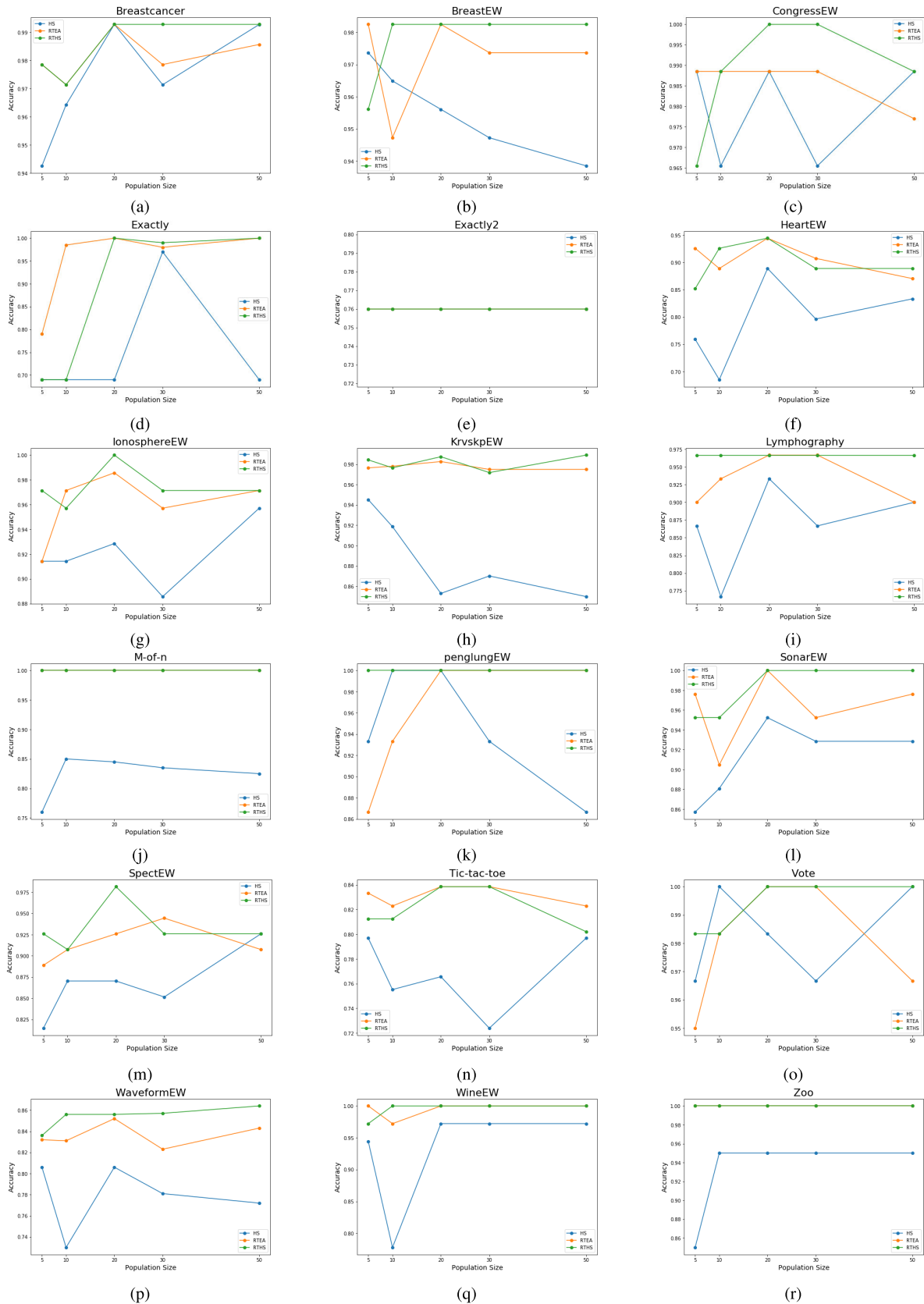
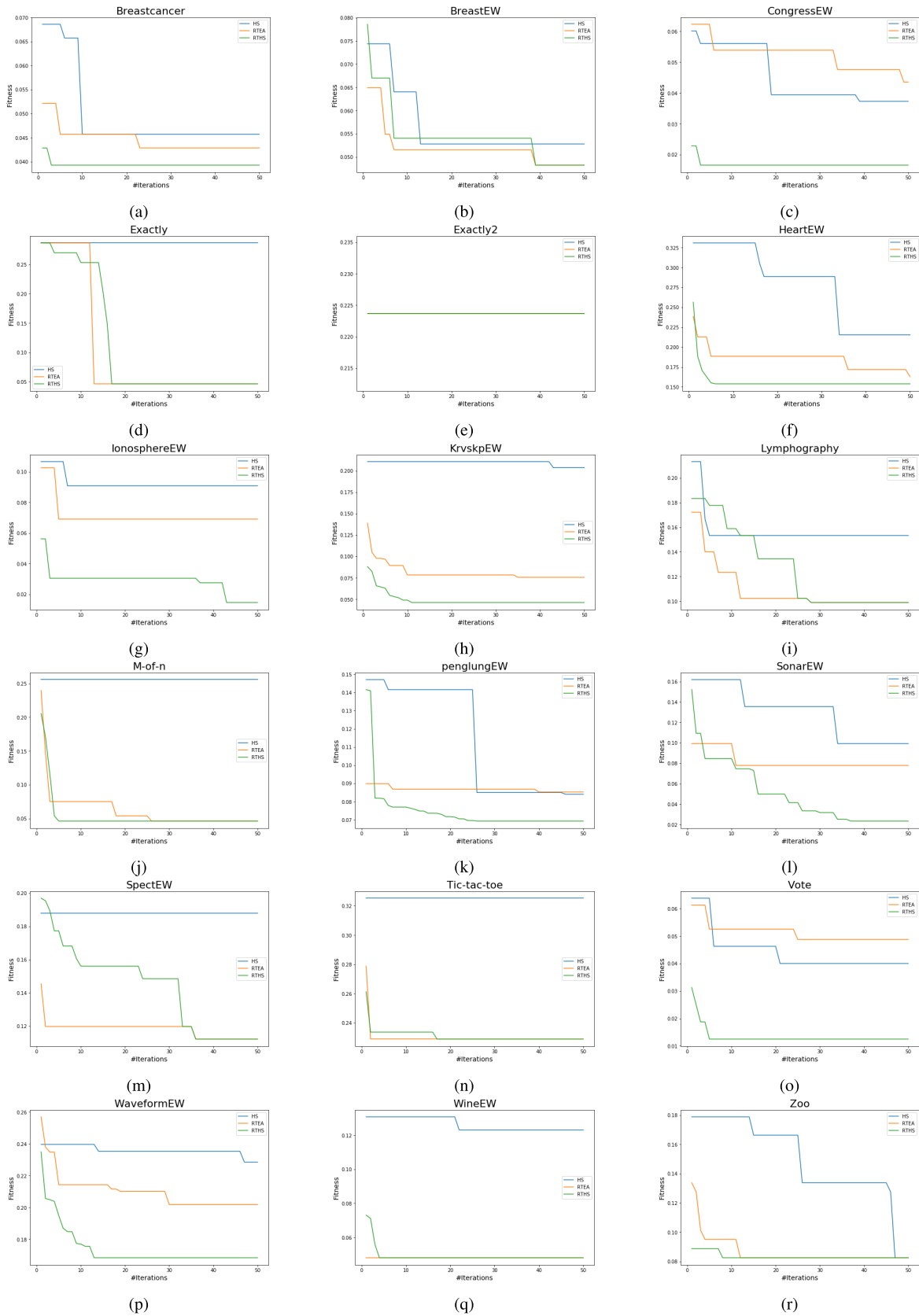


FIGURE 2. Effect of population size on classification accuracy for 18 UCI datasets using HS algorithm, RTEA and RTHS algorithm.

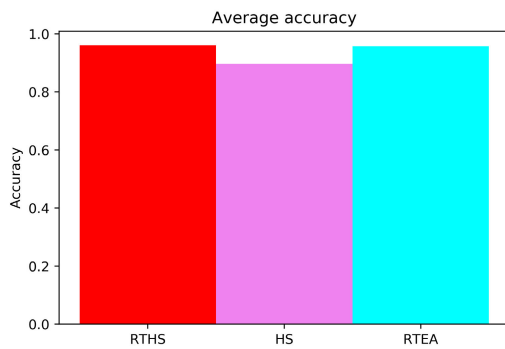




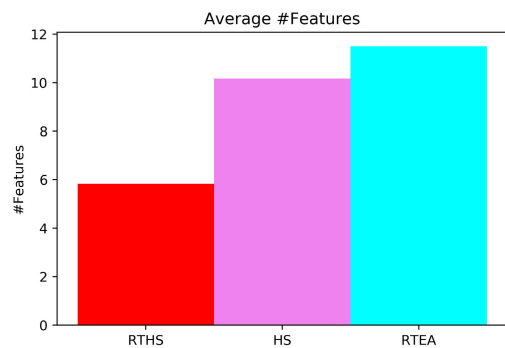
**FIGURE 3.** Best fitness value in each iteration for 18 UCI datasets using HS algorithm, RTEA and RTHS algorithm.

**TABLE 2.** Performance comparison of HS algorithm, RTEA and RTHS algorithm in terms of accuracy and number of selected features using KNN classifier.

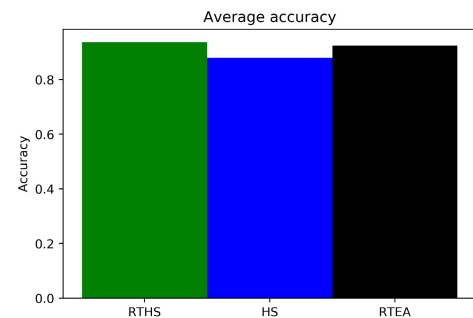
Sl. No.	Dataset	Original		HS		RTEA		RTHS	
		Accuracy	#Features	Accuracy	#Features	Accuracy	#Features	Accuracy	#Features
1	Breastcancer	96	9	<b>99.28</b>	5	<b>99.28</b>	<b>3</b>	<b>99.28</b>	<b>3</b>
2	BreastEW	92.63	30	95.61	6	98.24	<b>2</b>	<b>98.25</b>	<b>2</b>
3	CongressEW	92.18	16	98.85	<b>4</b>	98.85	7	<b>100</b>	<b>4</b>
4	Exatly	72.3	13	69	<b>1</b>	<b>100</b>	6	<b>100</b>	6
5	Exactly2	73.3	13	<b>76</b>	<b>1</b>	<b>76</b>	<b>1</b>	<b>76</b>	<b>1</b>
6	HeartEW	68.15	13	88.89	<b>5</b>	<b>94.4</b>	6	<b>94.4</b>	<b>5</b>
7	IonosphereEW	83.43	34	92.85	<b>3</b>	97.14	8	<b>100</b>	8
8	KrvskpEW	96.1	36	85.29	10	97.96	17	<b>98.75</b>	<b>9</b>
9	Lymphography	81.33	18	93.33	10	<b>96.67</b>	7	<b>96.67</b>	<b>4</b>
10	M-of-n	87.4	13	84.5	<b>6</b>	<b>100</b>	<b>6</b>	<b>100</b>	<b>6</b>
11	PenglungEW	81.33	325	<b>100</b>	77	<b>100</b>	80	<b>100</b>	<b>13</b>
12	SonarEW	80.95	60	95.23	16	<b>100</b>	16	<b>100</b>	<b>8</b>
13	SpectEW	82.22	22	87.03	8	96.29	10	<b>98.15</b>	<b>4</b>
14	Tic-tac-toe	81.1	9	76.56	<b>3</b>	<b>85.41</b>	6	83.8	9
15	Vote	92.33	16	98.33	3	<b>100</b>	4	<b>100</b>	<b>2</b>
16	WaveformEW	81.44	40	80.6	16	<b>84.1</b>	20	<b>84.1</b>	<b>15</b>
17	WineEW	66.67	13	97.22	3	<b>100</b>	3	<b>100</b>	<b>2</b>
18	Zoo	87	16	95	6	<b>100</b>	5	<b>100</b>	<b>4</b>



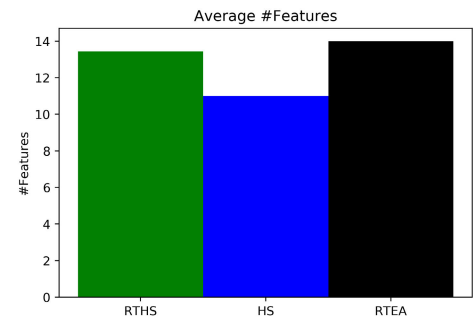
**FIGURE 4.** Graphical comparison of average accuracies achieved by RTHS algorithm, HS algorithm and RTEA using KNN classifier.



**FIGURE 5.** Graphical comparison of average #features selected by RTHS algorithm, HS algorithm and RTEA using KNN classifier.



**FIGURE 6.** Graphical comparison of average accuracies attained by RTHS algorithm, HS algorithm and RTEA using Random Forest classifier.



**FIGURE 7.** Graphical comparison of average #features selected by RTHS, HS and RTEA algorithms using Random Forest classifier.

for none of the cases, though it provides equivalent result in 4 cases (Breastcancer, BreastEW, Exactly2 and M-of-n datasets).

To visualize these results, bar charts showing the comparison of both average accuracies achieved and number of features utilized by the three algorithms, namely, RTHS algorithm, HS algorithm and RTEA, using KNN classifier have been plotted in Figure 4 and Figure 5 respectively. These bar charts show that all the three FS methods perform almost

equivalently in terms of average accuracy, but in terms of average number of features selected, the RTHS algorithm has the upper hand.

From Table 3, it is quite clear that the proposed FS method performs much better than both RTEA and HS algorithm. For 14 datasets (77.78%), it has achieved > 90% accuracy. Whereas for 7 datasets (38.89%), it has achieved an 100% accuracy. These datasets are: BreastEW, IonosphereEW, M-of-n, PenglungEW, Vote, WineEW and Zoo. For all the 18 datasets, it achieves the highest accuracy along with ties

**TABLE 3.** Performance comparison of HS algorithm, RTEA and RTHS algorithm in terms of accuracy and number of selected features using Random Forest classifier.

Sl no.	Dataset	Original		HS		RTEA		RTHS	
		Accuracy	#Features	Accuracy	#Features	Accuracy	#Features	Accuracy	#Features
1	Breastcancer	97.14	9	97.86	<b>2</b>	98.57	<b>2</b>	<b>98.86</b>	<b>2</b>
2	BreastEW	92.3	30	97.37	<b>4</b>	99.12	6	<b>100</b>	7
3	CongressEW	91.4	16	96.55	4	96.55	<b>2</b>	<b>97</b>	<b>2</b>
4	Exactly	65	13	<b>69</b>	<b>1</b>	<b>69</b>	<b>1</b>	<b>69</b>	<b>1</b>
5	Exactly2	72	13	<b>76</b>	<b>1</b>	<b>76</b>	<b>1</b>	<b>76</b>	<b>1</b>
6	HeartEW	67.77	13	85.19	5	85.19	<b>4</b>	<b>87.04</b>	5
7	IonosphereEW	81.84	34	91.43	12	<b>100</b>	<b>9</b>	<b>100</b>	<b>9</b>
8	KrvskpEW	94.27	36	95.46	13	<b>98.59</b>	<b>14</b>	<b>98.59</b>	<b>14</b>
9	Lymphography	83.33	18	90	<b>5</b>	90	9	<b>93.33</b>	9
10	M-of-n	74.64	13	80	<b>4</b>	<b>100</b>	8	<b>100</b>	6
11	PenglungEW	80.43	325	93.33	<b>95</b>	86.67	122	<b>100</b>	121
12	SonarEW	76.26	60	85.71	20	<b>90.48</b>	<b>17</b>	<b>90.48</b>	20
13	SpectEW	81.37	22	87.04	<b>6</b>	<b>96.3</b>	10	<b>96.3</b>	7
14	Tic-tac-toe	81.58	9	68.75	<b>1</b>	<b>93.75</b>	9	<b>93.75</b>	8
15	Vote	88.34	16	95	<b>1</b>	98.33	6	<b>100</b>	5
16	WaveformEW	78.5	40	79	<b>14</b>	84.6	23	<b>86.2</b>	16
17	WineEW	67.7	13	97.22	4	<b>100</b>	<b>3</b>	<b>100</b>	<b>3</b>
18	Zoo	90	16	<b>100</b>	<b>6</b>	<b>100</b>	<b>6</b>	<b>100</b>	<b>6</b>

**TABLE 4.** Performance comparison of HS algorithm, RTEA and RTHS algorithm in terms of accuracy and number of selected features using Naive Bayes classifier.

Sl no.	Dataset	Original		HS		RTEA		RTHS	
		Accuracy	#Features	Accuracy	#Features	Accuracy	#Features	Accuracy	#Features
1	Breastcancer	91.43	9	97.85	<b>2</b>	98.57	<b>2</b>	<b>98.86</b>	<b>2</b>
2	BreastEW	98.245	30	99.28	<b>6</b>	98.24	10	<b>100</b>	<b>6</b>
3	CongressEW	97.7	16	97.7	4	<b>98.85</b>	2	<b>98.85</b>	<b>1</b>
4	Exactly	<b>69</b>	13	<b>69</b>	<b>1</b>	<b>69</b>	<b>1</b>	<b>69</b>	<b>1</b>
5	Exactly2	<b>76</b>	13	<b>76</b>	<b>1</b>	<b>76</b>	<b>1</b>	<b>76</b>	<b>1</b>
6	HeartEW	92.59	13	92.59	<b>4</b>	<b>94.44</b>	7	<b>94.44</b>	7
7	IonosphereEW	94.29	34	94.28	<b>3</b>	97.14	5	<b>100</b>	9
8	KrvskpEW	66.2	36	80.75	17	93.58	14	<b>96.4</b>	<b>11</b>
9	Lymphography	93.33	18	93.33	9	96.67	<b>5</b>	<b>100</b>	6
10	M-of-n	96.5	13	83.5	<b>5</b>	<b>98</b>	9	<b>98</b>	6
11	PenglungEW	66.7	325	86.67	28	86.67	66	<b>100</b>	<b>29</b>
12	SonarEW	88	60	88	9	90.48	25	<b>97.61</b>	<b>6</b>
13	SpectEW	66.67	22	88.89	9	92.59	6	<b>94.44</b>	<b>4</b>
14	Tic-tac-toe	75	9	69.79	3	<b>72.91</b>	5	<b>72.91</b>	5
15	Vote	<b>100</b>	16	98.33	<b>2</b>	<b>100</b>	3	<b>100</b>	<b>2</b>
16	WaveformEW	81.5	40	82	16	83.6	18	<b>84.3</b>	<b>10</b>
17	WineEW	<b>100</b>	13	<b>100</b>	5	<b>100</b>	<b>3</b>	<b>100</b>	<b>3</b>
18	Zoo	<b>100</b>	16	<b>100</b>	5	<b>100</b>	4	<b>100</b>	<b>3</b>

in case of 10 datasets with RTEA and 3 datasets with HS algorithm. This proves that using Random Forest classifier, the RTHS algorithm still shows superior performance. Figure 6 illustrates the comparison of average accuracies achieved by the three algorithms using Random Forest classifier.

Now, if we consider the number of selected features, we can see that RTHS algorithm is slightly behind the HS algorithm and RTEA. For 8 datasets (44.44%), it selects the least number of features along with ties in 8 datasets with RTEA and in 4 datasets with HS algorithm. From this perspective, the RTHS algorithm may seem to be inefficient, but looking at the achieved classification accuracy, it can be said that the RTHS algorithm produces far better results than others. Figure 7 shows the comparison of number of features chosen by the three algorithms using Random Forest classifier. By looking at Figure 7, it is obvious that in terms

of average number of features selected, the RTHS algorithm selects least number of features than RTEA.

Going through Table 4, it can be easily said that the RTHS algorithm has achieved the highest accuracy for all the 18 UCI datasets along with ties for 4 datasets with HS algorithm and 9 datasets with RTEA. In case of 14 datasets (77.78%), the proposed FS method has achieved accuracy > 90%. On the other hand, for 7 datasets (38.89%), which are: BreastEW, IonosphereEW, Lymphography, PenglungEW, Vote, WineEW and Zoo, the proposed RTHS algorithm has produced an 100% accuracy. These prove that the RTHS algorithm produces the best results in terms of classification accuracy. Figure 8 illustrates the comparison of average accuracies achieved by the three algorithms using Naive Bayes classifier. Now, considering the number of selected features, Figure 9 tells us a lot. For 13 datasets (72.22%), the RTHS algorithm selects the least number of features along with ties

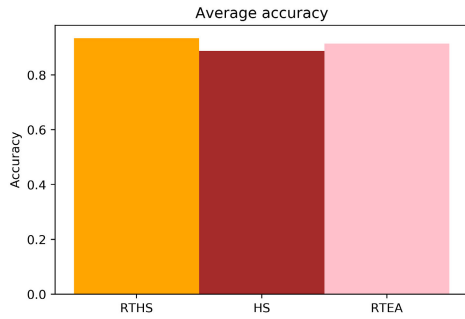


FIGURE 8. Graphical comparison of average accuracies attained by RTHS algorithm, HS algorithm and RTEA using Naive Bayes classifier.

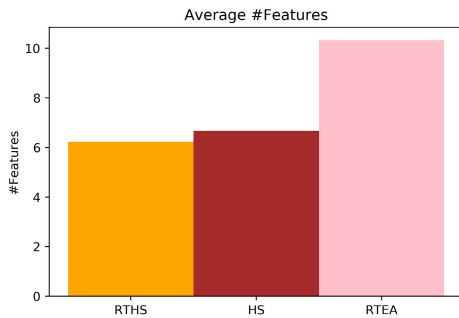


FIGURE 9. Graphical comparison of average #features selected by RTHS algorithm, HS algorithm and RTEA using Naive Bayes classifier.

in 5 datasets with HS algorithm and ties in 4 datasets with RTEA.

Based on the above discussion as well as observing Table 2, we can conclude that the RTHS algorithm significantly helps both HS algorithm and RTEA to explore different parts of the search space and to achieve better solution in terms of both achieved classification accuracy and selected number of features.

During this comparison, if the classification accuracy is found to be higher for any particular classifier, then that classifier is considered as better and if the classification accuracies remain the same for any two classifiers, then the number of selected features is used as the deciding factor to break the tie. While observing Table 2, Table 3 and Table 4, we can also conclude that the KNN classifier produces better results than both Random Forest and Naive Bayes classifiers.

Comparing the results obtained from both KNN and Random Forest classifiers, it can be observed that the KNN classifier performs better than Random Forest for exactly 13 datasets (72.22%) considering both the achieved classification accuracy and number of selected features. In case of BreastEW, Tic-tac-toe and WaveformEW datasets, Random Forest has upper hand over KNN classifier. Again, in case of Exactly2 and M-of-n datasets, two classifiers produce the same classification accuracy and choose the same number of features.

Furthermore, comparing the results obtained by KNN and Naive Bayes classifiers, it can be inferred that the KNN classifier outperforms the Naive Bayes classifier for exactly 12 datasets (66.67%) while considering the above mentioned

TABLE 5. Parameters setting for state-of-the-art FS methods used here for comparison.

Algorithm	Parameters
BGA	popSize = 8 maxIter = 70 Mutation and Crossover rate = 0.8
BPSO	popSize = 8 maxIter = 70 inertia factor = 0.1 individual-best acceleration factor = 0.1
BALO	popSize = 8 maxIter = 200
BGSA	popSize = 8 maxIter = 20 G=1 $\alpha = 20$
HSGW, RSGW, ASGW	popSize = 20 maxIter = 100 a in GWO and WOA = [2 0] $\alpha$ in GSA = 20 G0 in GSA = 100
WOASAT-2	popSize = 10 maxIter = 100 a in WOA = [2 0] selection pressure = 0.5
BGWOPSO	popSize = 10 maxIter = 100 c1 = c2 = c3 = 0.5 w = 0.5+rand()/2 l ∈ [1, 0]
WOA-CM	popSize = 10 maxIter = 100 a in WOA = [2 0]

factors. For 4 datasets, namely, BreastEW, Lymphography, WaveformEW and Zoo, the Naive Bayes classifier produces the best results. For Exactly2 and Vote datasets, the two classifiers result in a tie.

We can finally conclude from the above discussion that the KNN classifier produces better results than Random Forest and Naive Bayes classifier. So, for performing comparison of the proposed RTHS algorithm with state-of-the-art FS methods, we consider KNN classifier only.

## VI. COMPARISON

To check the the effectiveness of the proposed FS method, we have compared it with 10 state-of-the-art FS methods that include four popular meta-heuristic FS methods namely, GA, PSO, ALO, and GSA, and six hybrid meta-heuristic FS methods namely, Serial grey-whale optimizer (HSGW), random switching grey-whale optimizer (RSGW), adaptive switching grey-whale optimizer (ASGW), WOASAT-2, BGWOPSO and WOA-CM. HSGW, RSGW, and ASGW are three different FS strategies formed by hybridizing both GWO and WOA methods [86]. WOASAT-2 method [10] is hybrid of WOA and SA methods. BGWOPSO method [70] is developed by hybridizing GWO and PSO methods. In WOA-CM method [88], the performance of WOA is enhanced by using both crossover and mutation. The values of the control parameters of these FS methods are described in Table 5.

Table 6 shows the performance of the RTHS algorithm in terms of classification accuracy. From Table 6, it can

**TABLE 6.** Comparison of classification accuracies obtained by proposed FS method with 10 state-of-the-art FS methods over 18 UCI datasets.

Dataset	RTHS	BGA	BPSO	BALO	BGSA	HSGW	RSGW	ASGW	WOASAT-2	BGWOPSO	WOA-CM
Breastcancer	<b>0.9928</b>	0.9743	0.9629	0.974	0.9686	0.986	0.971	0.985	0.97	0.98	0.968
BreastEW	<b>0.9825</b>	0.9754	0.9719	0.974	0.9544	0.981	0.982	<b>1</b>	0.98	0.97	0.971
CongressEW	<b>1</b>	0.9679	0.9633	0.981	0.9633	0.975	0.961	0.994	0.98	0.98	0.956
Exactly	<b>1</b>	<b>1</b>	<b>1</b>	0.965	0.994	<b>1</b>	0.997	0.999	<b>1</b>	<b>1</b>	<b>1</b>
Exactly2	0.76	0.77	0.768	0.762	0.77	<b>0.815</b>	0.779	0.777	0.75	0.76	0.742
HeartEW	<b>0.9444</b>	0.8741	0.837	0.838	0.8296	0.923	0.848	0.831	0.85	0.85	0.807
IonosphereEW	<b>1</b>	0.9489	0.9489	0.904	0.9432	0.944	0.978	0.972	0.96	0.95	0.926
KrvskpEW	<b>0.9875</b>	0.985	0.9731	0.973	0.9549	0.973	0.972	0.971	0.98	0.97	0.972
Lymphography	<b>0.9667</b>	0.8378	0.8919	0.917	0.8649	0.934	0.893	0.884	0.89	0.92	0.852
M-of-n	<b>1</b>	<b>1</b>	<b>1</b>	0.967	0.994	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0.991
PenglungEW	<b>1</b>	0.9189	0.9189	0.827	0.8333	0.942	<b>1</b>	<b>1</b>	0.94	0.96	0.792
SonarEW	<b>1</b>	0.9904	0.9423	0.845	0.9135	0.964	0.979	0.948	0.97	0.96	0.919
SpectEW	<b>0.9815</b>	0.8955	0.8881	0.899	0.8433	0.862	0.815	0.87	0.88	0.88	0.866
Tic-tac-toe	0.838	0.7996	0.7996	0.783	0.7766	0.828	0.859	<b>0.865</b>	0.79	0.81	0.785
Vote	<b>1</b>	0.9733	0.96	0.972	0.96	0.983	0.996	0.984	0.97	0.97	0.939
WaveformEW	<b>0.841</b>	0.7836	0.756	0.797	0.7344	0.748	0.757	0.746	0.76	0.8	0.753
WineEW	<b>1</b>	0.9888	0.9775	0.972	0.9775	<b>1</b>	<b>1</b>	<b>1</b>	0.99	<b>1</b>	0.959
Zoo	<b>1</b>	0.902	0.9608	0.98	0.9804	<b>1</b>	<b>1</b>	<b>1</b>	0.97	<b>1</b>	0.98
<b>Average Rank</b>	1.5	4.38	5.72	5.67	7.33	3.61	3.77	3.77	4.61	4.22	7.44
<b>Assigned Rank</b>	<b>1</b>	5	8	7	9	2	3	3	6	4	10

**TABLE 7.** Comparison of number of selected features by the proposed FS method with 10 state-of-the-art FS methods over 18 UCI datasets.

Dataset	RTHS	BGA	BPSO	BALO	BGSA	HSGW	RSGW	ASGW	WOASAT-2	BGWOPSO	WOA-CM
Breastcancer	<b>3</b>	4	4	4.7	4	5	5.933	4.867	4.2	4.4	4.302
BreastEW	<b>2</b>	8	9	13.85	10	16.667	17.5	15.833	11.6	13.6	15.81
CongressEW	4	<b>2</b>	3	6.65	4	8.867	9.7	8.833	6.4	4.4	6.448
Exactly	6	6	6	5.75	<b>4</b>	6.7	7.1	6.867	6	6	6.045
Exactly2	<b>1</b>	<b>1</b>	<b>1</b>	1.5	<b>1</b>	9.033	9.2	7.933	2.8	1.6	5.252
HeartEW	5	5	<b>3</b>	8.6	<b>3</b>	8.767	6.133	6.367	5.4	5.8	6.995
IonosphereEW	8	<b>7</b>	<b>7</b>	11.75	9	18.167	20.5	17.3	12.8	13	14.416
KrvskpEW	9	11	12	16.15	14	24.8	24.8	24.5	18.4	15.8	18.54
Lymphography	<b>4</b>	5	5	7.35	6	10.567	10.567	11.2	7.2	9.2	8.208
M-of-n	6	6	6	6	<b>5</b>	6.8	7.1	6.867	6	6	6.006
PenglungEW	<b>13</b>	84	130	133.1	140	165.333	181.2	170.3	127.4	130.8	128.05
SonarEW	<b>8</b>	19	22	26.6	24	34.3	36.433	35.3	26.4	31.2	35.64
SpectEW	4	5	6	7.65	5	10.233	13.3	10.167	9.4	8.4	8.052
Tic-tac-toe	9	5	6	5	<b>4</b>	7	7	7	6	5.2	6.903
Vote	<b>2</b>	5	<b>3</b>	6.6	4	7.567	8.8	8.967	5.2	3.4	7.408
WaveformEW	15	15	15	20.5	<b>14</b>	26.933	27.533	25.833	20.6	14.2	25.4
WineEW	<b>2</b>	4	5	5.4	4	4.533	5.867	5.933	6.4	6	6.799
Zoo	<b>4</b>	<b>4</b>	5	5.7	6	5.533	5.3	7.6	5.6	6.8	6
<b>Average Rank</b>	1.8	2.05	2.55	5	2.94	7.4	8.33	7.7	4.72	4.88	6.27
<b>Assigned Rank</b>	<b>1</b>	2	3	7	4	9	11	10	5	6	8

be observed that the RTHS algorithm performs the best in 16 cases (88.9%), which is quite impressive. In case of Tic-tac-toe dataset, it stands *third* position following ASGW and RSGW methods.

It is worth mentioning that the RTHS algorithm outperforms BGA method completely in 15 cases, and ties in 2 cases. For Exactly2 dataset, BGA performs better than RTHS algorithm. On the other hand, the RTHS algorithm outperforms BPSO in 15 cases and ties in 2 cases. For Exactly2 dataset, BPSO performs slightly better than RTHS algorithm. However, the proposed algorithm completely outperforms both BALO and BGSA methods for exactly 17 datasets except Exactly2.

Again, the HSGW method outperforms RTHS algorithm in case of Exactly2 dataset and ties in 4 cases. The RTHS algorithm outperforms HSGW in 13 cases. With respect to RSGW method, the RTHS algorithm has 4 ties, 12 wins

and 2 losses. ASGW ties with RTHS method in 4 cases whereas the former method outperforms the latter in 3 cases. Furthermore, it loses to RTHS algorithm for the rest 11 cases. In comparison with WOASAT2 method, the RTHS algorithm has 2 ties and 16 wins. Again, the BGWOPSO method has 3 ties and 15 loses when comparing with RTHS algorithm. Both WOA-CM and RTHS tie in case of Exactly dataset whereas for the rest 17 datasets, the proposed RTHS algorithm wins.

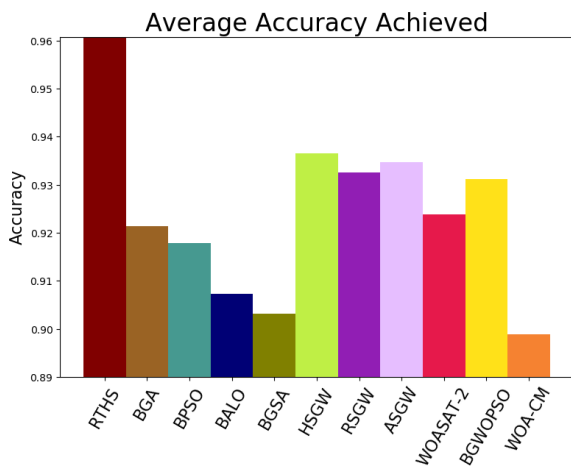
Table 7 shows the performance of the RTHS algorithm *w.r.t.* number of selected features. The RTHS algorithm selects the lowest number of features in case of 8 datasets. In case of Exactly2 dataset, the RTHS algorithm fails to achieve the highest accuracy but in terms of number of selected features, it performs the best followed by BGA, BPSO and BGSA methods. So, considering both Table 6 and Table 7, it can be concluded that the RTHS algorithm

**TABLE 8.** *p*-values produced by the Wilcoxon rank-sum test for the classification accuracies achieved by the RTHS algorithm, HS algorithm, and RTEA as compared to themselves as well as 10 other state-of-the-art FS methods considered here.

	RTHS	HS	RTEA	BGA	BPSO	BALO	BGSA	HSGW	RSGW	ASGW	WOASAT-2	BGWOPSO	WOA-CM
RTHS	-	0.000	0.000	0.001	0.001	0.000	0.000	0.008	0.011	0.048	0.000	0.001	0.000
HS	0.000	-	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RTEA	0.000	0.001	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

**TABLE 9.** Comparison of the proposed method with the top – 3 FS methods in terms of precision, recall, f-score and roc\_auc\_score for 18 UCI datasets.

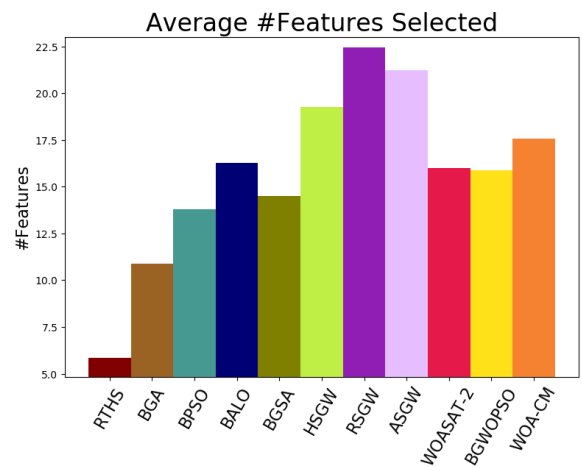
Dataset	HSGW				RSGW				ASGW				RTHS			
	precision	recall	fscore	roc_auc	precision	recall	fscore	roc_auc	precision	recall	fscore	roc_auc	precision	recall	fscore	roc_auc
Breastcancer	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0.9552	0.9678	0.9609	0.9925	0.98	0.9891	0.9843	0.9981	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
BreastEW	0.9539	0.9335	0.9422	0.9726	0.9672	0.9573	0.9619	0.9904	0.936	0.9315	0.9337	0.9487	<b>0.9741</b>	<b>0.9692</b>	<b>0.9716</b>	<b>0.9959</b>
CongressEW	0.9512	0.9512	0.9512	0.977	0.9662	0.9604	0.9632	0.9655	0.9714	<b>0.9815</b>	<b>0.9759</b>	0.9854	<b>0.9612</b>	0.9663	0.9636	<b>0.992</b>
Exactly	0.8143	0.8086	0.8113	0.8624	0.7779	0.7779	0.7779	0.8463	<b>0.9006</b>	<b>0.8734</b>	<b>0.8852</b>	<b>0.9704</b>	<b>0.9006</b>	<b>0.8734</b>	<b>0.8852</b>	<b>0.9704</b>
Exactly2	0.5505	0.5192	0.5007	0.5914	<b>0.6241</b>	<b>0.5718</b>	<b>0.5766</b>	<b>0.698</b>	0.5609	0.534	0.5291	0.6407	<b>0.6241</b>	<b>0.5718</b>	<b>0.5766</b>	<b>0.698</b>
HearEW	0.5907	0.5917	0.5903	0.6785	0.739	0.7417	0.7393	0.7792	<b>0.85</b>	<b>0.85</b>	<b>0.85</b>	<b>0.9</b>	<b>0.85</b>	<b>0.85</b>	0.82	<b>0.9</b>
IonosphereEW	0.9018	0.78	0.8045	0.9293	<b>0.9327</b>	0.86	0.8825	<b>0.9942</b>	0.9245	0.84	0.8639	0.9267	0.9252	<b>0.8889</b>	<b>0.9029</b>	0.948
KrvskpEW	0.9646	0.9634	0.9639	0.9925	0.9765	0.9764	0.9765	<b>0.9965</b>	0.9704	0.97	0.9702	0.9843	<b>0.9825</b>	<b>0.9834</b>	<b>0.9828</b>	0.9895
Lymphography	0.4234	0.4409	0.432	0.8923	0.3946	0.4006	0.3944	0.874	0.4039	0.4069	0.401	0.7794	<b>0.6132</b>	<b>0.668</b>	<b>0.6403</b>	<b>0.9279</b>
M-of-n	0.9379	0.9329	0.9353	0.9774	0.8527	0.8443	0.8481	0.935	0.9922	0.9865	0.9892	0.9996	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
PenglungEW	<b>0.881</b>	<b>0.9048</b>	<b>0.8762</b>	0.9512	0.7929	0.7857	0.7732	<b>0.998</b>	0.7571	0.8095	0.7651	0.964	<b>0.881</b>	<b>0.9048</b>	<b>0.8762</b>	0.9243
SonarEW	0.8413	0.8249	0.8286	0.9073	0.8612	0.8513	0.8542	0.881	0.8333	0.8295	0.8309	0.8936	<b>0.9273</b>	<b>0.9302</b>	<b>0.9282</b>	<b>0.9611</b>
SpectEW	0.7778	0.7378	0.7545	0.8076	0.7473	0.6924	0.7126	<b>0.8763</b>	0.7976	<b>0.8171</b>	<b>0.8066</b>	0.8594	<b>0.7999</b>	0.704	0.7333	0.7378
Tic-tac-toe	0.6864	0.645	0.6515	0.747	0.8995	<b>0.8204</b>	<b>0.8433</b>	<b>0.8703</b>	<b>0.9139</b>	0.806	0.8325	0.8057	0.812	0.7223	0.7394	0.7833
Vote	0.951	0.943	0.9467	0.97	0.9259	0.9459	0.9314	0.9636	<b>0.9647</b>	<b>0.9647</b>	<b>0.9647</b>	0.9959	<b>0.9647</b>	<b>0.9647</b>	<b>0.9647</b>	<b>0.998</b>
WaveformEW	0.811	0.8111	0.811	0.9365	0.8202	0.8195	0.8187	0.9508	<b>0.8367</b>	0.8361	0.836	0.951	0.8354	<b>0.8365</b>	<b>0.838</b>	<b>0.965</b>
WineEW	0.9697	0.9762	0.9718	0.9978	0.9744	0.9762	0.9743	1	0.9697	0.9762	0.9718	0.9794	<b>0.976</b>	<b>0.976</b>	<b>0.976</b>	0.99
Zoo	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0.9841	0.9286	0.944	0.9605	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>



**FIGURE 10.** Graphical comparison of average accuracies achieved by the proposed method and 10 state-of-the-art FS methods over the 18 UCI datasets.

performs the best *w.r.t.* the 10 state-of-the-art FS methods considered here for comparison. Figure 10 shows the graphical comparison of the average classification accuracies achieved by RTHS algorithm and 10 state-of-the-art FS methods. It clearly shows that the RTHS algorithm attains the highest average classification accuracy as compared to other state-of-the-art FS methods. Figure 11 depicts the graphical comparison of the average number of features selected by RTHS algorithm as well as 10 state-of-the-art FS methods. Figure 11 implies that the proposed RTHS algorithm selects the lowest number of features as compared to other state-of-the-art FS methods.

To determine the statistical significance of the RTHS algorithm, the Wilcoxon rank-sum test [89] has been performed. It is a non-parametric statistical test where pairwise comparison of the proposed FS method is done *w.r.t.* 10 other



**FIGURE 11.** Graphical comparison of average number of features selected by the proposed method and 10 state-of-the-art FS methods over the 18 UCI datasets.

state-of-the-art FS methods. Here, the null hypothesis states that the two sets of results follow the same distribution. If the distribution of two results are statistically different, then the generated *p*-value obtained from the test statistics will be  $< 0.05$ , when the test is performed at 0.05% significance level. If this condition is satisfied, then the null hypothesis is rejected. From the test results provided in Table 8, it can be concluded that the proposed RTHS algorithm is statistically significant *w.r.t.* 10 other state-of-the-art FS methods.

Observing both Table 6 and Table 7, we can conclude that the top – 3 FS methods which perform better as compared to the proposed RTHS algorithm are: HSGW, RSGW, and ASGW. Table 9 provides the detailed performance results attained by the proposed FS method as well as these top – 3 methods in terms of four statistical popular measures such

as precision, recall, f-score, and roc\_auc\_score for all the 18 UCI datasets considered in the present work.

## VII. CONCLUSION

In this paper, we have proposed a hybrid meta-heuristic method for FS, named as RTHS, based on a well-known meta-heuristic called HS algorithm and a recently proposed meta-heuristic called RTEA. The proposed RTHS has been applied on 18 standard UCI datasets and compared with 10 state-of-art meta-heuristic and hybrid meta-heuristic FS approaches. The obtained results prove the superiority of RTHS over other methods. Hence, we can say RTHS can be considered as a competent method for solving FS problems. Observing the results meticulously, we can understand that RTEA helps HS algorithm to overcome its limitations in terms of exploration and exploitation (as discussed in section IV). But, there may be some cases, where it may fail to find global optima as per the requirement of the problem, which is in accordance with *No Free Lunch* theorem [74]. At the same time, we need to perform a bit exhaustive experiments to find the ideal value of the parameters used in this algorithm for different problems, which is another shortcoming of the proposed work. As future scope of the work, we can apply the proposed RTHS on other popular and interesting research problems, like facial emotion recognition, musical symbol recognition, handwritten or printed script recognition, etc. RTHS can be applied on high dimensional datasets like gene expression data. It would be interesting to hybridize this with other recently proposed or classical meta-heuristic algorithms.

## REFERENCES

- [1] L. Xiong, R.-S. Chen, X. Zhou, and C. Jing, "Multi-feature fusion and selection method for an improved particle swarm optimization," *J. Ambient Intell. Humanized Comput.*, pp. 1–10, Dec. 2019, doi: 10.1007/s12652-019-01624-4.
- [2] H. Faris, A. A. Heidari, A. M. Al-Zoubi, M. Mafarja, I. Aljarah, M. Eshtay, and S. Mirjalili, "Time-varying hierarchical chains of salps with random weight networks for feature selection," *Expert Syst. Appl.*, vol. 140, Feb. 2020, Art. no. 112898, doi: 10.1016/j.eswa.2019.112898.
- [3] H. Jia, J. Li, W. Song, X. Peng, C. Lang, and Y. Li, "Spotted hyena optimization algorithm with simulated annealing for feature selection," *IEEE Access*, vol. 7, pp. 71943–71962, 2019, doi: 10.1109/access.2019.2919991.
- [4] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.
- [5] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Machine Learning Proceedings*. Amsterdam, The Netherlands: Elsevier, 1992, pp. 249–256, doi: 10.1016/b978-1-55860-247-2.50037-1.
- [6] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.
- [7] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Proc. 18th Int. Conf. Neural Inf. Process. Syst.* Cambridge, MA, USA: MIT Press, 2005, p. 507–514.
- [8] Z. Zheng, X. Wu, and R. Srihari, "Feature selection for text categorization on imbalanced data," *ACM SIGKDD Explor. Newslett.*, vol. 6, no. 1, p. 80, Jun. 2004, doi: 10.1145/1007730.1007741.
- [9] Q. Gu, Z. Li, and J. Han, "Generalized Fisher score for feature selection," in *Proc. 27th Conf. Uncertainty Artif. Intell. (UAI)*. Arlington, VA, USA: AUAI Press, 2011, p. 266–273.
- [10] M. M. Mafarja and S. Mirjalili, "Hybrid whale optimization algorithm with simulated annealing for feature selection," *Neurocomputing*, vol. 260, pp. 302–312, Oct. 2017, doi: 10.1016/j.neucom.2017.04.053.
- [11] M. Ghosh, S. Adhikary, K. K. Ghosh, A. Sardar, S. Begum, and R. Sarkar, "Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods," *Med. Biol. Eng. Comput.*, vol. 57, no. 1, pp. 159–176, Aug. 2018, doi: 10.1007/s11517-018-1874-4.
- [12] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014, doi: 10.1016/j.advengsoft.2013.12.007.
- [13] B. Chatterjee, T. Bhattacharyya, K. K. Ghosh, P. K. Singh, Z. W. Geem, and R. Sarkar, "Late acceptance hill climbing based social ski driver algorithm for feature selection," *IEEE Access*, vol. 8, pp. 75393–75408, 2020, doi: 10.1109/ACCESS.2020.2988157.
- [14] Z. Woo Geem, J. Hoon Kim, and G. V. Loganathan, "A new heuristic optimization algorithm: Harmony search," *Simulation*, vol. 76, no. 2, pp. 60–68, Feb. 2001, doi: 10.1177/003754970107600201.
- [15] Y. He, X. Wang, and S. Gao, "Ring theory-based evolutionary algorithm and its application to D{0–1} KP," *Appl. Soft Comput.*, vol. 77, pp. 714–722, Apr. 2019, doi: 10.1016/j.asoc.2019.01.049.
- [16] K. K. Ghosh, S. Ahmed, P. K. Singh, Z. W. Geem, and R. Sarkar, "Improved binary sailfish optimizer based on adaptive  $\beta$ -hill climbing for feature selection," *IEEE Access*, vol. 8, pp. 83548–83560, 2020, doi: 10.1109/access.2020.2991543.
- [17] D. Dua and C. Graff. (2017). *UCI Machine Learning Repository*. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [18] M. Gendreau and J.-Y. Potvin, "Metaheuristics in combinatorial optimization," *Ann. Oper. Res.*, vol. 140, no. 1, pp. 189–213, Nov. 2005, doi: 10.1007/s10479-005-3971-7.
- [19] I. Fister Jr., X.-S. Yang, I. Fister, J. Brest, and D. Fister, "A brief review of nature-inspired algorithms for optimization," 2013, *arXiv:1307.4186*. [Online]. Available: <http://arxiv.org/abs/1307.4186>
- [20] M. Abdel-Basset, L. Abdel-Fatah, and A. K. Sangaiah, "Metaheuristic algorithms: A comprehensive review," in *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications*. Amsterdam, The Netherlands: Elsevier, 2018, pp. 185–231, doi: 10.1016/b978-0-12-813314-9.00010-4.
- [21] A. F. Nematollahi, A. Rahiminejad, and B. Vahidi, "A novel meta-heuristic optimization method based on golden ratio in nature," *Soft Comput.*, vol. 24, no. 2, pp. 1117–1151, Apr. 2019, doi: 10.1007/s00500-019-03949-w.
- [22] R. Leardi, "Application of genetic algorithm-PLS for feature selection in spectral data sets," *J. Chemometrics*, vol. 14, nos. 5–6, pp. 643–655, 2000, doi: 10.1002/1099-128X(200009/12)14:5/6<643::AID-CEM621>3.0.CO;2-E.
- [23] P. Ghamisi and J. A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 309–313, Feb. 2015, doi: 10.1109/lgrs.2014.2337320.
- [24] J. Huang, Y. Cai, and X. Xu, "A hybrid genetic algorithm for feature selection wrapper based on mutual information," *Pattern Recognit. Lett.*, vol. 28, no. 13, pp. 1825–1844, Oct. 2007, doi: 10.1016/j.patrec.2007.05.011.
- [25] S. Oreski and G. Oreski, "Genetic algorithm-based heuristic for feature selection in credit risk assessment," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 2052–2064, Mar. 2014, doi: 10.1016/j.eswa.2013.09.004.
- [26] B. Chakraborty, "Feature subset selection by particle swarm optimization with fuzzy fitness function," in *Proc. 3rd Int. Conf. Intell. Syst. Knowl. Eng.*, Nov. 2008, pp. 1038–1042, doi: 10.1109/iske.2008.4731082.
- [27] S. Lee, S. Soak, S. Oh, W. Pedrycz, and M. Jeon, "Modified binary particle swarm optimization," *Prog. Natural Sci.*, vol. 18, no. 9, pp. 1161–1166, Sep. 2008, doi: 10.1016/j.pnsc.2008.03.018.
- [28] X. Wang, J. Yang, X. Teng, W. Xia, and R. Jensen, "Feature selection based on rough sets and particle swarm optimization," *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 459–471, Mar. 2007, doi: 10.1016/j.patrec.2006.09.003.
- [29] L. Ke, Z. Feng, and Z. Ren, "An efficient ant colony optimization approach to attribute reduction in rough set theory," *Pattern Recognit. Lett.*, vol. 29, no. 9, pp. 1351–1357, Jul. 2008, doi: 10.1016/j.patrec.2008.02.006.
- [30] E. Rashedi, H. Nezamabadi-pour, and S. Saryzadi, "GSA: A gravitational search algorithm," *Inf. Sci.*, vol. 179, no. 13, pp. 2232–2248, Jun. 2009, doi: 10.1016/j.ins.2009.03.004.
- [31] L. Davis, *Handbook of Genetic Algorithms*. New York, NY, USA: Van Nostrand Reinhold, 1991.
- [32] W. D. Hillis, "Co-evolving parasites improve simulated evolution as an optimization procedure," *Phys. D, Nonlinear Phenomena*, vol. 42, nos. 1–3, pp. 228–234, 1990, doi: 10.1016/0167-2789(90)90076-2.

- [33] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, 1997, doi: [10.1023/a:1008202821328](https://doi.org/10.1023/a:1008202821328).
- [34] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA, USA: MIT Press, 1992.
- [35] D. Fogel, "Artificial intelligence through simulated evolution," in *Evolutionary Computation: The Fossil Record*. New York, NY, USA: Wiley, 2009, pp. 227–296, doi: [10.1109/9780470544600.ch7](https://doi.org/10.1109/9780470544600.ch7).
- [36] D. Simon, "Biogeography-based optimization," *IEEE Trans. Evol. Comput.*, vol. 12, no. 6, pp. 702–713, Dec. 2008, doi: [10.1109/tevc.2008.919004](https://doi.org/10.1109/tevc.2008.919004).
- [37] H. Salimi, "Stochastic fractal search: A powerful metaheuristic algorithm," *Knowl.-Based Syst.*, vol. 75, pp. 1–18, Feb. 2015, doi: [10.1016/j.knsys.2014.07.025](https://doi.org/10.1016/j.knsys.2014.07.025).
- [38] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. ICNN Int. Conf. Neural Netw.*, Nov. 1995, pp. 1942–1948, doi: [10.1109/ICNN.1995.488968](https://doi.org/10.1109/ICNN.1995.488968).
- [39] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 28–39, Nov. 2006, doi: [10.1109/mci.2006.329691](https://doi.org/10.1109/mci.2006.329691).
- [40] S. Mirjalili, "The ant lion optimizer," *Adv. Eng. Softw.*, vol. 83, pp. 80–98, May 2015, doi: [10.1016/j.advengsoft.2015.01.010](https://doi.org/10.1016/j.advengsoft.2015.01.010).
- [41] M. Eusuff, K. Lansey, and F. Pasha, "Shuffled frog-leaping algorithm: A memetic meta-heuristic for discrete optimization," *Eng. Optim.*, vol. 38, no. 2, pp. 129–154, Mar. 2006, doi: [10.1080/03052150500384759](https://doi.org/10.1080/03052150500384759).
- [42] K. M. Passino, "Biomimicry of bacterial foraging for distributed optimization and control," *IEEE Control Syst. Mag.*, vol. 22, no. 3, pp. 52–67, Mar. 2002, doi: [10.1109/MCS.2002.1004010](https://doi.org/10.1109/MCS.2002.1004010).
- [43] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," *J. Global Optim.*, vol. 39, no. 3, pp. 459–471, Apr. 2007, doi: [10.1007/s10898-007-9149-x](https://doi.org/10.1007/s10898-007-9149-x).
- [44] X.-S. Yang, "Firefly algorithms for multimodal optimization," in *Stochastic Algorithms: Foundations and Applications*. Berlin, Germany: Springer, 2009, pp. 169–178, doi: [10.1007/978-3-642-04944-6\\_14](https://doi.org/10.1007/978-3-642-04944-6_14).
- [45] A. S. Joshi, O. Kulkarni, G. M. Kakandikar, and V. M. Nandedkar, "Cuckoo search optimization—A review," *Mater. Today, Proc.*, vol. 4, no. 8, pp. 7262–7269, 2017, doi: [10.1016/j.matpr.2017.07.055](https://doi.org/10.1016/j.matpr.2017.07.055).
- [46] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016, doi: [10.1016/j.advengsoft.2016.01.008](https://doi.org/10.1016/j.advengsoft.2016.01.008).
- [47] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper optimisation algorithm: Theory and application," *Adv. Eng. Softw.*, vol. 105, pp. 30–47, Mar. 2017, doi: [10.1016/j.advengsoft.2017.01.004](https://doi.org/10.1016/j.advengsoft.2017.01.004).
- [48] A. Kaveh and N. Farhoudi, "A new optimization method: Dolphin echolocation," *Adv. Eng. Softw.*, vol. 59, pp. 53–70, May 2013, doi: [10.1016/j.advengsoft.2013.03.004](https://doi.org/10.1016/j.advengsoft.2013.03.004).
- [49] M. Jain, V. Singh, and A. Rani, "A novel nature-inspired algorithm for optimization: Squirrel search algorithm," *Swarm Evol. Comput.*, vol. 44, pp. 148–175, Feb. 2019, doi: [10.1016/j.swevo.2018.02.013](https://doi.org/10.1016/j.swevo.2018.02.013).
- [50] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983, doi: [10.1126/science.220.4598.671](https://doi.org/10.1126/science.220.4598.671).
- [51] P. J. M. van Laarhoven and E. H. L. Aarts, "Simulated annealing," in *Simulated Annealing: Theory and Applications*. Amsterdam, The Netherlands: Springer, 1987, pp. 7–15, doi: [10.1007/978-94-015-7744-1\\_2](https://doi.org/10.1007/978-94-015-7744-1_2).
- [52] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel type of phase transition in a system of self-driven particles," *Phys. Rev. Lett.*, vol. 75, no. 6, pp. 1226–1229, Aug. 1995, doi: [10.1103/physrevlett.75.1226](https://doi.org/10.1103/physrevlett.75.1226).
- [53] A. Hatamlou, "Black hole: A new heuristic optimization approach for data clustering," *Inf. Sci.*, vol. 222, pp. 175–184, Feb. 2013, doi: [10.1016/j.ins.2012.08.023](https://doi.org/10.1016/j.ins.2012.08.023).
- [54] A. Kaveh and S. Talatahari, "A novel heuristic optimization method: Charged system search," *Acta Mechanica*, vol. 213, nos. 3–4, pp. 267–289, Jan. 2010, doi: [10.1007/s00707-009-0270-4](https://doi.org/10.1007/s00707-009-0270-4).
- [55] S. Mirjalili, "SCA: A sine cosine algorithm for solving optimization problems," *Knowl.-Based Syst.*, vol. 96, pp. 120–133, Mar. 2016, doi: [10.1016/j.knsys.2015.12.022](https://doi.org/10.1016/j.knsys.2015.12.022).
- [56] S. Mirjalili, S. M. Mirjalili, and A. Hatamlou, "Multi-verse optimizer: A nature-inspired algorithm for global optimization," *Neural Comput. Appl.*, vol. 27, no. 2, pp. 495–513, Mar. 2015, doi: [10.1007/s00521-015-1870-7](https://doi.org/10.1007/s00521-015-1870-7).
- [57] R. V. Rao, V. J. Savsani, and D. P. Vakharia, "Teaching–learning-based optimization: A novel method for constrained mechanical design optimization problems," *Comput.-Aided Des.*, vol. 43, no. 3, pp. 303–315, Mar. 2011, doi: [10.1016/j.cad.2010.12.015](https://doi.org/10.1016/j.cad.2010.12.015).
- [58] E. Atashpaz-Gargari and C. Lucas, "Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition," in *Proc. IEEE Congr. Evol. Comput.*, Sep. 2007, pp. 4661–4667, doi: [10.1109/cec.2007.4425083](https://doi.org/10.1109/cec.2007.4425083).
- [59] T. Ray and K. M. Liew, "Society and civilization: An optimization algorithm based on the simulation of social behavior," *IEEE Trans. Evol. Comput.*, vol. 7, no. 4, pp. 386–396, Aug. 2003, doi: [10.1109/tevc.2003.814902](https://doi.org/10.1109/tevc.2003.814902).
- [60] A. Husseinzadeh Kashan, "League championship algorithm (LCA): An algorithm for global optimization inspired by sport championships," *Appl. Soft Comput.*, vol. 16, pp. 171–200, Mar. 2014, doi: [10.1016/j.asoc.2013.12.005](https://doi.org/10.1016/j.asoc.2013.12.005).
- [61] A. Kaveh, "Tug of war optimization," in *Advances in Metaheuristic Algorithms for Optimal Design of Structures*. Cham, Switzerland: Springer, Nov. 2016, pp. 451–487, doi: [10.1007/978-3-319-46173-1\\_15](https://doi.org/10.1007/978-3-319-46173-1_15).
- [62] R. Moghdani and K. Salimifard, "Volleyball premier league algorithm," *Appl. Soft Comput.*, vol. 64, pp. 161–185, Mar. 2018, doi: [10.1016/j.asoc.2017.11.043](https://doi.org/10.1016/j.asoc.2017.11.043).
- [63] E.-G. Talbi, "Taxonomy of hybrid metaheuristics," *J. Heuristics*, vol. 8, no. 5, pp. 541–564, 2002, doi: [10.1023/a:1016540724870](https://doi.org/10.1023/a:1016540724870).
- [64] I.-S. Oh, J.-S. Lee, and B.-R. Moon, "Hybrid genetic algorithms for feature selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1424–1437, Nov. 2004, doi: [10.1109/tpami.2004.105](https://doi.org/10.1109/tpami.2004.105).
- [65] O. C. Martin and S. W. Otto, "Combining simulated annealing with local search heuristics," *Ann. Oper. Res.*, vol. 63, no. 1, pp. 57–75, Feb. 1996, doi: [10.1007/bf02601639](https://doi.org/10.1007/bf02601639).
- [66] M. Ghosh, T. Kundu, D. Ghosh, and R. Sarkar, "Feature selection for facial emotion recognition using late hill-climbing based memetic algorithm," *Multimedia Tools Appl.*, vol. 78, no. 18, pp. 25753–25779, Jun. 2019, doi: [10.1007/s11042-019-07811-x](https://doi.org/10.1007/s11042-019-07811-x).
- [67] M. Mafarja and S. Abdullah, "Investigating memetic algorithm in solving rough set attribute reduction," *Int. J. Comput. Appl. Technol.*, vol. 48, no. 3, p. 195, 2013, doi: [10.1504/ijcat.2013.056915](https://doi.org/10.1504/ijcat.2013.056915).
- [68] M. Tubishat, N. Idris, L. Shuib, M. A. M. Abushariah, and S. Mirjalili, "Improved salp swarm algorithm based on opposition based learning and novel local search algorithm for feature selection," *Expert Syst. Appl.*, vol. 145, May 2020, Art. no. 113122, doi: [10.1016/j.eswa.2019.113122](https://doi.org/10.1016/j.eswa.2019.113122).
- [69] J. Jona and N. Nagaveni, "A hybrid swarm optimization approach for feature set reduction in digital mammograms," *WSEAS Trans. Inf. Sci. Appl.*, vol. 9, no. 11, pp. 340–349, 2012.
- [70] Q. Al-Tashi, S. J. Abdul Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, "Binary optimization using hybrid grey wolf optimization for feature selection," *IEEE Access*, vol. 7, pp. 39496–39508, 2019, doi: [10.1109/access.2019.2906757](https://doi.org/10.1109/access.2019.2906757).
- [71] X. Lai and M. Zhang, "An efficient ensemble of GA and PSO for real function optimization," in *Proc. 2nd IEEE Int. Conf. Comput. Sci. Inf. Technol.*, Aug. 2009, pp. 651–655, doi: [10.1109/iccscit.2009.5234780](https://doi.org/10.1109/iccscit.2009.5234780).
- [72] M. E. Basiri and S. Nemati, "A novel hybrid ACO-GA algorithm for text feature selection," in *Proc. IEEE Congr. Evol. Comput.*, May 2009, pp. 2561–2568, doi: [10.1109/cec.2009.4983263](https://doi.org/10.1109/cec.2009.4983263).
- [73] E. Zorapaci and S. A. Özel, "A hybrid approach of differential evolution and artificial bee colony for feature selection," *Expert Syst. Appl.*, vol. 62, pp. 91–103, Nov. 2016, doi: [10.1016/j.eswa.2016.06.004](https://doi.org/10.1016/j.eswa.2016.06.004).
- [74] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, Apr. 1997, doi: [10.1109/4235.585893](https://doi.org/10.1109/4235.585893).
- [75] X.-S. Yang, "Harmony search as a metaheuristic algorithm," in *Music-Inspired Harmony Search Algorithm*. Berlin, Germany: Springer, 2009, pp. 1–14, doi: [10.1007/978-3-642-00185-7\\_1](https://doi.org/10.1007/978-3-642-00185-7_1).
- [76] M. Mahdavi, M. Fesanghary, and E. Damangir, "An improved harmony search algorithm for solving optimization problems," *Appl. Math. Comput.*, vol. 188, no. 2, pp. 1567–1579, May 2007, doi: [10.1016/j.amc.2006.11.033](https://doi.org/10.1016/j.amc.2006.11.033).
- [77] K. S. Lee and Z. W. Geem, "A new meta-heuristic algorithm for continuous engineering optimization: Harmony search theory and practice," *Comput. Methods Appl. Mech. Eng.*, vol. 194, nos. 36–38, pp. 3902–3933, Sep. 2005, doi: [10.1016/j.cma.2004.09.007](https://doi.org/10.1016/j.cma.2004.09.007).
- [78] M. G. H. Omran and M. Mahdavi, "Global-best harmony search," *Appl. Math. Comput.*, vol. 198, no. 2, pp. 643–656, May 2008, doi: [10.1016/j.amc.2007.09.004](https://doi.org/10.1016/j.amc.2007.09.004).



- [79] Z. W. Geem, "Optimal cost design of water distribution networks using harmony search," *Eng. Optim.*, vol. 38, no. 3, pp. 259–277, Apr. 2006, doi: [10.1080/03052150500467430](https://doi.org/10.1080/03052150500467430).
- [80] J. B. Fraleigh, *A first course in Abstract Algebra*. London, U.K.: Pearson Education India, 2003.
- [81] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statistician*, vol. 46, no. 3, p. 175, Aug. 1992, doi: [10.2307/2685209](https://doi.org/10.2307/2685209).
- [82] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
- [83] I. Rish, "An empirical study of the naive Bayes classifier," in *Proc. IJCAI Workshop Empirical Methods Artif. Intell.*, vol. 3, no. 22, 2001, pp. 41–46.
- [84] G. van Rossum and F. L. Drake, *The Python Language Reference Manual*. Boston, MA, USA: Network Theory, 2011.
- [85] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- [86] M. Mafarja, A. Qasem, A. A. Heidari, I. Aljarah, H. Faris, and S. Mirjalili, "Efficient hybrid nature-inspired binary optimizers for feature selection," *Cognit. Comput.*, vol. 12, no. 1, pp. 150–175, Jul. 2019, doi: [10.1007/s12559-019-09668-6](https://doi.org/10.1007/s12559-019-09668-6).
- [87] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, Jan. 2016, doi: [10.1016/j.neucom.2015.06.083](https://doi.org/10.1016/j.neucom.2015.06.083).
- [88] M. Mafarja and S. Mirjalili, "Whale optimization approaches for wrapper feature selection," *Appl. Soft Comput.*, vol. 62, pp. 441–453, Jan. 2018, doi: [10.1016/j.asoc.2017.11.006](https://doi.org/10.1016/j.asoc.2017.11.006).
- [89] F. Wilcoxon, "Individual comparisons by ranking methods," in *Breakthroughs in Statistics*. New York, NY, USA: Springer, 1992, pp. 196–202, doi: [10.1007/978-1-4612-4380-9\\_16](https://doi.org/10.1007/978-1-4612-4380-9_16).



**PAWAN KUMAR SINGH** (Member, IEEE) received the B.Tech. degree in information technology from the West Bengal University of Technology, in 2010, and the M.Tech. degree in computer science and engineering and the Ph.D. (engineering) degree from Jadavpur University (J.U.), in 2013 and 2018, respectively. He received the RUSA 2.0 Fellowship for pursuing his postdoctoral research in J.U., in 2019. He is currently working as an Assistant Professor with the Department of Information Technology, J.U. He has published more than 50 research papers in peer-reviewed journals and international conferences. His current research interests include computer vision, pattern recognition, handwritten document analysis, image processing, machine learning, and artificial intelligence. He is a member of the Institution of Engineers (India) and the Association for Computing Machinery (ACM) and a Life Member of the Indian Society for Technical Education (ISTE, New Delhi) and the Computer Society of India (CSI).



**ZONG WOO GEEM** (Member, IEEE) received the B.Eng. degree from Chung-Ang University, the Ph.D. degree from Korea University, and the M.Sc. degree from Johns Hopkins University. He researched at Virginia Tech, the University of Maryland at College Park, and Johns Hopkins University. He is an Associate Professor with the Department of Energy IT, Gachon University, South Korea. He invented a music-inspired optimization algorithm, harmony search, which has been applied to various scientific and engineering problems. His research interest includes phenomenon-mimicking algorithms and their applications to energy, environment, and water fields. He served as an Associate Editor for *Engineering Optimization* and a Guest Editor for *Swarm and Evolutionary Computation*, the *International Journal of Bio-Inspired Computation*, the *Journal of Applied Mathematics*, *Applied Sciences*, and *Sustainability*.



**RAM SARKAR** (Senior Member, IEEE) received the B.Tech. degree in computer science and engineering from the University of Calcutta, in 2003, and the M.E. degree in computer science and engineering and the Ph.D. degree in engineering from Jadavpur University, in 2005 and 2012, respectively. He joined the Department of Computer Science and Engineering, Jadavpur University, as an Assistant Professor, in 2008, where he is currently working as an Associate Professor. He received the Fulbright-Nehru Fellowship (USIEF) for postdoctoral research with the University of Maryland at College Park, USA, from 2014 to 2015. His current research interests include image processing, pattern recognition, machine learning, and bioinformatics.



**SHAMEEM AHMED** is currently pursuing the sophomore bachelor's degree in computer science and engineering with Jadavpur University, Kolkata, India. His research interests include machine learning, optimization, and image processing.



**KUSHAL KANTI GHOSH** is currently pursuing the bachelor's degree in computer science and engineering with Jadavpur University, Kolkata, India. His research interests include machine learning, optimization, game theory, and image processing.

...