





Digital Object Identifier

Mayfly in Harmony: A New Hybrid Meta-heuristic Feature Selection Algorithm

TRINAV BHATTACHARYYA¹, BITANU CHATTERJEE¹, PAWAN KUMAR SINGH ² (Member, IEEE), JIN HEE YOON ³ (Member, IEEE), ZONG WOO GEEM ⁴ (Member, IEEE), RAM SARKAR ¹ (Senior Member, IEEE)

¹Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India

²Department of Information Technology, Jadavpur University, Kolkata 700106, India

³School of Mathematics and Statistics, Sejong University, Seoul 05006, South Korea

⁴Department of Energy IT, Gachon University, Seongnam 13120, South Korea

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

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2020R1A2C1A01011131)

ABSTRACT Feature selection is a process to reduce the dimension of a dataset by removing redundant features, and to use the optimal subset of features for machine learning or data mining algorithms. This helps to minimize the time requirement to train a learning algorithm as well as to lessen the storage requirement by ignoring the less-informative features. Feature selection can be considered as a combinatorial optimization problem. In this paper, the authors have presented a new feature selection algorithm called Mayfly-Harmony Search (MA-HS) based on two meta-heuristics namely Mayfly Algorithm and Harmony Search. Mayfly Algorithm has not hitherto been used for feature selection problems to the best of the author's knowledge. An S-shaped transfer function is incorporated for converting it into a binary version of Mayfly Algorithm. When different candidate solutions obtained from various regions of the search space using Mayfly Algorithm are taken into the harmony memory and processed by Harmony Search, a superior solution can be ensured. This is the primary reason for proposing a hybrid of Mayfly Algorithm and Harmony Search. Thus, combining harmony search with Mayfly Algorithm leads to an increased exploitation of the search space and an overall improvement in the performance of Mayfly-Harmony Search (MA-HS) algorithm. The proposed algorithm has been applied on 18 UCI datasets and compared with 12 other state-of-the-art meta-heuristic FS methods. Experiments have also been performed on three high-dimensional microarray datasets. The results obtained support the superior performance of the algorithm compared to the other methods. The source code of the proposed algorithm can be found using the link as follows: <https://github.com/trin07/MA-HS>.

INDEX TERMS MA-HS algorithm, Mayfly Optimization, Harmony Search, Feature selection, Meta-heuristic, Hybrid method, UCI datasets

I. INTRODUCTION

In recent years, with the improvement in data collection methods in various fields, the amount of data available has increased dramatically. This has led to greater complexity in terms of both computational time and space required to execution of algorithms in the domains of machine learning and data mining. Much of this available data is redundant and decreases the efficiency of the algorithm. For classification tasks, the irrelevant data causes a significant drop in accuracy and performance. This is why feature selection (FS) has gained importance in the scientific community in recent

times.

A. FEATURE SELECTION

FS aims to select only those features in the dataset which will be useful for classification and discard the irrelevant ones. FS has been used for various purposes such as neuro-imaging [1], gene prediction [2], text mining, image processing and fault diagnosis in industrial applications [3].

FS methods can be broadly classified into filters, wrappers and embedded. Filter techniques [4] use various scoring metrics to rank the features and choose the most important

ones so no classifier is required to gauge the effectiveness of the filter methods. Wrapper techniques [5] use a classification algorithm to evaluate different feature subsets and the one which produces the best result is declared as the solution. This is why wrapper methods are generally considered to be superior as filter methods completely ignore the relation between the chosen subset and the inductive algorithm performance [6]. The optimal feature subset is influenced by the specific bias and heuristic of the inductive algorithm. Embedded techniques [7] are similar to wrapper methods but instead use an intrinsic model building metric during learning.

FS is an NP-hard problem as there are 2^n possible solutions for a dataset having n features in total. Evaluating each of these solutions is not a feasible option because of the high computational cost involved. A feature subset which is a possible solution to the problem at hand can be selected using random search. This can be done using a heuristic strategy which performs a guided search over the entire solution space to find a reasonably good feature subset which may not be the optimal solution but is acceptable within computational constraints. Higher-level heuristics or meta-heuristics have become quite popular in recent years to solve FS problems in different fields [8] such as handwriting recognition [9], benchmark problems [10], gene selection [11], medical diagnosis [12], financial problems, network intrusion and security .

B. MOTIVATION

Numerous meta-heuristic algorithms have been proposed to solve the FS problem in various domains in the literature. Each of these algorithms has its own pros and cons in dealing with the particular problem at hand. The performance of an algorithm depends on its exploration and exploitation capabilities, as well as the inherent characteristics of the dataset used. Hybridized algorithms seek to combine two or more meta-heuristics so that the shortcomings of a particular method can be dealt with by another method. For instance, an algorithm with good exploration ability can be hybridized with one with good exploitation ability in order to set a good trade-off between exploration and exploitation capabilities of the overall system. This is an active area of research today. The No Free Lunch (NFL) theorem [13] states that the performance, averaged over all problems in the class, is the same for any solution method. This means that on a particular problem, different algorithms may obtain different results, but over all problems, they are equivalent. This has encouraged researchers to build on previous work and come up with new problem-specific approaches for improved performance. In this paper, the authors have proposed a hybrid algorithm based on Harmony Search (HS) algorithm [14] and Mayfly Algorithm (MA) [15] for solving the FS problem. HS algorithm has been applied for optimization in various domains in the past. MA is a recently proposed meta-heuristic which has been shown to perform well in dealing with optimization problems. Despite being similar to PSO, MA is seen to

have greater ability in finding out a more optimal solution than PSO and hence it has better chances of finding the globally optimal solution. However, in certain cases, premature convergence diminishes the quality of the final solution. As explained earlier, an algorithm with good exploration ability can be hybridized with another algorithm having high exploitation ability. MA is seen to effectively explore the search space while HS algorithm exploits and improves the existing feature subsets. This provides a compelling reason to hybridize MA with HS algorithm. If different solutions obtained from various regions of the search space using MA are put into the harmony memory, then a superior solution can be found by HS algorithm. Thus, combining HS algorithm with MA can lead to an increased exploitation of the search space and an overall improvement in the performance of the hybrid algorithm. This has been proven in this work which is evident from the results obtained.

C. CONTRIBUTIONS

The main contributions of this work are as follows:

- Hybridization of MA with HS for the first time to the best of the authors' knowledge.
- Modification of MA to enhance its exploration ability.
- Use of HS algorithm to increase the exploitation ability of the hybrid MA-HS algorithm.
- Use of S-shaped transfer function to convert continuous search space into a binary search space.
- Validate the effectiveness of MA-HS by applying it on 18 standard UCI datasets and comparison with 12 state-of-the-art FS algorithms, and then providing a statistical analysis of the results obtained.
- Prove the robustness of the algorithm by applying it on 3 high dimensional microarray datasets.

II. LITERATURE SURVEY

FS is essentially a combinatorial optimization problem. Often, in real-world optimization problems, it is not feasible to perform an exhaustive search due to the large amount of time required for processing. This is where heuristic methods are particularly useful. These methods aim to find good solutions, which may not be optimal, in reasonable computational time. Meta-heuristics are advanced heuristics which provide a set of guidelines or strategies to develop heuristic optimization algorithms [16]. The crux of a meta-heuristics algorithm is formed by the processes of intensification and diversification, which correspond to exploitation and exploration of the search space. Numerous meta-heuristic algorithms have been proposed in recent years to solve optimization problems in various domains.

Meta-heuristics can be nature inspired or non-nature inspired. Nature inspired meta-heuristic algorithms can be broadly classified into the four categories of evolutionary algorithms, physics-based algorithms, swarm-based algorithms and human-based algorithms, as shown in Figure 1.

- *Evolutionary algorithms*: These techniques are based on biological evolution involving genetic recombination,

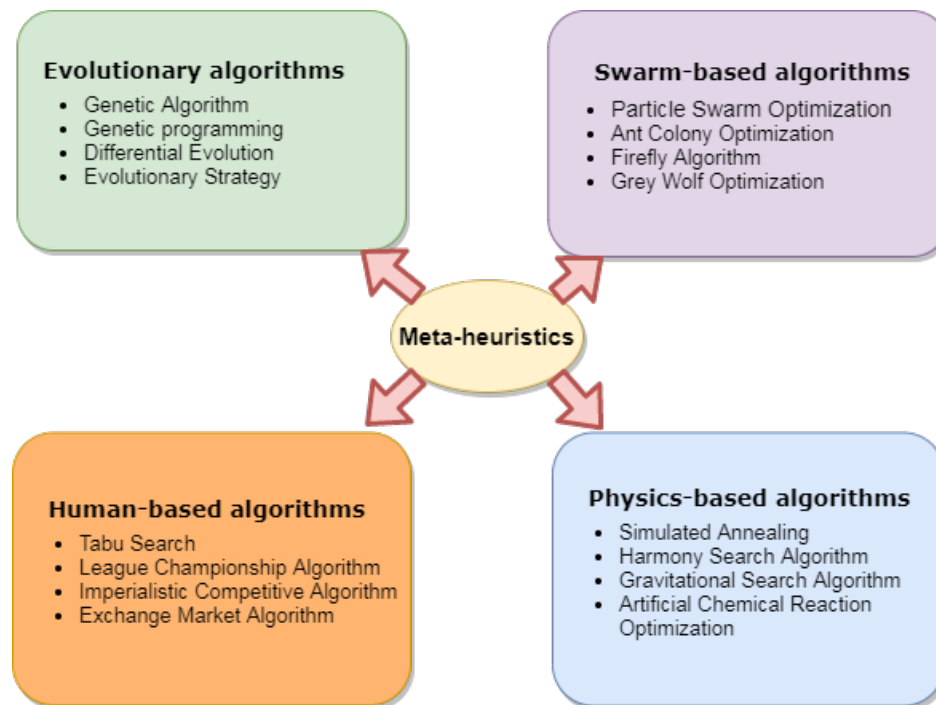


FIGURE 1: Different categories of meta-heuristic algorithms

mutation and natural selection. Examples include Genetic Algorithm (GA) [17], Evolutionary Strategy (ES) [18] and Genetic Programming (GP) [19].

- *Swarm-based algorithms*: These algorithms are based on the behavior of a population of social organisms as a whole. Here, individuals seek to interact with one another and try to improve themselves based on the knowledge gained by the entire swarm. Examples include Particle Swarm Optimization (PSO) [20], Ant Colony Optimization (ACO) [21], Bat-inspired Algorithm (BA) [22] and Firefly Algorithm (FA) [23].
- *Physics-based algorithms*: These processes are based upon physical phenomenon like gravity, inertia and electromagnetic force. Some algorithms under this category are Simulated Annealing (SA) [24], Gravitational Search Algorithm (GSA) [25], Harmony Search (HS) [14], Ray Optimization (RO) [26] and Central Force Optimization (CFO) [27].
- *Human-based algorithms*: These are meta-heuristic techniques inspired by human behavior. Examples include Imperialist Competitive Algorithm (ICA) [28], League Championship Algorithm (LCA) [29] and Exchange Market Algorithm (EMA) [30].

Meta-heuristics can also be classified based on the number of solutions dealt with by the algorithm [31]. *Trajectory-based algorithms* involve a single solution and work on improving it as it navigates through the search space. These are often hybridized with other diversification techniques to prevent the solution from being stuck in local optima. Examples include SA, TS, Greedy Randomized Adaptive Search

Procedure (GRASP) [32] and Variable Neighborhood Search (VNS) [33]. *Population-based algorithms*, on the other hand, deal with a set of solutions simultaneously. These solutions (called agents) are allowed to improve using the information obtained from mutual interaction. New and better solutions are produced by merging different solutions. Examples include GA, PSO, ACO, Red Deer Algorithm (RDA) [34] and Shuffled Frog Leaping Algorithm (SFLA) [35].

The number of new meta-heuristic algorithms developed in recent years and used in the field of optimization has been visualized in Figure 2. Most of these methods have been used to solve the FS problem in the literature. Hybridized algorithms, where two or more independent methods are combined together, have grown in popularity over recent years. In combination, one method can effectively overcome the shortcomings of another, leading to improved results. GA was used for FS in automatic feature classifiers in [36], and hybridized in hybridized GA [37]. Binary PSO was used to solve the FS problem in [38]. It was further improved as Sentiment Fitness Sum Binary PSO (SCO-FS-BPSO) in [39]. A hybrid version of the Grasshopper Optimization Algorithm (GOA) was used in [40] for FS and parameter tuning of the SVM classifier. Mirjalili et al. [41] developed Ant Lion Optimizer (ALO) for FS, based on the behavior of ant lions. A binary GSA, based on Newtonian laws of gravity and motion, was proposed by Rashedi et al. [42]. The authors of [43] proposed a hybrid binary CRO with SA for FS in biomedical datasets. A hybrid Harris Hawks Optimizer (HHO) with SA was proposed by the authors of [44]. Zhang et al. [45] proposed a new approach for multi-objective FS called Binary

differential evolution with self-learning (MOFS-BDE). An efficient method to prevent premature convergence called Return cost based binary firefly algorithm (Rc-BBFA) was proposed by the authors in [46]. Hegazy et al. proposed an improved version of the Salp Swarm Algorithm (SSA) [47], based on the swarming mechanism of salps, for FS. Other notable meta-heuristics include artificial bee colony [48, 49], modified PSO [citezhang2019filter,zhang2015multi].

The authors of [50] used PSO to select informative text features for improved text clustering. In [51] the authors proposed an enhanced Krill Herd Algorithm (KHA) for FS in the domain of text document clustering. A hybrid version of this method (MMKHA) was used in [52]. The methods proposed by the authors of [53] mimicked the inner workings of the brain for symbolic optimization and were used for object detection, categorization and visual tracking. The communication method used by honeybees inspired the work of [54] for 3D reconstruction which is robust to outliers. The authors of [55] used to swarm behaviour of the Honeybee Search Algorithm (HSA) for object tracking.

HS was proposed by Geem et al. [14]. Since its initial development in 2000, HS has been applied to a wide range of computational problems. Applications include water distribution networks [56, 57]; structural design [58]; benchmark optimization [59]; soil stability analysis [60, 61]; transport-related problems [62]; energy studies [63], image enhancement [64] and medical diagnosis [65]. The MA [15] is inspired by the flight patterns and mating behavior of mayflies. It has been used for solving optimization problems which are encountered in the real world. effectively. In this paper, the authors have proposed a novel hybrid algorithm for FS named as MAHS algorithm, based on MA and HS algorithm.

III. SOME PRELIMINARIES

A. MAYFLY OPTIMIZATION ALGORITHM

Mayflies are insects that belong to the order Ephemeroptera, part of a group of insects known as Palaeoptera. These insects appear mainly during the month of May in the UK, thus having the name Mayfly. Immature mayflies spend several years growing as aquatic nymphs until they are ready to go to the surface as adult mayflies. Most male adults assemble in swarms a few metres above the water to attract the females. They perform a nuptial dance which involves characteristic up and down movement generating a pattern. Female mayflies go to these swarms for mating. The mating process lasts only for a few seconds after which they drop the eggs in the water and the cycle continues. A detailed description of the above process is mentioned in the works of Allan et al. [66] and Peckarsky et al. [67].

MA has been developed by Zervoudakis et al. [15] and is a new method for solving FS problems. It is a hybrid method combining the advantages of classical optimization methods such as PSO [68], GA [69] and FA. In the work of [70], it has been shown that PSO needed modifications as it is likely to get stuck in a local optimum, especially for problems

having a high dimension. The MA performs the necessary modifications, thereby enabling the algorithm to have a better performance across small and large scale feature sets. The components of MA are given as follows:

- **Movement of male mayflies:** The position of a male mayfly is updated as in Equation 1:

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

where x_i^t is the present position of the male mayfly and the new position x_i^{t+1} is obtained by adding the current position with the velocity v_i^{t+1} . The male mayflies are always a few meters above the surface of water and develop great speeds. A male mayfly's velocity is calculated as in Equation 2:

$$v_{kj}^{t+1} = g * v_{kj}^t + a_1 * e^{-\beta r_p^2} * (pbest_{kj} - x_{kj}^t) + a_2 * e^{-\beta r_g^2} * (gbest_j - x_{kj}^t) \quad (2)$$

where v_{kj}^t is the velocity of mayfly k in dimension j at time t , x_{kj}^t is the position of the same mayfly at time t , a_1 and a_2 are positive attraction constants which are used for measuring the contribution of the cognitive and social components respectively, g is a gravitational coefficient and β is a fixed visibility coefficient used to limit a mayfly's visibility to others. $pbest_k$ is the most optimal position that the particular mayfly k ever visited and $gbest_j$ is the j^{th} component of the position of the best male mayfly. Since this is a minimization problem so $pbest_k$ is updated as follows:

$$pbest_k = \begin{cases} x_k^{t+1} \\ if fitness(x_k^{t+1}) < fitness(pbest_k) \end{cases} \quad (3)$$

where $fitness(x_k^t)$ gives the fitness value of a position as given in subsection IV-B, i.e., the quality of a solution. Finally r_p is the Cartesian distance between x_k and $pbest_k$, while r_g is the Cartesian distance between x_k and $gbest$. These are calculated as shown:

$$|x_k - X_k| = \sqrt{\sum_{j=1}^n (x_{kj} - X_{kj})^2} \quad (4)$$

where x_{kj} represents the position of the j^{th} element of k^{th} mayfly and X_k either represents $pbest_k$ or $gbest$. It is essential that the best mayflies at a particular time keep performing the nuptial dance which gives a stochastic element to the algorithm. Mathematical representation of this dance is given in Equation 5.

$$v_{kj}^{t+1} = g * v_{kj}^t + d * r \quad (5)$$

where d is the coefficient of nuptial dance and r is a random value $\in [-1, 1]$. The nuptial dance coefficient progressively reduces as $d_{itr} = d_0 \times \delta^{itr}$. In this equation d_0 is the initial value of the nuptial dance coefficient, itr is the current number of iterations and δ is a random value $\in [0, 1]$.

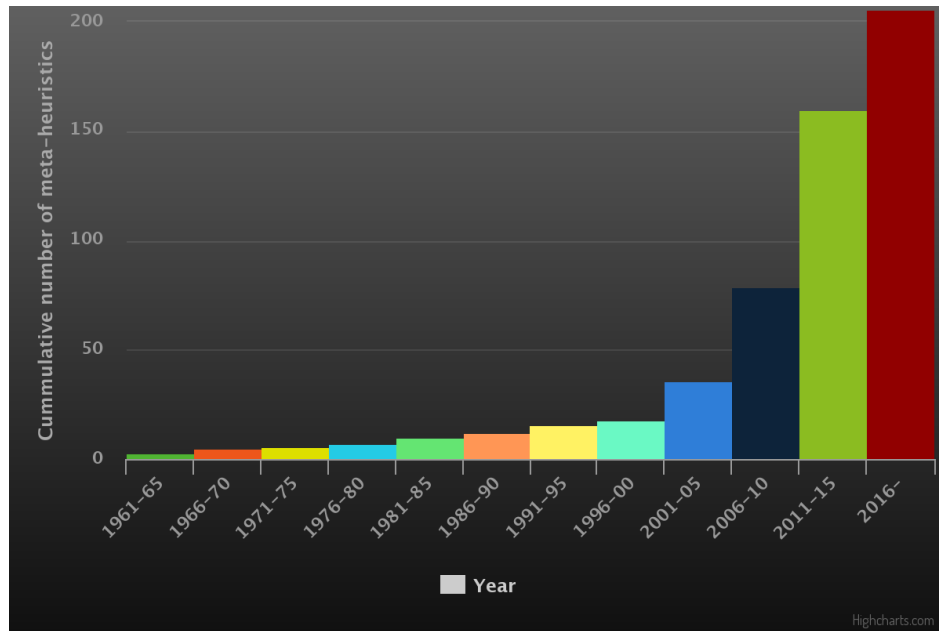


FIGURE 2: A histogram showing the rise in the cumulative number of meta-heuristic algorithms developed in recent years

- **Movement of female mayflies:** The female mayflies move towards the males for breeding. The position of a female mayfly is updated as follows:

$$y_i^{t+1} = y_i^t + v_i^{t+1} \quad (6)$$

where y_i^t is the current position of the female mayfly at time t and it is updated by adding its velocity v_i^{t+1} . The attraction process between males and females depends on the quality of the current solution, i.e., the best performing female is attracted to the best performing male and so on. The velocity of a female is updated as in Equation 7

$$v_{kj}^{t+1} = \begin{cases} \text{if } fitness(y_k) > fitness(x_k) \\ g * v_{kj}^t + a_2 * e^{-\beta r_{mf}^2} * (x_{kj}^t - y_{kj}^t) \\ \text{elseif } fitness(y_k) \leq fitness(x_k) \\ g * v_{kj}^t + fl * r \end{cases} \quad (7)$$

where v_{kj}^t is the j^{th} component of the k^{th} female mayfly's velocity at time t , y_{kj}^t is the position of female mayfly k in dimension j at time t , x_{kj}^t is the j^{th} component of the position of male mayfly k at time t , a_2 and β are previously defined attraction constant and visibility coefficient respectively, g is the gravity coefficient defined before in Equation 2, r is a random value $\in [-1, 1]$, and r_{mf} is the Cartesian distance between male and female mayflies which is given in Equation 4. fl is a random walk coefficient in the instance of a female not being attracted with a male and $fl_{itr} = fl_0 \times \delta^{itr}$. Here itr and δ are two previously defined variables in Equation 5.

- **Crossover between mayflies:** The crossover operation is done by first selecting a male mayfly and then a

female. This selection is done based on their fitness value, i.e., the best male breeds with the best female. Two offsprings are produced after a crossover as shown in Equation 9.

$$offspring1 = r_{of} * male + (1 - r_{of}) * female \quad (8)$$

$$offspring2 = r_{of} * female + (1 - r_{of}) * male \quad (9)$$

Here *male* is the parent male mayfly, *female* is the female parent and r_{of} is a stipulated value between 0 and 1. The initial velocities of the offspring are set as 0.

- **Mutation of mayflies:** The newly generated offspring are mutated to enhance the exploration ability of the algorithm. A normally distributed random number is added to the offspring's variable as described in

$$offspring'_n = offspring_n + \kappa \quad (10)$$

where κ is the normally distributed random value.

B. HARMONY SEARCH

Harmony in nature is a special relationship between several sound waves that have different frequencies. Musical performances tend to have a soothing effect on humans. The amount of soothing effect is determined by its aesthetic estimation. A musical performer strives to achieve the best aesthetic state which is analogous to a global optimum. Thus finding the best aesthetic state is an optimization problem determined by an objective function. Aesthetic estimation is calculated by the set of sounds played by simultaneously joined instruments. The aesthetic value of the sound generated can be improved by practice.

A novel method called HS was introduced by Geem et al. [14] to generate a set of harmonies which are improvised

similar to an evolutionary algorithm. The algorithm initialises a Harmony Memory (HM) and subsequently modifies it to create new harmonies. The new harmonies replace the previous ones if they are found to be better with respect to an objective function. In this way HS performs improvisation upon existing musical harmonies and achieves an optimal combination of harmonies. The HS algorithm is a reliable tried-and-tested method and has been previously used for optimization in the work of [57]. The pseudo code of HS is given in Algorithm 1.

Algorithm 1 Pseudo code for Harmony Search

Input: MaxHS, NumVariables, r_{accept} , r_{pa} , b_{range}
 Output: Best Solution $X = (x_1, x_2, \dots, x_d)$

```

Initialize HM
for itr ← 1 to MaxHS do
  for i ← 1 to NumVariables do
    if rand1 <  $r_{accept}$  then
      Choose a value from HM for the  $i$ th variable
    if rand2 <  $r_{pa}$  then
      Adjust the value by an amount of  $b_{range} * rand3$ 
    end if
  else
    Choose a random value
  end if
end for
Accept the new solution if better
end for
Return the current best solution
  
```

IV. PROPOSED WORK

FS is a binary optimization problem having restricted solutions which are represented by either 0 or 1. Each agent or solution is represented by a binary vector where a value 1 shows that the chosen feature is selected and value of 0 represents that the feature is not selected. The size of this solution vector corresponds to the number of features in a particular dataset. In FS, the selected feature subset needs to be evaluated during every iteration of the algorithm. This is therefore a difficult task, especially the wrapper-based methods.

A. PROPOSED MA-HS ALGORITHM

The MA described in subsection III-A was originally developed for continuous optimization problems. The re-designed algorithm for FS problems, called **MA-HS** is described in Algorithm 2. Each solution vector in MA is converted into its binary form, i.e., consisting only of 0's and 1's and then evaluated. To achieve this conversion, the S-shaped transfer function is used. This function gives the probability of whether to choose a particular feature in a solution vector. It is a reliable function and has previously been used by many researchers [71, 72]. The S-shaped transfer function

Algorithm 2 Pseudo code for the proposed MA-HS Algorithm

Input: PopSize, MaxIter

Output: Best Agent $X = (x_1, x_2, \dots, x_d)$

```

Initiate population and velocity of male and female mayflies randomly
Evaluate population and then find gbest
for itr ← 1 to MaxIter do
  for i ← 1 to PopSize do
    Update pbest
    Evaluate and update the velocities of male and female mayflies
  end for
  Sort the mayflies and rank them
  Perform crossover and generate male and female offspring
  Mutate the offspring
  Replace worst mayflies with the best new offspring generated
  Perform HS on male mayflies
  Update gbest
end for
  
```

used in this algorithm is depicted in Figure 4 and shown in Equation 11.

$$S(x) = \frac{1}{1 + e^{-x}} \quad (11)$$

During the conversion process, the agent's feature is updated according to Equation 12.

$$P_d^{t+1} = \begin{cases} 1 & \text{if } S(P_d^{t+1}) > rand \\ 0 & \text{if } S(P_d^{t+1}) \leq rand \end{cases} \quad (12)$$

where P_d^{t+1} is the updated feature subset of the agent, $rand$ is a random number between 0 and 1, and $S(P_d^{t+1})$ is the S-shaped transfer function as previously defined in Equation 11. A broad outline of the proposed framework is given in Figure 3.

B. FITNESS FUNCTION

The algorithm evaluates the quality of a solution in this section. A learning algorithm has been used since it is a wrapper-based method. The K-Nearest Neighbor (KNN) classifier [73] is used for finding out the classification accuracy. The fitness function consists of the classification error and the number of features. In FS, the authors intend to increase the accuracy and simultaneously reduce the number of features. For this purpose, classification error is used instead of accuracy. This is because both the error and number of features needs to be decreased. Combining these two, the fitness function will therefore be reduced to a single objective function. Equation 13 gives the function for evaluating a feature subset.

$$\downarrow Fitness = \gamma \times \lambda + (1 - \gamma) \times \frac{|f|}{|F|} \quad (13)$$

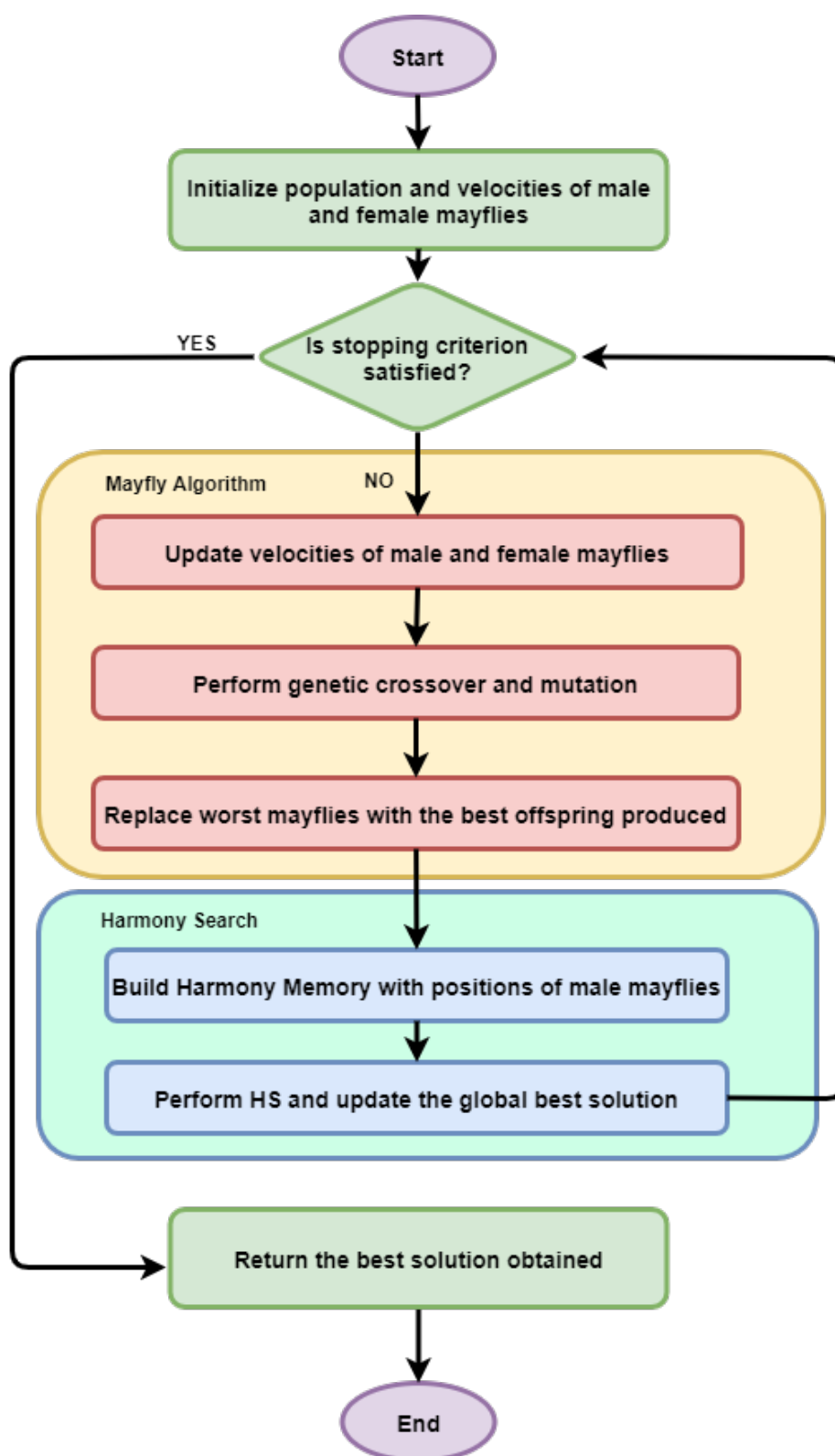


FIGURE 3: An outline of the proposed MA-HS algorithm used for solving FS problems

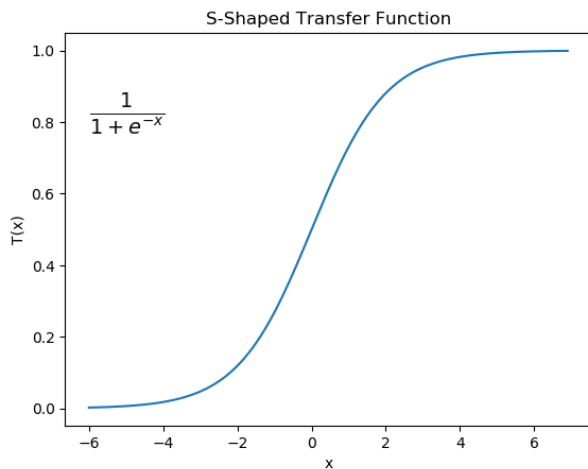


FIGURE 4: S-shaped transfer function for converting continuous Mayfly search space into binary

where $|f|$ is the number of features in the feature subset, $|F|$ is the number of features in the given dataset, λ is the classification error and $\gamma \in [0, 1]$ is a parameter that gives a relative contribution between the classification error and the number of features.

V. EXPERIMENTAL RESULTS

This section deals with the results which support the effectiveness of the MA-HS algorithm for solving the FS problem. KNN classifier [73] has been used for finding out the classification accuracy. The authors have set $K = 5$ as per recommendations reported by Mafarja et al. [74]. In each dataset, 80% of the instances are used for training and rest 20% are used for testing the classification model. The method proposed here is implemented using Python3 [75] and graphs are obtained using Matplotlib [76]. The experiments are conducted on a PC with 3.30 GHz Intel(R) Pentium(R) CPU G4400 and 8 GB Memory. The time complexity of the algorithm is found out to be $O(MaxIter * PopSize * MaxHS * (time_{fitness} + d))$ where $MaxIter$ is the maximum number of iterations, $PopSize$ is the population size, $MaxHS$ is the maximum number of iterations of HS, $time_{fitness}$ is the complexity of calculating the fitness of an agent and d is the dimension of the dataset.

A. DATASET DESCRIPTION

For this work, the authors have chosen 18 different datasets from the UCI data repository [77]. They are utilized for evaluating the efficacy of the proposed method. The 18 datasets consist of 15 bi-class and 3 multi-class datasets as described in Table 1.

B. TUNING OF PARAMETERS

The aim of the proposed method is to maximize the classification accuracy and try to reduce the number of features at the same time. In other words both the classification error as

well as the number of features must be reduced. That is why the fitness function includes components of classification error and the number of features. A single parameter, in this case γ can be used for determining the relative weightage given to error and the number of features. Hence the objective is to minimize the fitness function which in turn will minimize the classification error and the number of features. In the proposed work, greater importance has been given to increase the classification accuracy. So the value of γ is set as 0.9 in Equation 13. The gravity coefficient g is set as 0.98 to imitate the effect of gravity on mayflies. Furthermore, the values of d , fl , a_1 and a_2 have been set to a comparatively high values to allow the mayflies to move freely. This has been done to enhance the exploration ability of the algorithm. The final list of parameter values is given in Table 2.

Figure 5 provides a graph which demonstrates the effect of variation in $PopSize$ on classification accuracy. Five different population sizes of [5, 10, 20, 30, 50] have been used with the number of iterations being 20 for each population size. With the exception of Ionosphere, KrVsKpEW, SpectEW and Tic-tac-toe, all the other remaining 14 datasets achieve their maximum value at 20 using MA-HS algorithm. Taking this into account, the authors have set the optimal population size as 20 for the MA-HS algorithm.

C. ANALYSIS OF RESULTS

Convergence of the best fitness value with progressive increase in iterations is graphically shown in Figure 7. In several cases, the best fitness obtained by MA and HS is nowhere close to that of MA-HS. Moreover MA-HS achieves its best fitness value at near 10 iterations which is about 50% of the total number of iterations. MA and HS take longer time to achieve their optimal fitness values. Thus MA-HS is better than both MA and HS as it finds out the optimal fitness value and converges quicker. A similar conclusion can also be drawn from Figure 6 where it is evident that the accuracy converges quickly for 80% of the datasets. Here too the final accuracy is achieved by MA-HS within 10 iterations or 50% of the total. So MA and HS help each other thereby increasing the effectiveness of MA-HS over individual MA and HS methods.

A comparison has been given in Table 3 and Table 4 to show the effectiveness of the proposed MA-HS algorithm. Table 3 deals with the classification accuracy and Table 4 with the number of features selected. Both the tables show that the MA-HS method performs far better than the original MA as well as HS. MA-HS achieves the first rank in both the cases. In Table 3, the accuracy of MA-HS is found to be better than MA and HS in all cases giving MA-HS a substantial edge in terms of accuracy. Let us now look at the results of Table 4. MA-HS gives least number of features for 15 out of 18 or 83% of the datasets. Overall, the average rank of MA-HS is by far better than both MA and HS. Comparing MA with HS, it is seen that MA outperforms HS in terms of accuracy but the opposite is true after analyzing the number of features selected. It can be safely inferred that

TABLE 1: Basic information related to the UCI datasets used in the present work

Sl. No.	Dataset	No. of Attributes	Sample Size	No. of Labels	Domain of Dataset
1	Breastcancer	9	699	2	Biology
2	Tic-tac-toe	9	958	2	Game
3	Exactly	13	1000	2	Biology
4	Exactly2	13	1000	2	Biology
5	HeartEW	13	270	2	Biology
6	M-of-n	13	1000	2	Biology
7	WineEW	13	178	3	Chemistry
8	CongressEW	16	435	2	Politics
9	Vote	16	300	2	Politics
10	Zoo	16	101	6	Artificial
11	Lymphography	18	148	2	Biology
12	SpectEW	22	267	2	Biology
13	BreastEW	30	569	2	Biology
14	Ionosphere	34	351	2	Electromagnetic
15	KrVsKpEW	36	3196	2	Game
16	WaveformEW	40	5000	3	Physics
17	Sonar	60	208	2	Biology
18	PenglungEW	325	73	2	Biology

TABLE 2: Parameters along with its corresponding values used in the proposed MA-HS algorithm

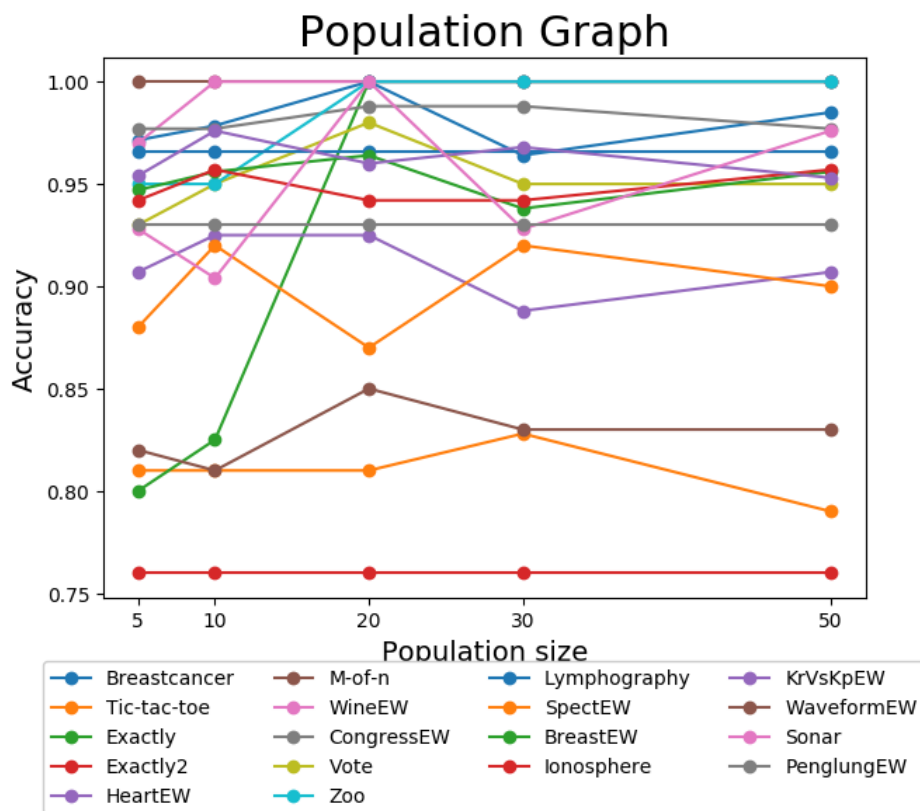
Parameter	Meaning	Value
PopSize	Population size	20
MaxIter	Maximum number of iterations	20
γ	Relative weightage used for fitness value	0.9
a_1	Positive attraction constant	3
a_2	Positive attraction constant	3.5
β	Visibility coefficient	0.1
d_0	Initial nuptial dance coefficient	3
fl_0	Initial random walk coefficient	3
g	Gravitational coefficient	0.98
r_{of}	Random value for crossover	0.95
$MaxHS$	Maximum iterations in HS	30
r_{accept}	Rate of acceptance	0.8
r_{pa}	Rate of pitch adjustment	0.3
b_{range}	Range of pitch adjustment	2

TABLE 3: Comparison of MA, HS and MA-HS algorithms with respect to classification accuracy

Dataset	MA	HS	MA-HS
Breastcancer	0.9785	0.942	0.9929
Tic-tac-toe	0.82	0.776	0.8385
Exactly	0.99	0.685	1
Exactly2	0.76	0.76	0.76
HeartEW	0.87	0.851	0.926
M-of-n	1	0.75	1
WineEW	0.97	0.916	1
CongressEW	0.977	0.954	1
Vote	0.983	0.933	1
Zoo	1	1	1
Lymphography	0.86	0.833	0.9667
SpectEW	0.87	0.796	0.907
BreastEW	0.94	0.956	0.9824
Ionosphere	0.91	0.9	0.98
KrVsKpEW	0.971	0.938	0.974
WaveformEW	0.813	0.788	0.835
Sonar	0.928	0.857	1
PenglungEW	1	1	1
Average rank	1.77	2.72	1
Final rank	2	3	1

TABLE 4: Comparison of MA, HS and MA-HS algorithms with respect to the number of selected features

Dataset	MA	HS	MA-HS
Breastcancer	2	3	3
Tic-tac-toe	5	4	4
Exactly	7	3	6
Exactly2	1	3	1
HeartEW	4	5	4
M-of-n	6	3	6
WineEW	4	3	2
CongressEW	3	6	3
Vote	5	5	3
Zoo	10	10	5
Lymphography	8	7	6
SpectEW	9	5	5
BreastEW	6	10	5
Ionosphere	12	12	3
KrVsKpEW	21	17	13
WaveformEW	17	16	13
Sonar	28	20	16
PenglungEW	150	165	63
Average rank	2.16	2	1.22
Final rank	3	2	1

FIGURE 5: Graph showing the variation of population sizes on classification accuracy obtained by the proposed **MA-HS** algorithm for 18 UCI datasets

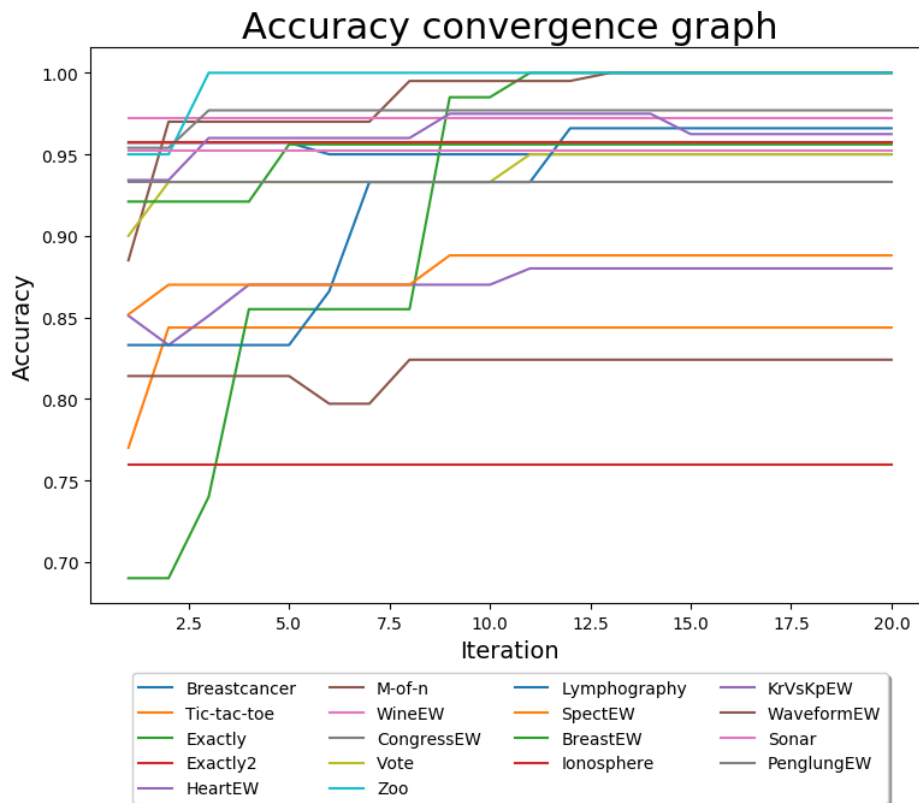


FIGURE 6: Graph showing the variation of classification accuracy obtained using the **MA-HS** algorithm (considering MaxIter=20) for 18 UCI datasets

in MA-HS, both MA and HS mutually benefit from each other and ultimately it improves the classification accuracy as well as the number of features. The sole MA method is strengthened by incorporating HS and vice versa, and the combined MA-HS algorithm is an effective technique for solving FS problems.

VI. COMPARISON AND DISCUSSION

The proposed method is compared with 12 state-of-the-art meta-heuristic FS methods for demonstrating its performance. The other methods described below are tried-and-tested algorithms and have previously been used for the 18 benchmark UCI datasets. The other given meta-heuristics are GA, BPSO [78], ALO, BGSA, binary Dragonfly Algorithm (BDA) [74], binary Salp Swarm Algorithm (BSSA) [79], Whale Optimization Algorithm (WOA) [80], BGWOPSO [81], Embedded Chaotic Whale Survival Algorithm (ECWSA) [82], binary Grasshopper Optimization Algorithm (BGOA) [83], WOASAT [84] and bALO-QR [83]. Among these, BGWOPSO is a hybrid of Grey Wolf Optimization (GWO) [85] and PSO, WOASAT stands for WOA with SA, and bALO-QR is binary ALO hybridized with QuickReduct (QR) [86]. The parameter settings of the other algorithms used for comparison is given in Table 5. Let us first look at the results of Table 6. It can be seen that the MA-HS algorithm performs way better than its peers. The average rank of the

proposed algorithm is better than the second algorithm i.e., BDA by a considerable margin. Its average rank is 1.6 while BDA has an average rank of 2.4. In 13 out of 18 or 72% of the datasets, the proposed MA-HS algorithm gives the best accuracy. For the remaining 5, it achieves the second rank for 3 of them. Only Exactly2 and KrVsKpEW datasets have a worse rank.

Now let us analyze the results of Table 7. For 11 out of 18 datasets or 61% of the datasets, the proposed MA-HS method gives the best result, i.e., least number of features. It obtains the second rank for 6 of the remaining 7 datasets and gets third rank for the other KrVsKpEW dataset. Therefore MA-HS comes within second rank for 17 out of 18 or 95% of the datasets. As a whole, MA-HS algorithm outperforms all the compared algorithms by a wide margin and is assigned the first rank.

From both Table 6 and Table 7, it can be seen that the proposed MA-HS algorithm is successfully able to increase the classification accuracy as well as reduce the number of features. The versatility of the method is evident as it outperforms many state-of-the-art meta-heuristic FS algorithms for all kinds of datasets, be it small or large. The MA-HS algorithm is able to achieve this performance because of the combination of its underlying characteristics through hybridization. The exploratory ability of MA coupled with exploitative ability of HS algorithm gives it an edge over

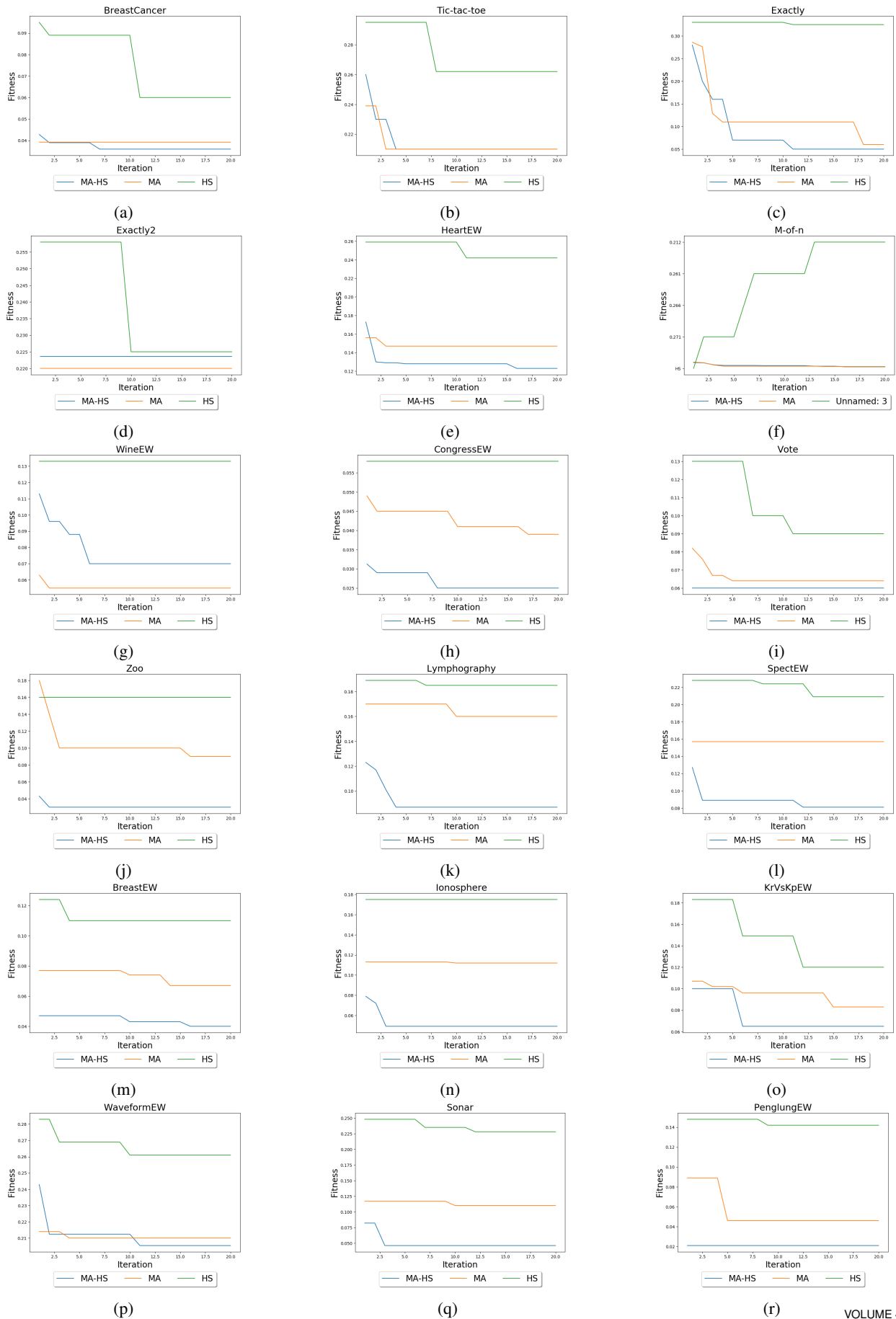


FIGURE 7: Convergence graphs obtained using MA, HS and MA-HS algorithms for 18 UCI datasets

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see <https://creativecommons.org/licenses/by/4.0/>

TABLE 5: Parameter settings of 12 state-of-the-art meta-heuristic FS methods used for comparison

Algorithm	PopSize	MaxIter
GA	8	70
BPSO	8	70
ALO	8	200
BGSA	8	20
BDA	8	20
BSSA	10	100
WOA	10	100
BGWOPSO	10	100
ECWSA	10	100
BGOA	8	20
WOASAT	10	100
bALO-QR	10	100

the others. In fact, the average rank obtained by the MA-HS algorithm in Table 7 is substantially less making it a highly suitable algorithm for FS.

A. STATISTICAL SIGNIFICANCE TEST

In order to find out the statistical significance of the results, Friedman test [87, 88] has been performed. It is a non-parametric statistical test and the procedure involves ranking each row, and then considering the values of ranks by columns. The final result gives a probability or p -value. The null hypothesis states that the two sets of results have the same distribution. If the two results are statistically significant, then the p -value will be less than 0.05 which is a universal threshold. If this comes out to be true then the null hypothesis will be rejected. From the results given in Table 8, it can be concluded that the results are statistically significant as the p -values are less than 0.05 for each and every case.

VII. EXPERIMENTS ON HIGH DIMENSIONAL DATA

The authors have also performed experiments on high dimensional microarray datasets to demonstrate the effectiveness of the proposed method. Microarray data represent the states of a cell at the molecular level and is used in the field of medical research for cancer classification. These datasets typically have a huge number of features (greater than 1000). The proposed method has been used on three publicly available microarray datasets, the details of which have been given in Table 9. The obtained results have been compared with state-of-the-art methods: GA, PSO, GSA, SSA, ALO and HHO.

Table 10 shows the classification accuracy obtained by the proposed method and the state-of-the-art algorithms on the three microarray datasets and the average accuracy obtained. Table 11 demonstrates the comparison of the number of selected features for each of the three datasets and the average number of selected features. It is quite evident that on average the proposed method outperforms the other algorithms in terms of classification accuracy, while selecting fewer features, as shown in Figure 8 and Figure 9. For the Leukaemia2 dataset, binary SSA achieves a slightly higher accuracy than MA-HS, but selects a much greater number

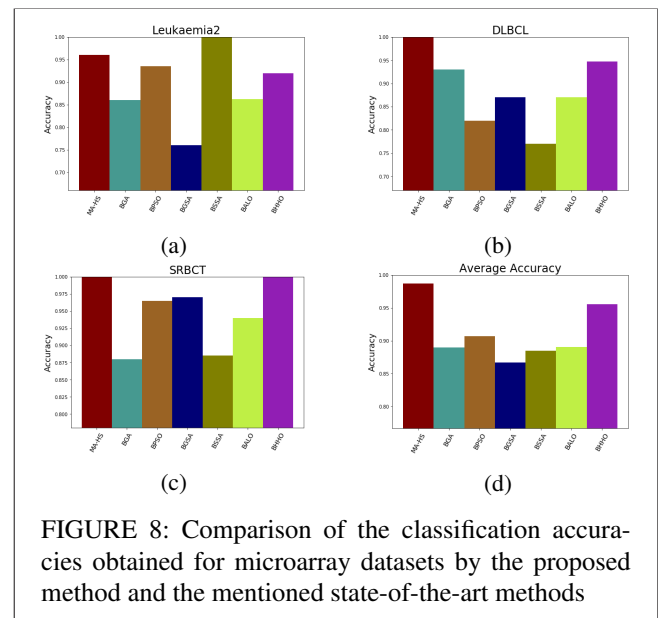


FIGURE 8: Comparison of the classification accuracies obtained for microarray datasets by the proposed method and the mentioned state-of-the-art methods

of features. Binary GSA selects a slightly less number of features but with far lower accuracy. In all other cases, the MA-HS algorithm proves to be clearly superior to the rest of the methods. These experiments on high dimensional data further establish the robustness of the proposed method.

VIII. CONCLUSION AND FUTURE DIRECTION

In this work, a novel meta-heuristic feature selection (FS) algorithm called Mayfly-Harmony Search (MA-HS) has been developed which is based on Mayfly Algorithm (MA) [15] and Harmony Search (HS) [14]. To the best of the author's knowledge, the proposed hybridization is a new one, and has never been used for solving FS problems in the past. The authors have taken the help of S-shaped transfer function (Equation 12) to change the continuous search space to a binary one consisting of only 0's and 1's. MA has been chosen due to its high exploration ability and enhanced exploitation ability is achieved by including HS as a reliable hybrid method. FS is a binary optimization problem and the fitness function is designed as such to maximize the classification

TABLE 6: Comparison of the proposed MA-HS based FS algorithm with some state-of-the-art FS methods in terms of **classification accuracy**

Dataset	MA-HS	GA	BPSO	ALO	BGSA	BDA	BSSA	WOA	BGWOPSO	ECWSA	BGOA	WOASAT	bALO-QR
Breastcancer	0.9929	0.974	0.963	0.974	0.9686	0.9928	0.9768	0.9571	0.98	0.9518	0.9737	0.97	0.974
Tic-tac-toe	0.8385	0.7996	0.7996	0.783	0.7766	0.8469	0.8205	0.7511	0.81	0.7878	0.8038	0.79	0.8
Exactly	1	1	1	0.965	0.994	1	0.9803	0.75	1	0.7811	0.9999	1	0.912
Exactly2	0.76	0.77	0.768	0.762	0.77	0.7725	0.7582	0.69	0.76	0.7912	0.7565	0.75	0.76
HeartEW	0.926	0.8741	0.837	0.838	0.8296	0.8759	0.8605	0.76	0.85	0.8556	0.8635	0.85	0.884
M-of-n	1	1	1	0.967	0.994	1	0.9918	0.85	1	0.9213	1	1	1
WineEW	1	0.9888	0.9775	0.972	0.9775	1	0.9933	0.928	1	0.9802	0.9985	0.99	1
CongressEW	1	0.9679	0.9633	0.981	0.9633	0.9866	0.9628	0.929	0.98	0.9619	0.9772	0.98	0.972
Vote	1	0.9733	0.96	0.972	0.96	0.9894	0.9511	0.9387	0.97	0.95	0.9484	0.97	0.948
Zoo	1	0.902	0.9608	0.98	0.9804	1	1	0.9647	1	0.98	0.9778	0.97	0.961
Lymphography	0.9667	0.8378	0.8919	0.917	0.8649	0.9922	0.89	0.785	0.92	0.8739	0.8586	0.89	0.886
SpectEW	0.907	0.8955	0.8881	0.899	0.8433	0.8519	0.8361	0.787	0.88	0.7988	0.8565	0.88	0.9
BreastEW	0.9824	0.9754	0.9719	0.974	0.9544	0.9792	0.9484	0.9553	0.97	0.9733	0.9606	0.98	0.962
Ionosphere	0.9857	0.9489	0.9489	0.904	0.9432	0.9911	0.9182	0.89	0.95	0.8672	0.922	0.96	0.869
KrVsKpEW	0.974	0.985	0.9731	0.973	0.9549	0.9794	0.9644	0.9151	0.97	0.9392	0.9708	0.98	0.975
WaveformEW	0.835	0.7836	0.756	0.797	0.7344	0.758	0.7335	0.712	0.8	0.7985	0.7562	0.76	0.894
SonarEW	1	0.9904	0.9423	0.845	0.9135	0.9841	0.9372	0.8543	0.96	0.7638	0.9497	0.97	0.84
PenglungEW	1	0.9189	0.9189	0.827	0.8333	1	0.8775	0.729	0.96	0.8766	0.7712	0.94	0.665
Average Rank	1.6	4.5	6.1	6	7.5	2.4	6.6	10.3	4	7.7	6.3	4.7	6.2
Assigned Rank	1	4	6	5	10	2	9	12	3	11	8	5	7

TABLE 7: Comparison of the proposed MA-HS based FS algorithm with some state-of-the-art FS methods in terms of **number of features selected**

Dataset	MA-HS	GA	BPSO	ALO	BGSA	BDA	BSSA	WOA	BGWOPSO	ECWSA	BGOA	WOASAT	bALO-QR
Breastcancer	3	4	4	4.7	4	5	3.8	5.4	4.4	4.5	4.2	4.2	4.05
Tic-tac-toe	4	5	6	5	4	7	6	10.8	5.2	7.74	6	5.2	6.48
Exactly	6	6	6	5.75	4	6	7.2	6	6	7.1	6	6	5.85
Exactly2	1	1	1	1.5	1	7.1	2.7	5.7	1.6	9	5.4	2.8	5.85
HeartEW	4	5	3	8.6	3	5.7	5.8	8.6	5.8	9.4	7.4	5.4	5.8
M-of-n	6	6	6	6	5	6	7.1	9.7	6	5	6	6	5.85
WineEW	2	4	5	5.4	4	3.6	6.3	8.8	6	6	6.8	6.4	5.46
CongressEW	3	2	3	6.6	4	5.5	5.7	10.3	4.4	5.6	4.1	6.4	4.32
Vote	3	5	3	6.6	4	3.4	4.8	7.6	3.4	6	4.9	5.2	5.6
Zoo	5	4	5	5.7	6	4.4	6.7	9.9	6.8	8	5.4	5.6	5.92
Lymphography	6	5	5	7.3	6	8.2	10.3	10.5	9.2	7.7	7.4	7.2	9
SpectEW	5	5	6	7.6	5	6.8	10.9	11.2	8.4	7.8	8.7	9.4	7.7
BreastEW	5	8	9	13.8	10	11.5	16.7	21	13.6	15	13.2	11.6	12.6
IonosphereEW	3	7	7	11.7	9	11.5	15.8	21.4	13	9.52	9.1	12.8	13.6
KrVsKpEW	13	11	12	16.1	14	20.7	20.4	28	15.8	13	16.9	18.4	14.4
WaveformEW	13	15	15	20.5	14	23	22.9	33.2	14.2	14	21.5	20.6	20.8
SonarEW	16	19	22	26.6	24	25.6	33.4	43.4	31.2	20	30.5	26.4	24
PenglungEW	63	84	130	133.1	140	121.2	171.6	144.3	130.8	65	95.7	127.4	130
Average Rank	1.6	2.4	2.8	6.2	3.1	5.7	8	10.2	6.2	6.3	6	6.1	6.05
Assigned Rank	1	2	3	9	4	5	11	12	9	10	6	8	7

TABLE 8: p -values obtained using Friedman statistical test considering classification accuracy of the proposed MA-HS algorithm over all other FS methods for 18 UCI datasets

MA-HS	MA	HS	GA	BPSO	ALO	BGSA	BDA	BSSA	WOA	BGWOPSO	ECWSA	BGOA	WOASAT	bALO-QR
MA-HS	0.001	0	0.004	0.001	0	0	0.04	0	0	0.002	0	0	0.001	0

TABLE 9: **Description of the microarray datasets** used in the present work

Sl. No.	Dataset	No. of Instances	No. of Attributes	No. of Classes	Domain
1	Leukaemia2	72	11225	3	Biology
2	DLBCL	77	5469	2	Biology
3	SRBCT	82	2308	4	Biology

TABLE 10: Comparison of the proposed MA-HS based FS algorithm with some state-of-the-art FS methods in terms of **classification accuracy** on microarray datasets

Dataset	MA-HS	BGA	BPSO	BGSA	BSSA	BALO	BHHO
Leukaemia2	0.9613	0.9352	0.8605	0.7611	1	0.8623	0.9223
DLBCL	1	0.93	0.82	0.87	0.77	0.87	0.947
SRBCT	1	0.88	0.9652	0.97	0.885	0.94	1
Average Rank	1.33	4.67	4	4.33	4.33	4.67	2.33
Assigned Rank	1	3	5	4	4	3	2

TABLE 11: Comparison of the proposed MA-HS based FS algorithm with some state-of-the-art FS methods in terms of **number of selected features** on microarray datasets

Dataset	MA-HS	BGA	BPSO	BGSA	BSSA	BALO	BHHO
Leukaemia2	1508	5111	5198	1472	5228	5073	3979
DLBCL	1147	2452	2505	2558	2802	3476	1741
SRBCT	497	949	865	3872	1042	1071	798
Average Rank	1.33	4	4.33	4.33	6	6.67	2.33
Assigned Rank	1	3	4	4	5	6	2

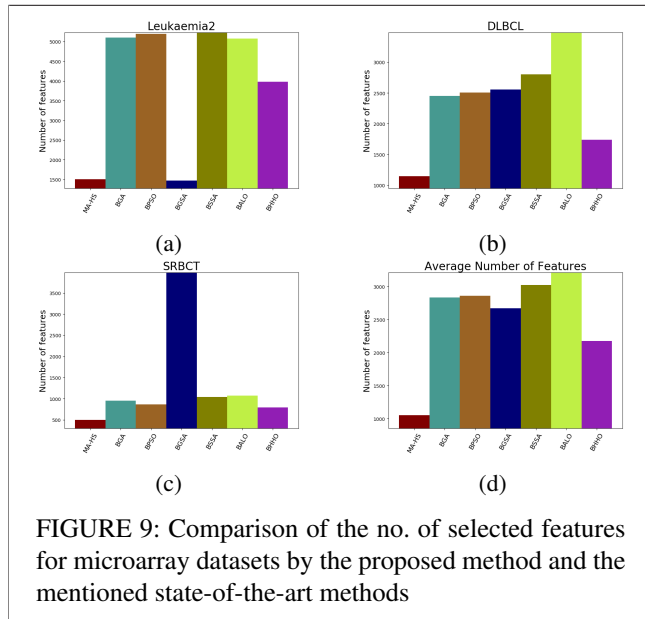


FIGURE 9: Comparison of the no. of selected features for microarray datasets by the proposed method and the mentioned state-of-the-art methods

accuracy and minimize the number of features selected. Furthermore, MA-HS has been applied on 18 UCI datasets and 3 high-dimensional microarray datasets, and compared with many well-known meta-heuristics. It proves that MA-HS is able to obtain a desired high classification accuracy and low number of features in comparison to the other methods. MA-HS converges quickly to an optimal value as shown in Figure 7 and Figure 6. It may be noted that the algorithm can perform even better by optimizing the set of values of the parameters. The parameter values have been tuned by experimenting with a range of values and these have also been discussed in subsection V-B. The set of parameters can be further optimized but will require more experimentation. According to the No Free Lunch Theorem [13], the algorithm may not give the best results for all types of problems. This is evident as the MA-HS algorithm does not give the best results for all the datasets. The algorithm is sometimes prone to premature convergence for datasets like Exactly2. As a future scope of this work, more advanced classifiers may be used for determining the fitness value. The mutation process may be strengthened which will increase the exploration ability of the algorithm. Other transfer functions such as V-shaped or U-shaped can be used here. The meta-heuristic can be combined with other state-of-the-art methods. The proposed work can be used in various fields such as text mining,

bio-informatics such as genomics and neuro-imaging, image processing and for industrial purposes such as fault diagnosis.

References

- [1] Kittipat Kampa et al. "Sparse optimization in feature selection: application in neuroimaging". In: *Journal of Global Optimization* 59.2-3 (2014), pp. 439–457.
- [2] Mairead L Bermingham et al. "Application of high-dimensional feature selection: evaluation for genomic prediction in man". In: *Scientific reports* 5 (2015), p. 10312.
- [3] Alan Jović, Karla Brkić, and Nikola Bogunović. "A review of feature selection methods with applications". In: *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*. Ieee. 2015, pp. 1200–1205.
- [4] Igor Kononenko. "Estimating attributes: Analysis and extensions of RELIEF". In: *Machine Learning: ECML-94*. Springer Berlin Heidelberg, 1994, pp. 171–182. DOI: [10.1007/3-540-57868-4_57](https://doi.org/10.1007/3-540-57868-4_57).
- [5] R. Leardi. "Genetic Algorithms in Feature Selection". In: *Genetic Algorithms in Molecular Modeling*. Elsevier, 1996, pp. 67–86. DOI: [10.1016/b978-012213810-2/50004-9](https://doi.org/10.1016/b978-012213810-2/50004-9).
- [6] Iñaki Inza et al. "Filter versus wrapper gene selection approaches in DNA microarray domains". In: *Artificial intelligence in medicine* 31.2 (2004), pp. 91–103.
- [7] Sebastián Maldonado and Julio López. "Dealing with high-dimensional class-imbalanced datasets: Embedded feature selection for SVM classification". In: *Applied Soft Computing* 67 (2018), pp. 94–105.
- [8] Qasem Al-Tashi et al. "Approaches to Multi-Objective Feature Selection: A Systematic Literature Review". In: *IEEE Access* (2020).
- [9] Manosij Ghosh et al. "Feature selection using histogram-based multi-objective GA for handwritten Devanagari numeral recognition". In: *Intelligent engineering informatics*. Springer, 2018, pp. 471–479.
- [10] Ritam Guha et al. "Deluge based Genetic Algorithm for feature selection". In: *Evolutionary intelligence* (2019), pp. 1–11.
- [11] Manosij Ghosh et al. "Recursive memetic algorithm for gene selection in microarray data". In: *Expert Systems with Applications* 116 (2019), pp. 172–185.
- [12] 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 (2019), pp. 159–176.
- [13] 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. DOI: [10.1109/4235.585893](https://doi.org/10.1109/4235.585893).
- [14] Zong Woo Geem, Joong Hoon Kim, and Gobichetti-palayam Vasudevan Loganathan. "A new heuristic optimization algorithm: harmony search". In: *simulation* 76.2 (2001), pp. 60–68.
- [15] Konstantinos Zervoudakis and Stelios Tsafarakis. "A mayfly optimization algorithm". In: *Computers & Industrial Engineering* 145 (July 2020), p. 106559. DOI: [10.1016/j.cie.2020.106559](https://doi.org/10.1016/j.cie.2020.106559). URL: <https://doi.org/10.1016/j.cie.2020.106559>.
- [16] Kenneth Sörensen and Fred Glover. "Metaheuristics". In: *Encyclopedia of operations research and management science* 62 (2013), pp. 960–970.
- [17] Eric Bonabeau et al. *Swarm intelligence: from natural to artificial systems*. 1. Oxford university press, 1999.
- [18] Hans-Georg Beyer and Hans-Paul Schwefel. In: *Natural Computing* 1.1 (2002), pp. 3–52. DOI: [10.1023/a:1015059928466](https://doi.org/10.1023/a:1015059928466). URL: <https://doi.org/10.1023/a:1015059928466>.
- [19] John R Koza and John R Koza. *Genetic programming: on the programming of computers by means of natural selection*. Vol. 1. MIT press, 1992.
- [20] 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.
- [21] Marco Dorigo, Mauro Birattari, and Thomas Stutzle. "Ant colony optimization". In: *IEEE computational intelligence magazine* 1.4 (2006), pp. 28–39.
- [22] Xin-She Yang. "A new metaheuristic bat-inspired algorithm". In: *Nature inspired cooperative strategies for optimization (NICSO 2010)*. Springer, 2010, pp. 65–74.
- [23] Xin-She Yang. "Firefly algorithm, stochastic test functions and design optimisation". In: *International journal of bio-inspired computation* 2.2 (2010), pp. 78–84.
- [24] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. "Optimization by simulated annealing". In: *science* 220.4598 (1983), pp. 671–680.
- [25] Esmat Rashedi, Hossein Nezamabadi-Pour, and Saeid Saryazdi. "GSA: a gravitational search algorithm". In: *Information sciences* 179.13 (2009), pp. 2232–2248.
- [26] A Kaveh and M Khayatizad. "A new meta-heuristic method: ray optimization". In: *Computers & structures* 112 (2012), pp. 283–294.
- [27] Richard A Formato. "Central force optimization". In: *Prog Electromagn Res* 77 (2007), pp. 425–491.
- [28] Esmaeil Atashpaz-Gargari and Caro Lucas. "Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition". In: *2007 IEEE congress on evolutionary computation*. Ieee. 2007, pp. 4661–4667.
- [29] Ali Husseinzadeh Kashan. "League championship algorithm: a new algorithm for numerical function optimization". In: *2009 international conference of soft computing and pattern recognition*. IEEE. 2009, pp. 43–48.
- [30] Naser Ghorbani and Ebrahim Babaei. "Exchange market algorithm". In: *Applied Soft Computing* 19 (2014), pp. 177–187.
- [31] Malti Baghel, Shikha Agrawal, and Sanjay Silakari. "Survey of metaheuristic algorithms for combinatorial optimization". In: *International Journal of Computer Applications* 58.19 (2012).
- [32] Thomas A Feo and Mauricio GC Resende. "Greedy randomized adaptive search procedures". In: *Journal of global optimization* 6.2 (1995), pp. 109–133.
- [33] Nenad Mladenović and Pierre Hansen. "Variable neighborhood search". In: *Computers & operations research* 24.11 (1997), pp. 1097–1100.
- [34] AM Fathollahi Fard and M Hajiaghaei-Keshteli. "Red deer algorithm (RDA); a new optimization algorithm inspired by red deer's mating". In: *International Conference on Industrial Engineering, IEEE., (2016 e)*. 2016, pp. 33–34.
- [35] Muzaffar Eusuff, Kevin Lansey, and Fayzul Pasha. "Shuffled frog-leaping algorithm: a memetic metaheuristic for discrete optimization". In: *Engineering optimization* 38.2 (2006), pp. 129–154.
- [36] W. Siedlecki and J. Sklansky. "A Note On Genetic Algorithms for Large-Scale Feature Selection". In: *Handbook of Pattern Recognition and Computer Vision*. World Scientific, Aug. 1993, pp. 88–107. DOI: [10.1142/9789814343138_0005](https://doi.org/10.1142/9789814343138_0005).
- [37] Jinjie Huang, Yunze Cai, and Xiaoming Xu. "A hybrid genetic algorithm for feature selection wrapper based on mutual information". In: *Pattern Recognition Letters* 28.13 (Oct. 2007), pp. 1825–1844. DOI: [10.1016/j.patrec.2007.05.011](https://doi.org/10.1016/j.patrec.2007.05.011).
- [38] Riccardo Poli, James Kennedy, and Tim Blackwell. "Particle swarm optimization". In: *Swarm Intelligence* 1.1 (Aug. 2007), pp. 33–57. DOI: [10.1007/s11721-007-0002-0](https://doi.org/10.1007/s11721-007-0002-0).
- [39] Lin Shang, Zhe Zhou, and Xing Liu. "Particle swarm optimization-based feature selection in sentiment classification". In: *Soft Computing* 20.10 (Mar. 2016), pp. 3821–3834. DOI: [10.1007/s00500-016-2093-2](https://doi.org/10.1007/s00500-016-2093-2).
- [40] Ibrahim Aljarah et al. "Simultaneous Feature Selection and Support Vector Machine Optimization Using the Grasshopper Optimization Algorithm". In: *Cognitive Computation* 10.3 (Jan. 2018), pp. 478–495. DOI: [10.1007/s12559-017-9542-9](https://doi.org/10.1007/s12559-017-9542-9).
- [41] Seyedali Mirjalili. "The Ant Lion Optimizer". In: *Advances in Engineering Software* 83 (May 2015), pp. 80–98. DOI: [10.1016/j.advengsoft.2015.01.010](https://doi.org/10.1016/j.advengsoft.2015.01.010).

- [42] Esmat Rashedi, Hossein Nezamabadi-Pour, and Saeid Saryazdi. "BGSA: binary gravitational search algorithm". In: *Natural Computing* 9.3 (2010), pp. 727–745.
- [43] Chaokun Yan et al. "Hybrid binary coral reefs optimization algorithm with simulated annealing for feature selection in high-dimensional biomedical datasets". In: *Chemometrics and Intelligent Laboratory Systems* 184 (2019), pp. 102–111.
- [44] Mohamed Abdel-Basset, Weiping Ding, and Doaa El-Shahat. "A hybrid Harris Hawks optimization algorithm with simulated annealing for feature selection". In: *Artificial Intelligence Review* (2020), pp. 1–45.
- [45] Yong Zhang et al. "Binary differential evolution with self-learning for multi-objective feature selection". In: *Information Sciences* 507 (2020), pp. 67–85.
- [46] Yong Zhang, Xian-fang Song, and Dun-wei Gong. "A return-cost-based binary firefly algorithm for feature selection". In: *Information Sciences* 418 (2017), pp. 561–574.
- [47] Ah. E. Hegazy, M.A. Makhlof, and Gh. S. El-Tawel. "Improved salp swarm algorithm for feature selection". In: *Journal of King Saud University - Computer and Information Sciences* 32.3 (Mar. 2020), pp. 335–344. DOI: [10.1016/j.jksuci.2018.06.003](https://doi.org/10.1016/j.jksuci.2018.06.003). URL: <https://doi.org/10.1016/j.jksuci.2018.06.003>.
- [48] Xiao-han Wang et al. "Multi-objective feature selection based on artificial bee colony: An acceleration approach with variable sample size". In: *Applied Soft Computing* 88 (2020), p. 106041.
- [49] Yong Zhang et al. "Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm". In: *Expert Systems with Applications* 137 (2019), pp. 46–58.
- [50] Laith Mohammad Abualigah, Ahamad Tajudin Khader, and Essam Said Hanandeh. "A new feature selection method to improve the document clustering using particle swarm optimization algorithm". In: *Journal of Computational Science* 25 (Mar. 2018), pp. 456–466. DOI: [10.1016/j.jocs.2017.07.018](https://doi.org/10.1016/j.jocs.2017.07.018). URL: <https://doi.org/10.1016/j.jocs.2017.07.018>.
- [51] Laith Mohammad Qasim Abualigah. *Feature selection and enhanced krill herd algorithm for text document clustering*. Springer, 2019.
- [52] Laith Mohammad Abualigah, Ahamad Tajudin Khader, and Essam Said Hanandeh. "Hybrid clustering analysis using improved krill herd algorithm". In: *Applied Intelligence* 48.11 (2018), pp. 4047–4071.
- [53] Daniel E. Hernández et al. "Evolutionary multi-objective visual cortex for object classification in natural images". In: *Journal of Computational Science* 17 (Nov. 2016), pp. 216–233. DOI: [10.1016/j.jocs.2015.10.011](https://doi.org/10.1016/j.jocs.2015.10.011). URL: <https://doi.org/10.1016/j.jocs.2015.10.011>.
- [54] G. Olague and C. Puente. "Honeybees as an Intelligent based Approach for 3D Reconstruction". In: *18th International Conference on Pattern Recognition (ICPR'06)*. Vol. 1. 2006, pp. 1116–1119.
- [55] Oscar E. Perez-Cham et al. "Parallelization of the Honeybee Search Algorithm for Object Tracking". In: *Applied Sciences* 10.6 (Mar. 2020), p. 2122. DOI: [10.3390/app10062122](https://doi.org/10.3390/app10062122). URL: <https://doi.org/10.3390/app10062122>.
- [56] Zong Woo Geem. "Harmony search in water pump switching problem". In: *International Conference on Natural Computation*. Springer. 2005, pp. 751–760.
- [57] Zong Woo Geem. "Optimal scheduling of multiple dam system using harmony search algorithm". In: *International Work-Conference on Artificial Neural Networks*. Springer. 2007, pp. 316–323.
- [58] Kang Seok Lee and Zong Woo Geem. "A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice". In: *Computer methods in applied mechanics and engineering* 194.36-38 (2005), pp. 3902–3933.
- [59] Mahamed GH Omran and Mehrdad Mahdavi. "Global-best harmony search". In: *Applied mathematics and computation* 198.2 (2008), pp. 643–656.
- [60] Liang Li et al. "New procedure for simulating arbitrary slip surface f soil slope in stability analysis". In: *Journal of Hydraulic Engineering* 16.4 (2008), pp. 535–541.
- [61] Liang Li, Shi-Chun Chi, and Xue-Song Chu. "Location of non-circular slip surface using the modified harmony search method based on correcting strategy". In: *Yantu Lixue(Rock and Soil Mechanics)* 27.10 (2006), pp. 1714–1718.
- [62] Zong Woo Geem, Kang Seok Lee, and Yongjin Park. "Application of harmony search to vehicle routing". In: *American Journal of Applied Sciences* 2.12 (2005), pp. 1552–1557.
- [63] Zong Woo Geem and Han Hwangbo. "Application of harmony search to multi-objective optimization for satellite heat pipe design". In: *Proceedings of. Cite-seer*. 2006, pp. 1–3.
- [64] Mohammed Azmi Al-Betar et al. "Gray image enhancement using harmony search". In: *International Journal of Computational Intelligence Systems* 9.5 (2016), pp. 932–944.
- [65] Aditya Panchal and B Tom. "Harmony search optimization for HDR prostate brachytherapy". In: *International Journal of Radiation Oncology• Biology• Physics* 75.3 (2009), S720–S721.
- [66] J.D. Allan and A.S. Flecker. "The mating biology of a mass-swarming mayfly". In: *Animal Behaviour* 37 (Mar. 1989), pp. 361–371. DOI: [10.1016/0003-3472\(89\)90084-5](https://doi.org/10.1016/0003-3472(89)90084-5). URL: [https://doi.org/10.1016/0003-3472\(89\)90084-5](https://doi.org/10.1016/0003-3472(89)90084-5).
- [67] Barbara L. Peckarsky et al. "PREDATOR CHEMICALS INDUCE CHANGES IN MAYFLY LIFE HISTORY TRAITS: A WHOLE-STREAM MANIPULATION". In: *Ecology* 83.3 (Mar. 2002), pp. 612–618.

- DOI: [10.1890/0012-9658\(2002\)083\[0612:pcicim\]2.0.co;2](https://doi.org/10.1890/0012-9658(2002)083[0612:pcicim]2.0.co;2). URL: [https://doi.org/10.1890/0012-9658\(2002\)083\[0612:pcicim\]2.0.co;2](https://doi.org/10.1890/0012-9658(2002)083[0612:pcicim]2.0.co;2).
- [68] R. Eberhart and J. Kennedy. "A new optimizer using particle swarm theory". In: *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*. IEEE. DOI: [10.1109/mhs.1995.494215](https://doi.org/10.1109/mhs.1995.494215). URL: <https://doi.org/10.1109/mhs.1995.494215>.
- [69] David E Goldberg and John Henry Holland. "Genetic algorithms and machine learning". In: (1988).
- [70] Mi Li et al. "A multi-information fusion "triple variables with iteration" inertia weight PSO algorithm and its application". In: *Applied Soft Computing* 84 (2019), p. 105677.
- [71] Wan-qiu Zhang, Yong Zhang, and Chao Peng. "Brain storm optimization for feature selection using new individual clustering and updating mechanism". In: *Applied Intelligence* 49.12 (2019), pp. 4294–4302.
- [72] Bitanu Chatterjee et al. "Late acceptance hill climbing based social ski driver algorithm for feature selection". In: *IEEE Access* 8 (2020), pp. 75393–75408.
- [73] 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](https://doi.org/10.1080/00031305.1992.10475879).
- [74] Majdi M. Mafarja et al. "Binary Dragonfly Algorithm for Feature Selection". In: *2017 International Conference on New Trends in Computing Sciences (ICTCS)*. IEEE, Oct. 2017. DOI: [10.1109/ictcs.2017.43](https://doi.org/10.1109/ictcs.2017.43).
- [75] Guido van Rossum and Fred L. Drake. *The Python Language Reference Manual*. Network Theory Ltd., 2011. ISBN: 1906966141.
- [76] J. D. Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- [77] Dheeru Dua and Casey Graff. *UCI Machine Learning Repository*. 2017. URL: <http://archive.ics.uci.edu/ml>.
- [78] Li-Yeh Chuang et al. "Improved binary PSO for feature selection using gene expression data". In: *Computational Biology and Chemistry* 32.1 (2008), pp. 29–38.
- [79] Hossam Faris et al. "An efficient binary salp swarm algorithm with crossover scheme for feature selection problems". In: *Knowledge-Based Systems* 154 (2018), pp. 43–67.
- [80] Seyedali Mirjalili and Andrew Lewis. "The whale optimization algorithm". In: *Advances in engineering software* 95 (2016), pp. 51–67.
- [81] Q. Al-Tashi et al. "Binary Optimization Using Hybrid Grey Wolf Optimization for Feature Selection". In: *IEEE Access* 7 (2019), pp. 39496–39508. ISSN: 2169-3536. DOI: [10.1109/ACCESS.2019.2906757](https://doi.org/10.1109/ACCESS.2019.2906757).
- [82] Ritam Guha et al. "Embedded chaotic whale survival algorithm for filter-wrapper feature selection". In: *arXiv preprint arXiv:2005.04593* (2020).
- [83] Majdi Mafarja et al. "Binary grasshopper optimisation algorithm approaches for feature selection problems". In: *Expert Systems with Applications* 117 (2019), pp. 267–286.
- [84] Majdi M Mafarja and Seyedali Mirjalili. "Hybrid whale optimization algorithm with simulated annealing for feature selection". In: *Neurocomputing* 260 (2017), pp. 302–312.
- [85] E. Emary, Hossam M. Zawbaa, and Aboul Ella Hassanien. "Binary grey wolf optimization approaches for feature selection". In: *Neurocomputing* 172 (Jan. 2016), pp. 371–381. DOI: [10.1016/j.neucom.2015.06.083](https://doi.org/10.1016/j.neucom.2015.06.083).
- [86] Javad Rahimipour Anaraki and Mahdi Eftekhari. "Improving fuzzy-rough quick reduct for feature selection". In: *2011 19th Iranian Conference on Electrical Engineering*. IEEE. 2011, pp. 1–6.
- [87] Pawan Kumar Singh, Ram Sarkar, and Mita Nasipuri. "Statistical validation of multiple classifiers over multiple datasets in the field of pattern recognition". In: *International Journal of Applied Pattern Recognition* 2.1 (2015), pp. 1–23.
- [88] Pawan Kumar Singh, Ram Sarkar, and Mita Nasipuri. "Significance of non-parametric statistical tests for comparison of classifiers over multiple datasets". In: *International Journal of Computing Science and Mathematics* 7.5 (2016), pp. 410–442.



TRINAV BHATTACHARYYA (email: trinav0711@gmail.com) is currently an undergraduate student of Computer Science and Engineering at Jadavpur University, Kolkata, India. His areas of interest include Machine Learning, Optimization, and Graph Theory.



BITANU CHATTERJEE (email: bitanu.ch@gmail.com) is currently an undergraduate student of Computer Science and Engineering at Jadavpur University, Kolkata, India. His areas of interest include Machine Learning, Optimization, and Graph Theory.



PAWAN KUMAR SINGH (email: pk Singh.it@jadavpuruniversity.in)

received his B. Tech degree in Information Technology from West Bengal University of Technology in 2010. He received his M. Tech in Computer Science and Engineering and Ph.D. (Engineering) degrees from Jadavpur University (J.U.) in 2013 and 2018 respectively. He also received RUSA 2.0 fellowship for pursuing his post-doctoral research in J.U. in 2019. He is currently working as an Assistant Professor in the Department of Information Technology in J.U. He has published more than 50 research papers in peer-reviewed journals and international conferences. His areas of current research interest are Computer Vision, Pattern Recognition, Handwritten Document Analysis, Image & Video Processing, Feature Optimization, Machine Learning and Artificial Intelligence. He is a member of the IEEE (U.S.A.), The Institution of Engineers (India) and Association for Computing Machinery (ACM) as well as a life member of the Indian Society for Technical Education (ISTE, New Delhi) and Computer Society of India (CSI).



RAM SARKAR (email: rsarkar@ieee.org) received his B. Tech degree in Computer Science and Engineering from University of Calcutta in 2003. He received his M.E. degree in Computer Science and Engineering and PhD (Engineering) degree from Jadavpur University in 2005 and 2012 respectively. He joined in the department of Computer Science and Engineering of Jadavpur University as an Assistant Professor in 2008, where he is now working as an Associate Professor. He received Fulbright-Nehru Fellowship (USIEF) for post-doctoral research in University of Maryland, College Park, USA in 2014-15. His areas of current research interest are Image Processing, Pattern Recognition, Machine Learning, and Bioinformatics. He is a senior member of the IEEE, U.S.A.



JIN HEE YOON (email: jin9135@sejong.ac.kr)

received B.S., M.S. and Ph.D. degree in Mathematics from Yonsei University, South Korea. She is currently a faculty of School of Mathematics and Statistics at Sejong University, Seoul, South Korea. Her research interests are fuzzy regression analysis, fuzzy time series, optimizations, intelligent systems and deep learning. She is a board member of KIIS (Korean Institute of Intelligent Systems) and has been working as an associate editor, guest editor and editorial board member of several journals including SCI and SCIE journals. In addition, she has been regularly working as an organizer and committee member of several international conferences.



ZONG WOO GEEM (email: geem@gachon.ac.kr)

is an Associate Professor in Department of Energy IT at Gachon University, South Korea. He has obtained B.Eng. in Chung-Ang University, Ph.D. in Korea University, and M.Sc. in Johns Hopkins University, and researched at Virginia Tech, University of Maryland - College Park, and Johns Hopkins University. 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 has served for various journals as an editor (Associate Editor for Engineering Optimization; Guest Editor for Swarm & Evolutionary Computation, Int. Journal of Bio-Inspired Computation, Journal of Applied Mathematics, Applied Sciences, and Sustainability).