Inteligencia Artificial 25(70), 33-49
doi: 10.4114/intartif.vol25iss70pp33-49

INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

Feature Selection: Binary Harris Hawk Optimizer Based Biomedical Datasets

Hadeel Tariq Ibrahim [1,A], Wamidh Jalil Mazher [2,B] , Enas Mahmood Jassim[3,C]
[1] E-Learning Unit, College of Education for Women, University of Thi-Qar, Thi-Qar, Iraq
[2] Electrical Department, Al-Shatra Technical Institute, Southern Technical University, Basra, Iraq
[3] Computer Sc. Dept., College of Basic Education, University of Diyala, Diyala, Iraq
[A] Hadeel.tariq@utq.edu.iq, [B] wamidh.mazher@stu.edu.iq, [C] enasmahmoud@uodiyala.edu.iq

Abstract. Feature selection (FS) is an essential preprocessing step in utmost solutions for the high-dimensional problem to reduce the number of features by deleting irrelevant and redundant data that preserve a suitable grade of classification accuracy. Feature selection can be treated as an optimization problem. Heuristic optimization algorithms are hopeful approaches to solve feature selection problems because of their difficulty, especially in high-dimensional data. Binary Harris hawk optimization (BHHO) is one of the lately suggested metaheuristic algorithms that has been demonstrated to be used more efficiently in facing some optimization problems. Support vector machines (SVMs) are a vital technique that are employed competently to resolve classification issues. We modified the BHHO algorithm with SVM classifier to solve the feature selection issue. This study suggests BHHO-FS to fix the feature selection problem in biomedical datasets. We ran the proposed approach BHHO-FS on real biomedical datasets with 17 types of cancer for Iraqi patients in 2010-2012. The experimental results demonstrate the supremacy of the proposed BHHO-FS in terms of three performance metrics: feature selection accuracy, runtime and number of selected features compared to four other state-of-art algorithms: Fire Fly (FF) algorithm, Genetic Algorithm (GA), Grasshopper Optimization Algorithm (GOA) and Particle Swarm Algorithm (PSO). Comparative experiments designate the importance of the proposed approach in comparison with the other four mentioned algorithms. The implementation of the proposed BHHO-FS approach on 17 datasets for different types of cancers reveals 99.967% average accuracy.

Keywords: Feature selection, Harris hawk optimizer, Fire fly, Genetic algorithm, Grasshopper optimization algorithm, Particle swarm.

1 Introduction

High-dimensional biomedical datasets comprise a huge number of irrelevant and redundant features. In the case of treating all the features equally, the accuracy of treating is reduced accordingly. Feature selection is an important technique that uses an evaluation criterion to pick a meaningful feature subset in order to reduce the potential of classification model overfitting [1]. Consequently, feature selection is recommended to be a crucial pre-process in diagnosing disease-based high-dimensional biomedical datasets [2]. Feature selection is a pre-processing step that excludes unnecessary and excessive features and specifies only the valuable subset of features. This step leads to higher performance metrics (i.e., accuracy, runtime), especially in critical classification problems such as cancer disease diagnoses [3].

Machine learning algorithms are having difficulty categorizing datasets with a large number of attributes as the dimensionality of public data has increased in recent years. Feature selection is an essential data pre-processing method that can help machine learning algorithms solve such problems [4]. The practice of stating linked features and eliminating unrelated, duplicated, or biased material increases the effectiveness of data mining algorithms.

ISSN: 1137-3601 (print), 1988-3064 (on-line)

Whereas duplicate features provide no new information for existing features, unrelated or irrelevant features provide vital information [5]. Feature selection optimization is an important pre-processing strategy in data mining for developing and removing duplicate and unneeded characteristics[6].

Support vector machines (SVMs) are significant machine learning techniques that are used to handle classification and regression issues[7]. SVM utilization has becoming required as complex applications mature[8]. SVM is a robust machine learning technique for solving classification and regression problems[9]. To forecast SVM kernels, researchers employed several kernel functions. The radial basis function (RBF) is preferable since it only adjusts one parameter [10]. RBF modifies the SVM parameters cost (C) and gamma (γ) [8]. SVM has been utilized in image retrieval in the literature[11], human emotion recognition [12], pattern recognition [13], spam categorization [14], gender classification [15], cancer diagnoses [9] feature selection [16], etc.

We modified the binary Harris hawk optimization algorithm (BHHO) to launch a proposed approach, BHHO-FS. BHHO-FS was executed and analysed to optimize feature selection. By comparing BHHO-FS with four other state-of-art algorithms, we demonstrated its high presentation. The additional algorithms are the Fire Fly (FF) algorithm[17], Genetic Algorithm (GA)[18], Grasshopper Optimization Algorithm (GOA)[19], and Particle Swarm Algorithm (PSO)[20].

The BHHO algorithm works fast because it runs with a speed Levy and greedy choosing [21]. The suggested method, BHHO-FS, is tested on seventeen (17) actual biological datasets from Iraqi cancer patients[22], as listed in Table 1. The proposed BHHO-FS results attained higher feature selection accuracy, lower runtime and fewer selected features compared to the other four algorithms.

Table 1: List of datasets used in experiments

No	Dataset	Number of instances	Number of features
1	Abdomen	471	16
2	Bladder	4288	16
3	Blood	4788	16
4	Bones	950	16
5	Brain	2935	16
6	Breast	10670	16
7	Colon	3258	16
8	Eye	179	16
9	Glands	1655	16
10	Heart	183	16
11	Liver	2842	16
12	Lungs	4984	16
13	Lymph	5448	16
14	Naso	1818	16
15	Nerve	1175	16
16	Skin	1920	16
17	Stomach	2222	16

The rest of this paper is structured as follows: the next section provides an outline of what has been done in the literature on some algorithms that have been employed in feature selection. Section 3 presents the basics of the binary Harris hawk optimizer (BHHO). The proposed BHHO-FS paradigm is discussed in Section 4. In Section 5, the experimental results are presented and analysed. Finally, in Section 6, conclusions and future work are presented.

2 Literature Review

Many heuristic optimization techniques are used in feature selection; a few heuristic optimization algorithms are discussed in this section. [18] suggested and examined the usage of a genetic algorithm for instantaneously first choosing an optimum feature subset and second optimizing support vector regression factors (SVR) to increase the accuracy of the software power estimations. They described tests executed with two datasets of software plans. The simulations in both datasets showed that the suggested GA-based algorithm was capable of considerably improving the SVR performance. [23] modified Differential Evolution (DE) algorithm and proposed DEFS for feature selection. DEFS greatly decreased the computational costs and demonstrated robust performance. The DEFS approach was employed in a brain computer interface (BCI) application and compared with additional dimensionality lessening methods. Their results confirmed the importance of the proposed DEFS by obtaining an optimum solution and using less memory.

The BAT algorithm (BA) is a feature selection method that is inspired by bat activity in conducting routes. BA does not need the use of complex operators like mutation and crossover. It essentially alters bat placements, loudness, and frequency. This method provides accurate classification while also reducing the size of the feature set [24]. In [17], the firefly algorithm (FF) was modified to propose a feature selection system. The modified FF was balanced adaptively to speed up the exploration and exploitation phases and find the optimum solution accordingly.

To create a filter feature selection method, the BCOA (binary cuckoo optimization algorithm) is combined with information theory. It is based on BCOA and the shared information of each pair of characteristics to assess the relevance and repeating in the objective function. It achieves a significant decrease in feature selection in a high-dimensional data collection [25].

By [26], a gray wolf optimizer was employed to find the optimum feature subset. In this paper, a comparison was performed with particle swarm optimization (PSO) and genetic algorithms (GAs) using a set of UCI data repositories. The authors approved the supremacy of the proposed algorithm in both classification accuracy and feature size minimization. Furthermore, the grey wolf optimization algorithm is more powerful than initialization in both PSO and GA optimizers.

The salp swarm algorithm [27] was developed to be used in feature selection. The accuracy and runtime of the proposed SSA-FS are compared with particle swarm optimization and differential evolution. Breast, bladder, and colon cancers for Iraqi patients, as well as synthetic datasets for evaluation, were used in this study. When compared to other selected algorithms, the suggested SSA-FS achieved the highest accuracies with the least amount of runtime.

[19] A Grasshopper Optimization Algorithm improved SVM parameters and selected features (GOA). It validated its capacity to address real-world problems in an unknown search field. GOA's strength rests in its high degree of exploitation, which is led by the interactions of all the individuals in the swarm [28].

3 Binary Harris Hawk Optimizer

The Harris hawk bird is famed due to its distinctive supportive hunting activities altogether with other group members existing in the similar steady group, whereas other birds-of-prey habitually attack to determine a prey, solitary. In the southern half of Arizona, USA, Harris hawk birds were first discovered and renowned birds of prey that live in a rather stable category. Such clever birds can regulate supper celebrations comprising some individuals in nonbreeding feasts [29].

Harris hawks hunt animals primarily by "surprise pounce," often well-known as the "seven kills" tactic. In this clever tactic, some hawks move to hit from various directions while at the same time converging on a believed runaway animal out the covering. This hit can be completed quickly by detaining the astonished prey in a matter of seconds, but depending on the prey's run-away skills, the "seven kills" may consist of multiple short, rapid rushes close to the prey in minutes [30].

3.1 Exploration-phase

Harris hawks in HHO lounge at certain spots by chance and wait to see a hunted rabbit using two techniques. The first strategy is modeled in Eq. (1) with considering an equal probability p for each lounging strategy, they lounge depending on the other family members' locations and the hunted animal (i.e., the rabbit) [30].

$$A(t+1) = \begin{cases} A_{rnd}(t) - rd_1|A_{rnd}(t) - 2rd_2A(t)| & p \geq 0.5 \\ \left(A_{htd}(t) - A_{avg}(t)\right) - rd_3\left(L_{bnd} + rd_4(U_{bnd} - L_{bnd})\right) & p < 0.5 \end{cases} \quad (1)$$

where $A(t+1)$ represents the hawk position vector in the following iteration t, $A_{htd}(t)$ is the hunted rabbit location, $A(t)$ is the present hawk position vector, and rd_1, rd_2, rd_3, rd_4, and p are arbitrary numbers within (0,1) that are modified in every iteration. The upper and lower bounds of the parameters are represented by U_{bnd} and L_{bnd}, respectively. A randomly chosen hawk from the present population is denoted by $A_{rnd}(t)$, where A_{avg} represents the average location of the present hawk population. The average location of hawks is calculated by Eq. (2):

$$A_{avg}(t) = \frac{1}{M}\sum_{i=1}^{M} A_i(t) \quad (2)$$

where $A_i(t)$ represents each hawk location at iteration t and M indicates the entire number of hawks.

3.2 Exploration to exploitation transition

In the HHO algorithm, the transition from exploration to exploitation is done based on the prey escaping energy. The prey energy drops significantly through escape. In this step, the rabbit energy is demonstrated as:

$$P = 2P_0\left(1 - \frac{t}{T_{max}}\right) \quad (3)$$

where the rabbit run-out power is denoted by P, T_{max} is the maximum iteration, and the initial value of the rabbit power is denoted by P_0.

3.3 Exploitation Phase

In this phase, Harris hawk birds achieve the "surprise pounce or seven kills" [37] by launching the purposed prey marked in the exploration phase. However, prey usually try to run in risky situations. Later, diverse hunting styles occurred in actual situations. As stated by the escape conduct of the prey and hunting strategies of Harris hawk birds, four probable strategies are suggested in the HHO algorithm to state the launching stage. By nature, prey always tend to run away from dangerous situations. The opportunity to run away is denoted by rd ; if the prey successfully runs away, $rd < 0.5$; otherwise, $rd \geq 0.5$.

3.3.1 Soft blockade

If $rd \geq 0.5$ and $P \geq 0.5$ this means that the prey as yet has sufficient energy, and attempts to run away using some haphazard tricky rebounds but last, it cannot. Through these tries, Harris's hawks surround it quietly to turn the prey more tired and then do the "surprise pounce". This conduct is demonstrated via the following rubrics:

This action is modeled as following:

$$A(t+1) = \Delta A(t) - P|KA_{htd}(t) - A(t)| \quad (4)$$

$$\Delta A(t) = A_{htd}(t) - A(t) \quad (5)$$

The difference between the hawk's position vector and the current position is denoted by $\Delta A(t)$ in iteration t, rd_5 is an arbitrary number in (0,1), and $K = 2(1 - rd_5)$ indicates the arbitrary bounce force of the hunted animal throughout the run-out scenario. The K value varies randomly in every iteration to mimic the nature of hunted animal movements.

3.3.2 Hard-blockade

Now, $rd \geq 0.5$ and $P \leq 0.5$, which means that the prey is so tired and has a little run-away energy. Furthermore, Harris hawks strongly surround purposed prey to last achieve the "surprise pounce". In this case, the present locations are modified using the following equation:

$$\Delta A(t+1) = A_{htd}(t) - P|\Delta A(t)| \quad (6)$$

3.3.3 Soft-blockade with advanced quick plunges

Once P ≥0.5 but rd<0.5, the prey has sufficient energy to fruitfully escape, and a soft-blockage is established prior to the "surprise pounce." This method is smarter than the earlier one. In [30], the authors assumed that the hawks have the ability to adopt their subsequent step depending on the following law in Eq. (7) to execute the soft:

$$B = A_{htd}(t) - P|KA_{htd}(t) - A(t)| \qquad (7)$$

They assumed that the hawks would plunge depending on the LF-based forms utilizing the following law:

$$C = B + S_{rdm} \times LF(G) \qquad (8)$$

where G is the problem space, S_{rdm} is a arbitrary vector of size $1 \times F$ and LF is the Levy-flight function, that is considered by Eq. (9): blockade

$$LF(a) = 0.01 \times \frac{l \times \mu}{|m|^{\frac{1}{\alpha}}}, \alpha = \left(\frac{\theta(1+\alpha) \times sin\left(\frac{\pi\alpha}{2}\right)}{\theta\left(\frac{1+\alpha}{2}\right) \times \alpha \times 2\left(\frac{\alpha-1}{2}\right)} \right)^{\frac{1}{\alpha}} \qquad (9)$$

where l and m are arbitrary values in (0,1), and α is the default constant adjusted to 1.5.

As a result, Eq. (10) can represent the final approach for changing hawk positions during the soft blockage stage:

$$A(t+1) = \begin{cases} B \; if \; F(B) < F(A(t)) \\ C \; \; if \; F(C) < F(A(t)) \end{cases} \qquad (10)$$

where B and C are got by Eqs. (7) and (8), respectively.

3.3.4 Hard blockade with advanced quick plunges

When $P < 0.5$ and $rd < 0.5$, the prey does not have sufficient energy to run away, and a hard blockade is made before the "surprise pounce" to capture and murder the prey. The case of this phase in the prey side is similar to that in the soft blockade, but here, the hawks attempt to reduce the distance of their average position with the run-away prey. Thus, the following law is employed in the hard blockade case:

$$A(t+1) = \begin{cases} B \; if \; F(B) < F(A(t)) \\ C \; \; if \; F(C) < F(A(t)) \end{cases} \qquad (11)$$

where B and C are got by new rules in Eqs. (12) and (13).

$$B = A_{htd}(t) - P|KA(t) - A_n(t)| \qquad (12)$$

$$C = B + S_{rdm} \times LF(G) \qquad (13)$$

where $A_n(t)$ is got utilizing Eq. (2).

3.4 HHO Pseudocode

The pseudocode of HHO algorithm is listed below[30]:

HHO Pseudocode

Inputs: The size of population is N with maximum iterations T
Outputs: The hunted animal (rabbit) position with its fitness value
Initialize the arbitrary population A_i(i=1,2,……,N)
While (termination condition is not met) do
  Compute the hawks' fitness values
  Put A_{htd} as the best position for hunted rabbit
  For (every hawk (A_i) do
    Modify the initial power P_0 and bounce force K
    Modify P utilizing Eq. (3)
    If ($\lvert P\rvert \geq 1$) then (Exploration phase)
      Modify the position vector utilizing Eq. (1)
    If ($\lvert P\rvert < 1$) then (Exploitation phase)
      If ($rd \geq 0.5$ and $\lvert P\rvert \geq 0.5$) then (Soft blockade)
        Modify the position vector utilizing Eq. (4)
      Else if ($rd \geq 0.5$ and $\lvert P\rvert < 0.5$) then (Hard blockade)
        Modify the position vector utilizing Eq. (6)
      Else if ($rd < 0.5$ and $\lvert P\rvert \geq 0.5$) then
        Modify the position vector utilizing Eq. (10)
      Else if ($rd < 0.5$ and $\lvert P\rvert < 0.5$) then
        Modify the position vector utilizing Eq. (11)
 	
  ReturnA_{htd}

4 The proposed BHHO-FS paradigm

The main goal of the proposed BHHO-FS is to develop a feature selection accuracy rate. Here, not only collecting features in high-dimensional datasets requires time and money but also redundant information consequences in wasting time during classification. Accordingly, it is better to lessen the number of features to obtain a quick response and to find a good relationship between the features and the results.

In general, every wrapper feature selection strategy is built around three key components: a search procedure, an induction mechanism, and an assessment calculation [31]. The BHHO method is used as a search technique in our suggested approach, the BHHO-FS, to find the best feature subset. Support Vector Machine is employed as an induction algorithm, with computations for classification accuracy utilized as an evaluation calculation. Figure 1 depicts a high-level schematic of wrapper feature selection together with a basic simulation of our suggested technique, BHHO-FS.

Three factors must be considered while developing the BHHO-FS paradigm: encoding characteristics, the objective function, and system architecture. Such concerns will be discussed in depth in the following sections.

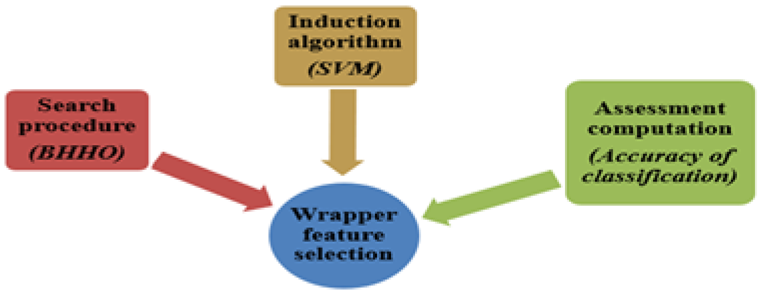

Figure 1: The feature selection method components and their correspondences in the planned BHHO-FS algorithm

4.1 Encoding features

The first step in the BHHO-FS paradigm is encoding the entered features (elements) in a vector formula. Such a vector is employed by the entered features to be optimized later until choosing the optimum (minimum) subset features. First, the entered features should be normalized to be in the [0, 1] period using the following equation [32]:

$$FB = \frac{FA - min_{FA}}{max_{FA} - min_{FA}} \qquad (14)$$

If the FB value was larger than or equal to 0.5 after applying Eq. (14), the value within the vector was adjusted to 1, and this feature was picked; else, it was put to 0, and such a feature was not picked. Figure 2 depicts the encoding feature processes.

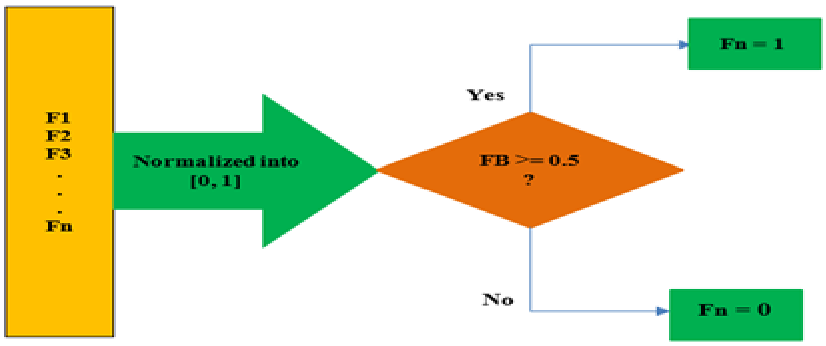

Figure 2: Encoding features steps

4.2 Objective function

The objective function is needed in wrapper feature selection to assess the specific solution. The main aim of feature selection is to improve the accuracy of prediction and consequently minimize the number of selected features. In each selection of our proposed BHHO-FS system, the objective function is used based on calculation accuracy, as shown in Eq. (15) [33]:

$$Accuracy = \frac{True_p + True_N}{True_p + False_N + False_p + True_N} \qquad (15)$$

Where:

$True_p$: the total correct predictions and real class is true.

$True_N$: the total correct predictions and real class are false.

$False_N$: the total incorrect predictions and real class is true.

$False_p$: the total incorrect predictions and real class are false.

4.3 System plan

The plan of the suggested system, BHHO-FS, is designated in this section, and its key parts are itemised below:

- **Data Normalization:** This is a public preceding processing act in feature selection. Features are adjusted to be restricted to the [0,1] range. Normalization is used to mitigate the terrible impact of misleading values of a few features; it was separated by identifying the picked feature depending on the FB value in Eq. (14).

- **Establishing training and testing sets**: Each of our biomedical datasets was divided into two parts: training and testing. The suggested BHHO-FS technique employed 80% of the whole dataset as the training set, with the remaining 20% used as the testing set. To create the model, we used the support vector machine (SVM) classifier on the training and testing sets[34].

- **Specifying a subset of attributes:** here, the 1's value attributes (features) have been picked from the training set.

- **Fitness assessment:** The coordinates of the predefined training set were used for SVM classifier learning purposes, and the performance of the classifier was determined by Eq(15).

- **Ending condition:** By deciding the top iteration, the entire process has been halted. In practice, the highest iteration was fixed to 5.

The proposed BHHO-FS plan is depicted in Fig. 3, which reveals the relationships among the system key parts.

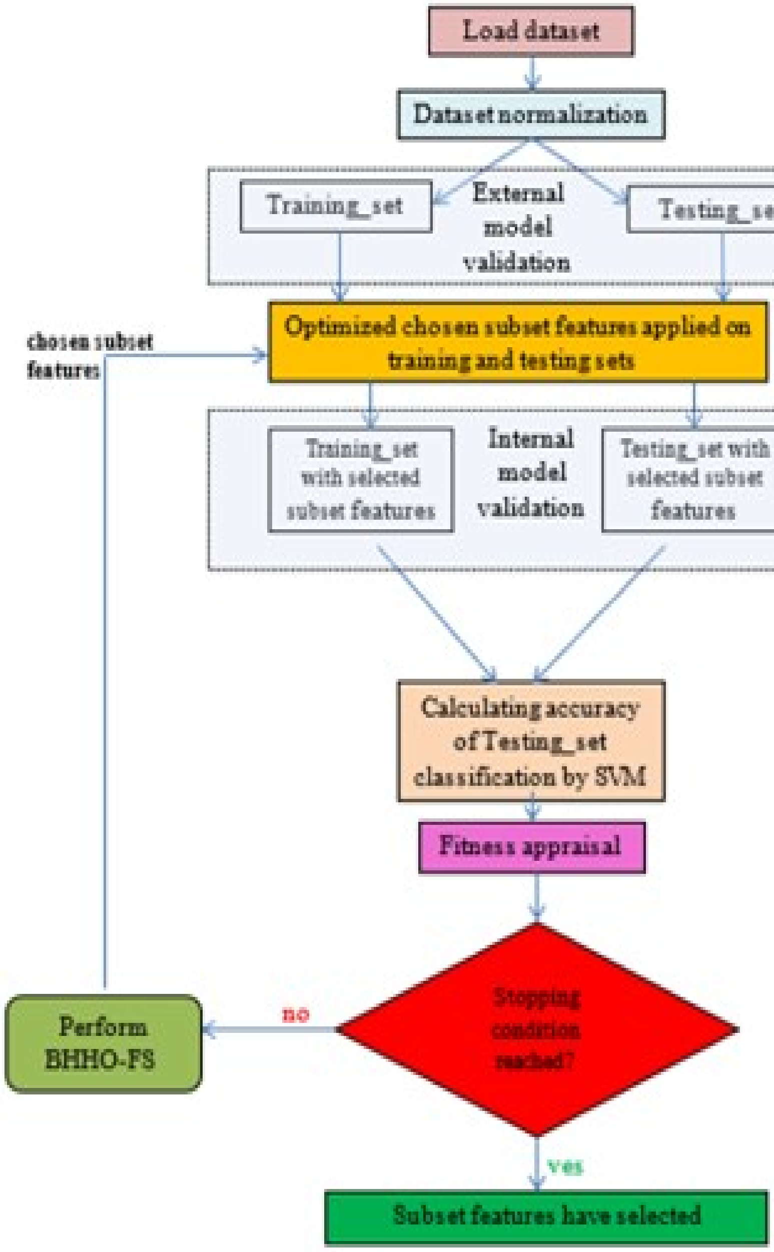

Figure 3: Proposed BHHO-FS workflow

5 Experimental results and analysis

This section summarizes the results of all tests. The BHHO-FS algorithm was compared with four other state-of-art algorithms: FF, GA, GOA and PSO. The comparisons between BHHO-FS and the other mentioned algorithms depended on three metrics, as listed below:

- Feature selection accuracy
- Run time (minutes: seconds: milliseconds)
- Number of selected features

We ran our proposed approach using the following:

1- MATLAB R2015a
2- Windows 10.

3- Intel(R) Core(TM) i7-5500U CPU with 2.40 GHz and 8 GB RAM

5.1 Datasets description

Actual Iraqi cancer patients' datasets (2010- 2012) years were used in this study [22]. For all forms of cancer, such datasets are gathered from all Iraqi governorates' hospitals. After removing superfluous includes and unfairness values, the last datasets comprised 16 features with varying amounts of occurrences. The details of the employed datasets are listed in Table 1 and depicted in Fig. 4.

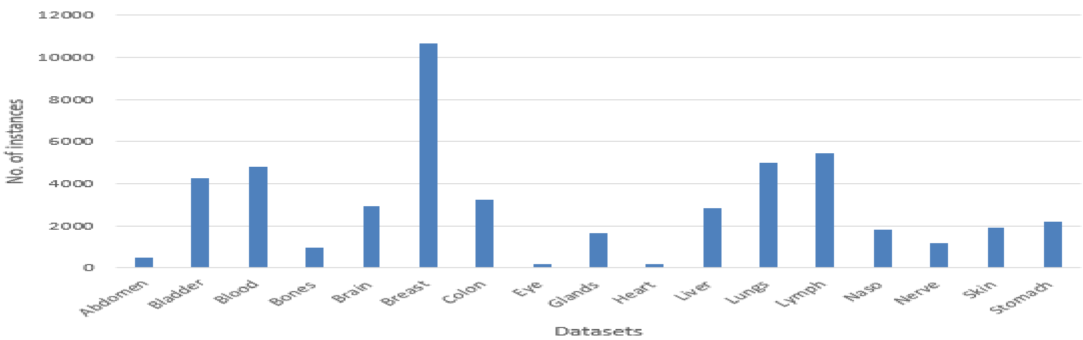

Figure 4: used datasets in experiments

5.2 Comparisons of BHHO-FS with other state-of-art algorithms: FF, GA, GOA and PSO

Feature selection accuracy: Table 2 shows the comparisons of feature selection accuracy between BHHO-FS and the other four state-of-art algorithms with 5 iterations by each algorithm. Furthermore, the SVM classifier is employed in such a comparison without any optimization. Table 2 is depicted by Fig. 5. It is evident that BHHO-FS outperformed other optimization algorithms over 14 datasets out of 17 datasets in feature selection accuracy (100%), as denoted by bold font. Accordingly, BHHO-FS achieved the highest average accuracy of 99.967%, as shown in Fig. 6. Moreover, GOA achieved 100% over eight datasets, and GA outperformed other algorithms over three datasets only. The progressive choice plan allows search agents to gradually grow their location and only select a better site, which can improve the value of solutions and concentrating capabilities of HHO over the sequence of iterations[30]. GA sometimes quickly detects worthy solutions even for complex search spaces, and the procedure has some drawbacks associated with it. The fundamental disadvantage is that the fitness function of the associated issue must be clearly specified; otherwise, the GA may converge to local optima rather than the global-optimal solution[35]. This explains why the GA algorithm sometimes achieved high classification accuracies but other times was not. The FF algorithm achieved the lowest accuracies because the FF algorithm needs an appropriate parameter setting with a big number of iterations to reach the optimum solution [36]. Due to the speedy convergence rate of PSO, it performs well and subsequently attains high accuracy [37].

Table 2: Comparison between proposed BHHO-FS and state-of-art algorithms based on classification accuracy in 5 iterations

Dataset	BHHO-FS	FF-FS	GA-FS	GOA-FS	PSO-FS	SVM
Abdomen	**100**	81.528	99.954	91.549	99.921	92.958
Bladder	**100**	87.523	99.956	**100**	99.887	90.278
Blood	99.746	86	**99.909**	99.653	99.867	92.014
Bones	**100**	73.894	99.934	78	99.865	88.667
Brain	**100**	75.604	99.949	**100**	99.833	82.222
Breast	**100**	69.736	99.970	**100**	99.926	85.294

Colon	99.613	82.473	**99.954**	99.612	99.866	94.574
Eye	**100**	70.391	99.966	89.655	99.939	86.207
Glands	**100**	82.356	99.970	87.097	99.933	87.097
Heart	**100**	78.688	99.961	93.939	99.956	93.939
Liver	95	80.225	**99.916**	94.444	99.799	78.363
Lungs	**100**	89.626	99.941	**100**	99.843	76.823
Lymph	**100**	79.331	99.952	**100**	99.911	88.71
Naso	**100**	84.488	99.963	**100**	99.923	94.954
Nerve	**100**	74.893	99.969	95.429	99.932	96
Skin	**100**	81.354	99.978	**100**	99.959	99.545
Stomach	**100**	80.378	99.927	**100**	99.820	63.514
Average accuracy	**99.967**	79.911	99.951	95.846	99.893	87.715

Figure 5 : comparison of feature selection accuracies between BHHO-FS and FF-FS, GA-FS, GOA-FS, PSO-FS and SVM over 17 datasets

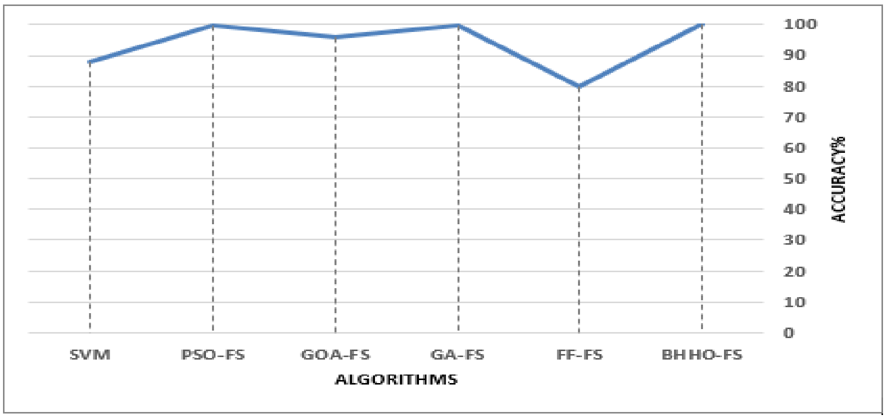

Figure 6: comparison of feature selection average accuracies between BHHO-FS and FF-FS, GA-FS, GOA-FS, PSO-FS and SVM over 17 datasets

Runtime: Obviously, runtime is extremely important to choose the right heuristic optimization algorithm, especially in higher dimensional search spaces [38]. Accordingly, in this study, we take into account calculating the runtime for all applied algorithms. As presented in Table 3, BHHO-FS confirmed its superiority to the FF, GA, GOA and PSO algorithms by consuming fewer runtimes over 8 datasets out of 17 datasets, as denoted by bold font. The minimum runtime has been achieved by BHHO-FS, as HHO performance is quick and competing in determining the right solutions [30]. In contrast, PSO outperformed the highest runtimes (as highlighted in Table 3) due its well-known stagnation ability into local optima, particularly in higher search space [37]. Accordingly, BHHO-FS achieved the lowest average runtime equal to 00:46:05 mm:ss:ms (minutes: seconds: millisecond), as shown in Fig. 7. The proposed BHHO-FS is dominant from the runtime average view, where it consumes the least runtime average in comparison with the other four algorithms because the HHO algorithm runs with a fast Levy and greedy choosing [21].

Table 3: Comparison between proposed BHHO-FS and state-of-art algorithms based on runtime (mm:ss:ms)

Dataset	BHHO-FS	FF-FS	GA-FS	GOA-FS	PSO-FS
Abdomen	**00:02:39**	00:03:11	00:37:84	00:03:11	01:56:77
Bladder	**00:52:91**	01:56:16	03:31:85	01:10:66	12:26:35
Blood	01:28:86	03:05:91	03:03:75	**01:20:01**	11:01:50
Bones	00:08:60	00:08:24	00:50:65	**00:07:00**	02:19.71
Brain	00:54:19	01:10:98	02:13:87	**00:23:85**	07:29:58
Breast	**01:33:20**	03:53:71	04:02:41	04:56:99	16:04:90
Colon	00:49:57	01:19:33	02:34:27	**00:42:53**	05:27:97
Eye	**00:01:19**	00:01:48	00:30:48	00:01:51	01:25:59
Glands	**00:11:70**	00:20:89	01:08:09	00:13:68	03:15:31
Heart	00:01:01	**00:00:28**	00:47:96	00:01:33	01:29:46
Liver	01:04:03	00:58:05	02:31:32	**00:26:51**	05:44:36
Lungs	**01:11:74**	02:50:99	04:19:56	01:15:03	09:16:83
Lymph	03:25:56	03:53:24	**03:23:50**	03:34:60	17:12:40
Naso	**00:13:40**	00:26:21	01:35:91	00:18:19	05:47:91
Nerve	**00:08:25**	00:10:54	00:53:71	00:08:33	02:18:29
Skin	00:16:70	00:31:26	01:17:62	**00:15:31**	04:03:57
Stomach	00:35:21	00:41:85	01:40:34	**00:22:12**	04:46:58
Average	**00:46:05**	01:16:20	02:04:07	00:54:26	06:28:00

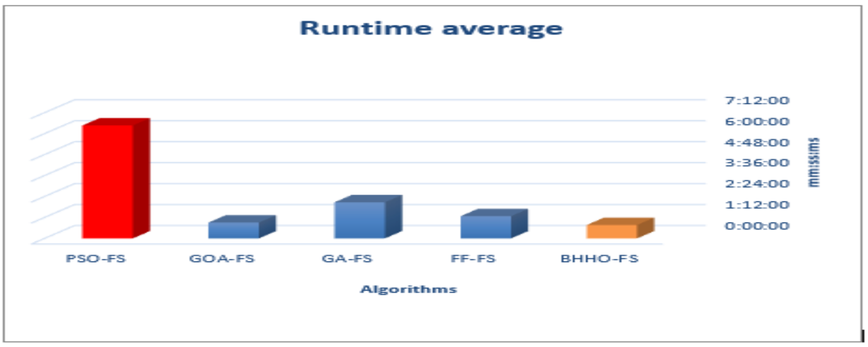

Figure 7: comparison of runtime average accuracies between BHHO-FS and FF-FS, GA-FS, GOA-FS and PSO-FS over 17 datasets

No. of selected features: In feature selection, the premium classifying algorithm must be able to outperform the smallest classification error rate by selecting the minimum number of features [39]. In Table 4 depicted with Figure 8, the minimum number of selected features is determined by the FF algorithm. FF outperformed the other algorithms on 10 datasets, and BHHO-FS outperformed the other algorithms on 8 datasets. As shown in Table 4, the FF and BHHO-FS algorithms achieved the lowest average of the selected features: 5.764 and 6, respectively. The comparison of selected features average between BHHO-FS and FF-FS, GA-FS, GOA-FS and PSO-FS over 17 datasets is depicted in Fig. 9.

Table 4: Comparison between proposed BHHO-FS and state-of-art algorithms based on number of selected features

Dataset	BHHO-FS	FF-FS	GA-FS	GOA-FS	PSO-FS
Abdomen	**4**	7	10	8	**4**
Bladder	**4**	5	8	9	10
Blood	6	6	**5**	11	11
Bones	8	**6**	9	10	10
Brain	**6**	**6**	7	8	7
Breast	8	**5**	**5**	12	9
Colon	7	**5**	6	6	9
Eye	6	**5**	8	11	10
Glands	**4**	7	6	9	8
Heart	7	7	8	10	9
Liver	**5**	**5**	8	7	11
Lungs	**5**	**5**	8	7	7
Lymph	7	6	11	9	10
Naso	**5**	6	7	6	8
Nerve	**5**	**5**	7	8	10
Skin	8	**6**	**6**	8	7
Stomach	7	**6**	8	9	6
Average	**6**	**5.764**	7.470	8.705	8.588

Obviously, BHHO-FS and GOA achieved higher accuracies, fewer runtimes and nearly fewer selected averages. Finally, the minimum average of selected features is obtained by the FF algorithm. To assess the performances of the five mentioned algorithms, we must consider all three metrics. In other words, the victorious algorithm should outperform higher accuracy, less runtime and minimum number of selected features.

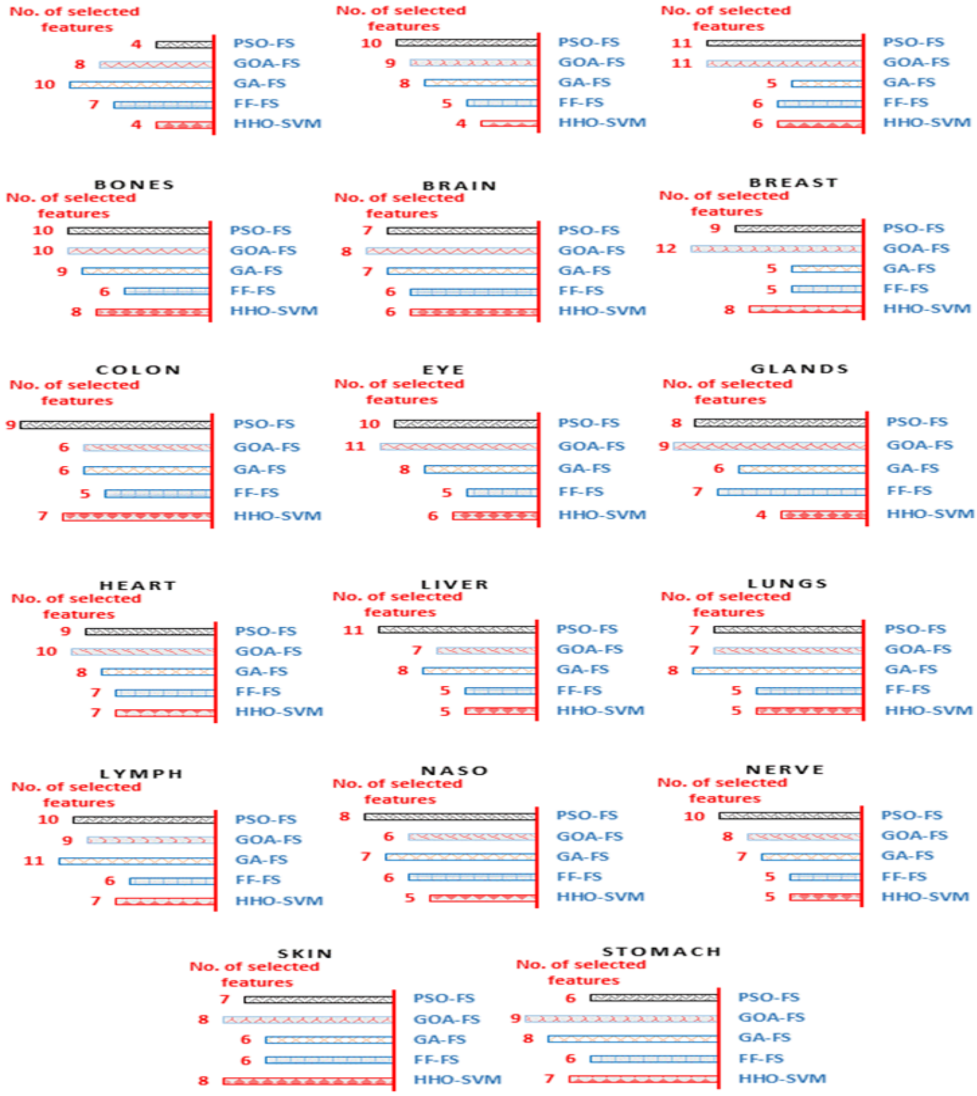

Figure 8 : comparison of no. of selected features between BHHO-FS and FF-FS, GA-FS, GOA-FS and PSO-FS over 17 datasets

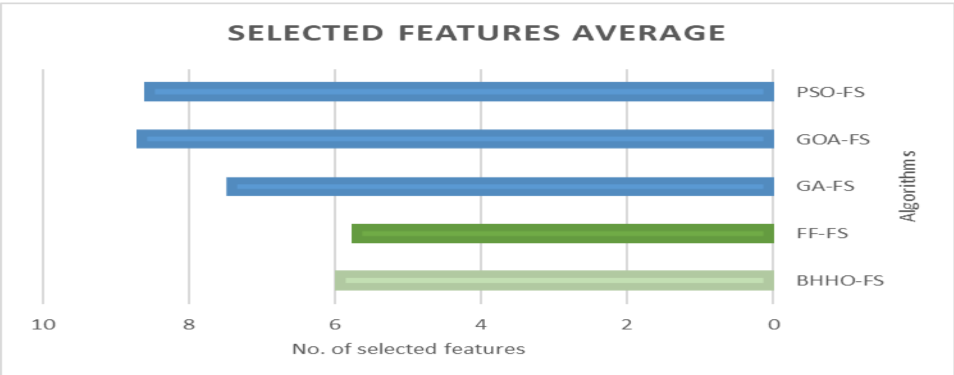

Figure 9: comparison of selected features average between BHHO-FS and FF-FS, GA-FS, GOA-FS and PSO-FS over 17 datasets

6 Conclusion

In this paper, we presented BHHO-FS, a novel strategy for feature selection based on the development of a recently organically inspired algorithm called binary Harris hawk optimization. We ran the proposed BHHO-FS algorithm over 17 real biomedical datasets for Iraqi cancer patients in 2010-2012. The proposed method was compared with four state-of-art algorithms based on three metrics: feature selection accuracy, runtime and number of selected features. The results demonstrate the superiority of the proposed BHHO-FS algorithm in most cases. Due to the high performances of BHHO and GOA, we suggest that future work combine them in one hybrid system for feature selection or any applicable field.

References

[1] S. Hengpraprohm and S. Jungjit, "Ensemble feature selection for breast cancer classification using microarray data," *Inteligencia Artificial*, vol. 23, no. 65, pp. 100–114, Jun. 2020, doi: 10.4114/intartif.vol23iss65pp100-114.

[2] B. Zhang and P. Cao, "Classification of high dimensional biomedical data based on feature selection using redundant removal," *PLoS One*, vol. 14, no. 4, pp. 1–19, 2019, doi: 10.1371/journal.pone.0214406.

[3] N. Maleki, Y. Zeinali, and S. T. A. Niaki, "A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection," *Expert Syst Appl*, vol. 164, no. September 2020, p. 113981, 2021, doi: 10.1016/j.eswa.2020.113981.

[4] V. Kumar, "Feature Selection: A literature Review," *The Smart Computing Review*, vol. 4, no. 3, 2014, doi: 10.6029/smartcr.2014.03.007.

[5] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 3, pp. 131–156, 1997, doi: 10.3233/IDA-1997-1302.

[6] M. Ghaemi and M.-R. Feizi-Derakhshi, "Feature selection using Forest Optimization Algorithm," *Pattern Recognit*, vol. 60, pp. 121–129, 2016, doi: 10.1016/j.patcog.2016.05.012.

[7] C. Staelin, "Parameter selection for support vector machines," *Hewlett-Packard Company, HPL-2002-354 (R.1)*, vol. 354, pp. 1–4, 2003, doi: 10.1016/j.aca.2011.07.027.

[8] V. N. Vapnik, M. Jordan, and J. F-, *The Nature of Statistical Learning Theory ISBN 0-387-98780-0 Springer-Wrlag New York Bertin Heidelberg SPIN 10713304*. 1999.

[9] N. H. Sweilam, A. A. Tharwat, and N. K. Abdel Moniem, "Support vector machine for diagnosis cancer disease: A comparative study," *Egyptian Informatics Journal*, vol. 11, no. 2, pp. 81–92, 2010, doi: 10.1016/j.eij.2010.10.005.

[10] S. Salcedo-Sanz, J. L. Rojo-Álvarez, M. Martínez-Ramón, and G. Camps-Valls, "Support vector machines in engineering: An overview," *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 4, no. 3, pp. 234–267, 2014, doi: 10.1002/widm.1125.

[11] L. Zhang, F. Lin, and B. Zhang, "Support vector machine learning for image retrieval," *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 2, no. x, pp. 721–724 vol.2, 2001, doi: 10.1109/ICIP.2001.958595.

[12] D. R. Tobergte and S. Curtis, "Review on Support Vector Machine (SVM) Classifier for Human Emotion Pattern Recognition from EEG Signals," *J Chem Inf Model*, vol. 53, no. 9, pp. 1689–1699, 2013, doi: 10.1017/CBO9781107415324.004.

[13] M. Lehtokangas, "PATTERN RECOGNITION WITH NOVEL SUPPORT VECTOR MACHINE LEARNING METHOD," in *Signal Processing Conference, 2000 10th European*, 2000.

[14] H. Drucker, D. Wu, and V. N. Vapnik, "Support vector machines for spam categorization," *IEEE Trans Neural Netw*, vol. 10, no. 5, pp. 1048–1054, 1999, doi: 10.1109/72.788645.

[15] B. Moghaddam, "Gender classification using support vector machines," *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, pp. 471–474 vol.2, 2000, doi: 10.1109/ICIP.2000.899454.

[16] L. Hermes and J. M. Buhmann, "Feature Selection for Support Vector Machines," *Proc. {IEEE} Intl. Conf. Pattern Recognition ({ICPR'00})*, vol. 2, pp. 716–719, 2000, doi: 10.1109/ICPR.2000.906174.

[17] E. Emary, H. M. Zawbaa, K. K. A. Ghany, A. E. Hassanien, and B. Parv, "Firefly Optimization Algorithm for Feature Selection," in *Proceedings of the 7th Balkan Conference on Informatics Conference - BCI '15*, 2015, pp. 1–7. doi: 10.1145/2801081.2801091.

[18] C. L. Huang and C. J. Wang, "A GA-based feature selection and parameters optimizationfor support vector machines," *Expert Syst Appl*, vol. 31, no. 2, pp. 231–240, 2006, doi: 10.1016/j.eswa.2005.09.024.

[19] H. T. Ibrahim, W. J. Mazher, O. N. Ucan, and O. Bayat, "A grasshopper optimizer approach for feature selection and optimizing SVM parameters utilizing real biomedical data sets," *Neural Comput Appl*, 2018, doi: 10.1007/s00521-018-3414-4.

[20] L.-Y. Chuang, H.-W. Chang, C.-J. Tu, and C.-H. Yang, "Improved binary PSO for feature selection using gene expression data," *Comput Biol Chem*, vol. 32, pp. 29–38, 2008, doi: 10.1016/j.compbiolchem.2007.09.005.

[21] S. Mirjalili, H. F. Faris, and I. Aljarah, *Machine Learning Techniques Algorithms and Applications*, no. January. 2019. doi: 10.1007/978-981-32-9990-0.

[22] Ministry of Health-Iraq-Iraqi Cancer Board, *Acceptance of Official Cancer datasets from Iraq*. 2017.

[23] R. N. Khushaba, A. Al-ani, A. Al-jumaily, and P. O. Box, "Differential Evolution based Feature Subset Selection," *Evolution (N Y)*, 2008.

[24] E. Emary, W. Yamany, and A. E. Hassanien, "New approach for feature selection based on rough set and bat algorithm," *2014 9th International Conference on Computer Engineering {&} Systems (ICCES)*, pp. 346–353, 2014, doi: 10.1109/ICCES.2014.7030984.

[25] M. Moghadasian and S. P. Hosseini, "Binary Cuckoo Optimization Algorithm for Feature Selection in High-Dimensional Datasets," pp. 18–21, 2014.

[26] E. Emary, HossamM.Zawbaa, C. Grosan, and A. E. Hassenian, *Feature Subset Selection Approach by Gray-Wolf Optimization*, vol. 334. 2015. doi: 10.1007/978-3-319-13572-4.

[27] H. T. Ibrahim, W. J. Mazher, O. N. Ucan, and O. Bayat, "Feature Selection using Salp Swarm Algorithm for Real Biomedical Datasets," *IJCSNS International Journal of Computer Science and Network Security*, vol. 17, no. 12, pp. 13–20, 2017.

[28] H. Deghbouch and F. Debbat, "A hybrid bees algorithm with grasshopper optimization algorithm for optimal deployment of wireless sensor networks," *Inteligencia Artificial*, vol. 24, no. 67, pp. 18–35, 2021, doi: 10.4114/intartif.vol24iss67pp18-35.

[29] J. C. Bednarz, "Cooperative hunting in Harris' hawks (Parabuteo unicinctus)," *Science (1979)*, vol. 239, no. 4847, pp. 1525–1527, 1988, doi: 10.1126/science.239.4847.1525.

[30] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019, doi: 10.1016/j.future.2019.02.028.

[31] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif Intell*, vol. 97, no. 1–2, pp. 273–324, 1997, doi: 10.1016/S0004-3702(97)00043-X.

[32] H. Faris, M. A. Hassonah, A. M. Al-Zoubi, S. Mirjalili, and I. Aljarah, "A multi-verse optimizer approach for feature selection and optimizing SVM parameters based on a robust system architecture," *Neural Comput Appl*, pp. 1–15, 2017, doi: 10.1007/s00521-016-2818-2.

[33] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Syst Appl*, vol. 36, no. 2 PART 2, pp. 3240–3247, 2009, doi: 10.1016/j.eswa.2008.01.009.

[34] A. Mammone, M. Turchi, and N. Cristianini, "Support vector machines," *Engineering*, no. December, pp. 1–39, 2009, doi: 10.1002/wics.049.

[35] J. Guo, J. White, G. Wang, J. Li, and Y. Wang, "A genetic algorithm for optimized feature selection with resource constraints in software product lines," *Journal of Systems and Software*, vol. 84, no. 12, pp. 2208–2221, 2011, doi: 10.1016/j.jss.2011.06.026.

[36] L. Zhang, L. Liu, X.-S. Yang, and Y. Dai, "A Novel Hybrid Firefly Algorithm for Global Optimization," *PLoS One*, vol. 11, no. 9, p. e0163230, 2016, doi: 10.1371/journal.pone.0163230.

[37] M. Li, W. Du, and F. Nian, "An adaptive particle swarm optimization algorithm based on directed weighted complex network," *Math Probl Eng*, vol. 2014, 2014, doi: 10.1155/2014/434972.

[38] J. A. Allen and S. Minton, "Selecting the right heuristic algorithm: Runtime performance predictors," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1081, pp. 41–53, 1996, doi: 10.1007/3-540-61291-2_40.

[39] I. Aljarah *et al.*, "A dynamic locality multi-objective salp swarm algorithm for feature selection," *Comput Ind Eng*, vol. 147, no. July, 2020, doi: 10.1016/j.cie.2020.106.